6-18-2022

ASPECT-BASED SENTIMENT ANALYSIS FOR UNIVERSITY TEACHING ANALYTICS

Tim Schurig
Copenhagen Business School, tim.schurig1996@gmail.com

Sine Zambach
Copenhagen Business School, sz.digi@cbs.dk

Raghava Rao Mukkamala
Copenhagen Business School, rrm.digi@cbs.dk

Malte Petry
Copenhagen Business School, malte.petry@gmx.de

ASPECT-BASED SENTIMENT ANALYSIS FOR UNIVERSITY TEACHING ANALYTICS

Research Paper

Tim Schurig, Copenhagen Business School, Copenhagen, Denmark, and Freie Universität Berlin, Berlin, Germany, tim.schurig@fu-berlin.de

Malte Petry, Copenhagen Business School, Copenhagen, Denmark, malte.petry@gmx.de

Sine Zambach, Copenhagen Business School, Copenhagen, Denmark, sz.digi@cbs.dk

Raghava Rao Mukkamala, Copenhagen Business School, Copenhagen, Denmark rrm.digi@cbs.dk

Abstract

Aspect-based sentiment analysis (ABSA) is a natural language processing method to analyze sentiments from large amounts of unstructured text in a much more fine-grained manner at the aspect level. In this research work, we apply it to analyze open text replies from surveys regarding online teaching. Like most other educational institutions, Copenhagen Business School (CBS) had to shift to online teaching from one day to the next. Using ABSA, we investigated the impact of this forced online learning experiment on teaching quality in the spring semester of 2020. Our findings reveal that students disliked online teaching due to insufficient information and unadjusted teaching methods. However, students liked its flexibility and possibility to learn at an individual pace. We show that ABSA can extract valuable information in an easily interpretable manner to support teaching and learning processes. Finally, our findings show that ABSA is a valuable tool to analyze unstructured text quantitatively.

Keywords: Aspect-Based Sentiment Analysis, Natural Language Processing, Covid-19, Open Text Surveys, Online Teaching.

1 Introduction

Schools and universities had to immediately close to confine the global outbreak of the Covid-19 pandemic. For instance, in April 2020, more than 90% of all enrolled students globally were affected by this.[1] Consequently, educational institutions were forced to provide alternative teaching formats to keep educating their students without much preparation. Correspondingly, at CBS, a Danish higher education institution, all activities continued from home for the remainder of the spring semester of 2020. Universities were not fully prepared for this radical shift as the situation was new for both students and teachers and raised questions on the quality, popularity, and acceptance of this kind of education (Aucejo et al., 2020; Gonzalez et al., 2020; Sahu, 2020).

Due to its unprecedented nature, the literature still lacks knowledge on implications of the Covid-19 pandemic on university teaching (Johnson et al., 2020; Said, 2021). However, analyzing the implications on teaching quality quantitatively with data-driven methods leads to more accurate and evidence-based decisions and thereby enables better future actions based on this (Brynjolfsson and McAfee, 2012).

[1]UNESCO (2020), “COVID-19: a global crisis for teaching and learning”, https://unesdoc.unesco.org/ark:/48223/pf0000373233, accessed: May 6, 2021.

Regarding the current situation, many blog posts or expert opinions were published[2,3], but there are not many verifiable findings even though most educational institutions regularly perform student survey evaluations. They often include open text questions, for evaluating their satisfaction and the perceived teaching quality (Chen and Hoshower, 2003; Spooren et al., 2013; Wachtel, 1998). Teaching evaluations are particularly important during these times as reports have mentioned online teaching's potential negative psychological, economic, and social implications.[4,5] Therefore, it is highly relevant for CBS to analyze how their employees managed to teach during Covid-19. As the gathered responses are too many to read individually, responsive actions are hindered if the responses are not analyzed automatically.

When dealing with open text survey data, problems arise from a time-consuming, subjective, and prone to error human coding process (Chen et al., 2018; Liew et al., 2014). It becomes infeasible to read through them manually when they become too many. Also, the focus in learning analytics (LA) has so far neglected quantitative text mining approaches such as aspect-based sentiment analysis (ABSA) (Aldowah et al., 2019). In order to demonstrate how to generate objective and data-driven insights from open text data, we use the specific situation related to Covid-19 as an illustrative context, as this field currently lacks quantifiable knowledge. To do so, we are using the natural language processing (NLP) technique ABSA, which builds on the commonly applied sentiment analysis, but goes beyond as it focuses on identifying fine-grained opinions towards aspects (Nazir et al., 2020).

At CBS, after the spring semester of 2020, a regular annual questionnaire was sent to the students to evaluate their courses and teachers. Furthermore, a second student survey was tailored to identify the students' experiences with Covid-19. Thus, we have access to student surveys reflecting on their experiences during the Covid-19 lockdown as well as general teaching comments. By leveraging the written student survey comments, we are answering the following research question:

How can we use ABSA to extract information from textual data of university evaluations in an easily interpretable manner to support teaching and learning in higher education?

To answer our research question, we furthermore defined the following two research propositions:

RP1: Which online teaching aspects did students like most and least?

RP2: Which aspects did the students like most (least) about the best (worst) performing teachers?

As a result, we developed a method that automatically scans through the students' opinions, analyzes them, and reveals positive and negative teaching-related factors. Therefore, our research will enable the organization to become more efficient and inclusive by incorporating the students' perspectives into higher-level decision-making processes. We do not only reveal aspects which students liked and disliked about forced online teaching during Covid-19, but on a more generalizable level, we also contribute with an illustration on how ABSA can be leveraged to reveal findings that are supporting decision-making processes in large (higher education) organizations using large amounts of open text replies.

The paper is structured as follows: We start by introducing related work regarding student surveys and difficulties when working with open text comments, the usage of ABSA in LA, and online teaching during Covid-19. Then, we explain the NLP method used in our paper and present our results. Finally, we discuss our results, reflect on the limitations of our study, and outline future research directions.

---

[2] Winkler, T. (2020), "10 Steps To Go Digital With Your Teaching In Times of Crisis" https://www.panopto.com/blog/10-steps-to-go-digital-with-your-teaching-in-times-of-crisis/, accessed: March 17, 2022.

[3] Rapanta, C., Botturi, L., Goodyear, P., Guàrdia, L. & Koole, M. (2020) "Online University Teaching During and After the Covid-19 Crisis: Refocusing Teacher Presence and Learning Activity", https://link.springer.com/article/10.1007/s42438-020-00155-y accessed: June 25, 2021.

[4] UNESCO (2022) "The impact of the COVID-19 pandemic on education: international evidence from the Responses to Educational Disruption Survey", https://unesdoc.unesco.org/ark:/48223/pf0000380398, accessed: March 17, 2022.

[5] OECD (2020), "Education responses to covid-19: Embracing digital learning and online collaboration", https://read.oecd-ilibrary.org/view/?ref=120_120544-8ksud7oaj2&title=Education_responses_to_Covid-19_%20Embracing_digital_learning_and_online_collaboration, accessed: June 25, 2021.

2 Related Work and Theoretical Background

In this section, we present the theoretical foundation of our work. We do so with the help of a literature review in which we outline the status quo in historical-conceptual terms to identify the mentioned research gaps (vom Brocke et al., 2015; Schryen, 2015; Webster and Watson, 2011).

2.1 Student Surveys, Open Text and ABSA

Student surveys and evaluations are among the most common data sources for quality assessment in higher education. They are often used for evidence-based university decision-making as they are inexpensive, fast and easy to process (Klemenčič and Chirikov, 2015; Williams, 2014). These surveys include student learning outcomes assessment, student course evaluations, and student experience, engagement, and satisfaction surveys (Chen and Hoshower, 2003; Klemenčič and Chirikov, 2015). Particularly, course evaluations are used to improve teaching and learning as a form of constructive alignment of the course's learning activities and the student's outcomes (Biggs, 1996).

In student evaluations and other surveys, the advantage of open-ended text questions compared to Likert-scale questions is that they can also identify factors that were not explicitly targeted in the questions (Spooren et al., 2013). This means that open-ended questions can express unconstrained opinions about some aspects, for example, the teacher, teaching methods, etc. Therefore, they provide important diagnostic information and insights that the survey designer might not have thought of (Calderon et al., 1996). Additionally, surveys mostly use Likert-scale data, but due to their ordinal character, they have been widely criticized when statistically interpreting them (Harpe, 2015; Ogden and Lo, 2012).

However, due to its unstructured nature, the open-ended text is more complicated to analyze, summarize and report (Chen and Hoshower, 2003; Popping, 2015). For instance, traditional methods such as human coding or structured content analysis are time-consuming, subjective, and prone to human errors (Chen et al., 2018; Liew et al., 2014; Strauss, 1987). Furthermore, they do not scale well with the increasing size of the corpus. The bigger the text corpora becomes, the less practical traditional human coding approaches are.

Therefore, the focus shifted to quantitative methods based on NLP and text mining approaches (Buenaño-Fernandez et al., 2020). However, they have only been researched in an educational context to a limited extent (Aldowah et al., 2019). Educational text mining explores the application of NLP techniques in an educational context. There, it has previously mainly been used for student evaluations, course completion prediction, student support or motivation, or analyzing the students' sentiment in discussion forums (Ferreira-Mello et al., 2019; Romero and Ventura, 2013). More concrete, it has been used to improve the teacher's ability to assess the progress of group discussions (Dringus and Ellis, 2005) or to gain deeper insights into online participation and achievement in online learning contexts (He, 2013).

However, only a few papers are available that applied NLP techniques to open text student surveys. Available papers only used qualitative coding approaches and counted word frequency and co-occurrences. These simple approaches do not leverage NLP's full potential. According to Aldowah et al. (2019), only 4.75% of the papers which used data mining techniques within higher education research leverage text mining approaches.

Additionally, ABSA has not received much attention in the educational context. Only a few papers in this area have used the method so far. For example, Ramesh et al. (2015) apply a weakly supervised technique on discussion forum texts from MOOCs by seeding words in a topic model and subsequently extracting aspects and assigning polarities. Chauhan et al. (2019) use rules on part-of-speech tags and a lexicon-based approach to student feedback from social media. Valakunde and Patwardhan (2013) evaluate faculty performance by analyzing student feedback on an aspect level and subsequently aggregating it using a weighted approach.

Due to the fragmented nature of text comments in students' evaluations and the vast amount of unstructured text, we apply ABSA. This method addresses the mentioned problems and makes it possible to identify fine-grained opinions towards self-reliantly identifiable aspects. We do so, as current research mainly focuses on traditional coding approaches when handling student evaluations and the small share of papers using NLP in an educational research context. Student evaluations produce vast amounts of unstructured text, which are too complex to be analyzed manually.

2.2 Online Teaching, Emergency Remote Teaching, and Increased Inequality

As our corpus is about online teaching during Covid-19, we provide a summary of existing online teaching literature. Online teaching classes should have well-structured course content, well-prepared instructors, clear instructions, and feedback elements involved (Bolliger, 2004; Sun and Chen, 2016). Virtually, social interactions and relations (both on a student and student-teacher level) are relevant for the students' learning satisfaction (Richardson et al., 2017). Students with less social interaction are less motivated and satisfied with their online courses (Trolian et al., 2016). Social engagement opportunities are important for students' learning outcomes (Redmond et al., 2018). The students' satisfaction with online learning depends on the teaching style and the teachers' ability to express themselves clearly (Bickle et al., 2019; Cheng et al., 2017). Online teaching is tailored towards the learners' flexibility and speed by being time- and location-independent (Dumford and Miller, 2018).

Studies have shown that students slightly prefer face-to-face courses over online courses due to, for instance, technology-related fear, anger, and helplessness in virtual environments (Allen et al., 2002; Tratnik et al., 2019). To make online learning successful students need to have the perception that they engage with other humans and have the possibility to develop personal relationships, which is supported if students have a sense of belonging to a learning community (Händel et al., 2020).

During the pandemic, the educational sector was forced to implement distance learning approaches to an extent it had not done before without having sufficient time to prepare (Bozkurt et al., 2020; Hodges et al., 2020). Therefore, the impact of the emergency remote learning situation on higher education should not be compared with carefully in advance planned online learning approaches (Hodges et al., 2020).

In the situation of this forced shift, there were concerns on equality and inclusion as in online learning environments, good students are becoming better, and weaker students are likely to become more vulnerable (Cavanaugh and Jacquemin, 2015). Also, it is reported that educational inequality increased in terms of the digital divide, which is defined as having vs. not having access to the infrastructure needed for online learning (Kang, 2021). Gan and Sun (2021) found that almost a third of the students at a US university experienced digital barriers when migrating to online platforms. Three-quarters of the digital barriers were reported by students coming from low-income households (Gan and Sun, 2021). During Covid-19, the skill gap between students from low- and high-income families increased, and lower-income students were 55% more likely than higher-income students to have a delayed graduation. The GPA gap almost doubled in the US between these two groups (Aucejo et al., 2020).

3 Methodology

ABSA builds on sentiment analysis, which describes the task of analyzing subjective expressions such as opinions from a text (Feldman, 2013; Liu, 2010). The goal is to classify a text into categories (e.g., positive, neutral, and negative) or attributing a score to it (e.g., -1 to +1). With the rising amount of opinionated text available on the internet, combined with its value for practical applications, sentiment analysis grew explosively in popularity (Liu, 2010). An example is to extract the sentiment about online teaching, as seen in Figure 1. Overall, the opinion about the course is good. Thus, the sentiment should be "positive". However, often it is not enough to calculate an overall sentiment of a text. It is also of interest which aspects of online teaching are good or bad. This is where ABSA comes into play. It was first discussed by Hu and Liu (2004) and does not just assign an overall sentiment to a text but assigns a sentiment to every aspect. Thus, different aspects of a corpus are revealed.

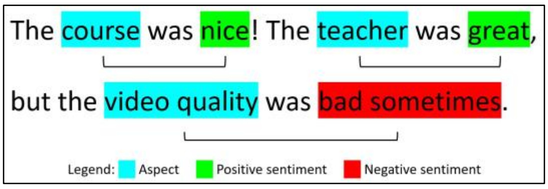

Figure 1 Example Sentence for Aspect-Based Sentiment Analysis

As ABSA works on the aspect-level and not the document level, aspects must be determined initially. According to Pavlopoulos and Androutsopoulos (2014), the steps for ABSA are as follows: Aspect term extraction, aspect term sentiment estimation, and aspect aggregation. Aspect term extraction identifies the different aspects of a text. Next, the sentiment for these aspects is determined. This includes the polarity, as well as the intensity with words, like, e.g., “very” or “extremely”. Lastly, similar extracted aspects are grouped together to achieve coarser aspects.

Datasets

We used data from two student surveys conducted at our university. The Semester Based Student Course Evaluation is sent out regularly. In addition, the Covid-19 Specific Student Survey was explicitly created by a project group to collect data on the experiences at the university during the Covid-19 lockdown in the spring of 2020. From the Covid-19 Specific Student Survey, we analyzed the two open text comments of the questions “Please explain what you liked the most about the online teaching” and “Please explain what you liked the least about the online teaching”. There were 1,150 comments (mean word count: 14.39) regarding what students liked most about online teaching and 1,209 comments (mean word count: 20.25) about what students liked least about online teaching. From the respondents to these text questions, 55% were studying a master’s degree, 38% a bachelor’s degree, and the remaining 7% a single subject. 60% of them are female and 40% male. 69% are Danish and 31% international.

To also find aspects students liked and disliked most about their teachers, we investigated how students evaluated them. For this, we use the course evaluations of the spring semester of 2020. They include a section that focuses on the teachers. We analyzed the question “What would you like to tell the teacher about the teaching?” and 52% of the respondents were bachelor’s, 35% master’s, and 13% single-subject students. 63% of them are female and 37% male. 76% are Danish, and the remaining 24% are international. Table 1 shows summary statistics for both surveys.

Table 1 Summary Statistics for both Surveys

Survey Name	**Sent to**	**Completed**	**Partly Answered**	**Teachers Evaluated**
Covid-19 Specific Student Survey	25,720	1,805 (7.0%)	134 (0.5%)	N/A
Semester Based Student Course Evaluations	40,571	9,348 (23.0%)	776 (1.9%)	725

We wanted to find out what the teachers with the best overall ratings did well and what the teachers who received the worst overall feedback did particularly bad. Therefore, we examined the students’ answers to the following question from the course evaluation: “Indicate to what extent you agree with the statement: Teacher x was overall a good teacher”. We calculated the overall metric for the courses a teacher taught. To ensure representative results, at least five students needed to answer the corresponding question of the evaluation. If a teacher taught in more than one course, we macro averaged the metrics over all his courses to limit the influence of large courses. If a teacher received a mean score of 5 from 10 students in the *course a* and a mean score of 3 in *course b* from 20 students, the overall rating is 4.

Based on this metric, we identified the 50 best teachers (average score: 4.14) and the 50 worst teachers (average score: 2.69). The number 50 was chosen as a trade-off between including too many mediocre teachers and having a too-small subset of data based on the exploration of our primary datasets. Then, we investigated the comments these teachers received on the question “What would you like to tell this teacher about the teaching?”. By doing so, we generated two corpora, one focusing on the feedback of the best-performing (531 comments; mean word count: 34.97) and the other one on the worst-performing teachers (568 comments; mean word count: 58.45).

Data Preprocessing

As many of the text fields contain a mix of comments in English and Danish, the first step was to translate everything into English. This was done via the Google Cloud API. As the translation was conducted via a third party, we were particularly diligent regarding personal data processing. All data was pseudonymized before receiving it. Nevertheless - to ensure that no personal data was missed in the text fields - before translating the text, we looked at a sample of the responses to ensure that no full names could identify a natural person in the text.

Our ABSA approach uses dependency parsing to extract the aspects and the corresponding sentiment. This requires light preprocessing of the text. First, we split the comments into sentences and lowercased all words. We extended contractions in the texts (e.g., “didn’t” -> “did not”) to ensure that negations were discovered by dependency parsing. To extract aspect-sentiment pairs from the comments, we used dependency parsing from the spaCy package[6] and added customized rules. To create the rules, we adapted an approach by a project that analyzed Amazon reviews[7]. We used three rules that account for different sentence structures. Each sentence is sequentially checked with all three rules. The algorithm extracts aspects and their sentiment towards the aspect, including the intensity and negation.

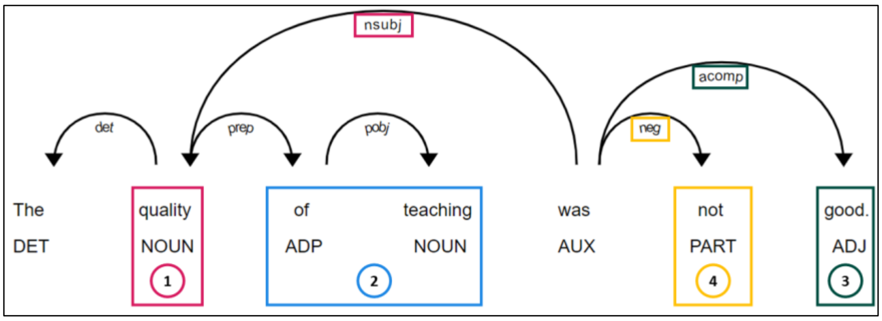

Figure 2: Process of dependency parsing rules.

Figure 2 describes the process of the rules with an example sentence for rule 2. In the first step, aspects are extracted by finding tokens that are objects or subjects of a sentence. Next, it is also checked if the aspect is a compound of two nouns (e.g., “quality of teaching”). Subsequently, it is searched for an adjectival complement that belongs to the object/subject. If one is found, we found an aspect-sentiment pair. A third rule implemented to find a pair just searches for adjectival modifiers and their corresponding nouns. Before returning the pair, further checks are implemented to account for negations

Table 2 Example Sentences for Dependency Parsing Rules

Rule	Sentence	Aspect	Sentiment	Negation
#1	There could have been a much better teaching quality.	teaching quality	much better	True
#2	The quality of teaching was not good.	teaching quality	Good	True
#3	It has been difficult to form an overview.	overview	Difficult	False

[6] spaCy documentation, https://spacy.io/usage/linguistic-features#dependency-parse, accessed: May 23, 2021.

[7] Aspect Extraction and Opinion Analysis - Building natural language models to extract insights from Amazon reviews, https://achyutjoshi.github.io/aspect_extraction/overview, accessed: May 2, 2021.

(e.g., "The connection was not good.") and intensity modifiers (e.g., "very good"). The negation check includes modal verbs to account for sentences like "The teaching could have been better." An example sentence for every rule used can be found in Table 2.

To determine a polarity score for each sentiment and negation, we used a combination of three libraries: TextBlob[8], NLTK[9], and flair[10]. TextBlob and NLTK's VADER are rule-based sentiment analyzers, while flair uses a pre-trained transformer-based model that detects positive and negative sentiment. TextBlob's sentiment analyzer and NLTK's VADER return a score between -1 and +1, while flair's sentiment analyzer only returns a positive or negative label and a confidence level. We tested the three libraries on a part of our corpus and noticed two things. First, flair often returns high confidence levels for positive/negative even if the sentiment would be neutral. Thus, we decided to use +1 if flair returned "positive" and -1 if flair returned "negative" with a confidence level of at least 0.9. Second, there is not a single library that is superior for all sentences that we tested. Therefore, to limit the influence of the strengths and weaknesses of a single library, we used an average of the three libraries. This is similar to ensemble methods which make use of the wisdom of the crowd (Polikar, 2012). For every score, if we found a negation for our sentiment, the sign will be flipped. Continuing with our example sentence for rule 2, we extracted "good" and identified a negation clause. TextBlob returns a polarity score of 0.7, NLTK's VADER 0.44, and flair 1. As the boolean variable for negation is true, the sign will be flipped, and the resulting scores are -0.7, -0.44, and -1, respectively. After averaging the scores, we obtain -0.71 as the final polarity score. Therefore, from the sentence "The quality of teaching was not good." our approach extracts the aspect "teaching quality" and assigns a polarity score of -0.71 to it.

After the whole corpus is checked according to the rules, we lemmatize the aspects and then group them by (sum up by) the lemmatized aspects. Lemmatization reduces a word down to its base form. This has the advantage that, e.g. plural and singular forms of a noun will end up being the same word. Therefore, lemmatization ensures that aspects are grouped together when the students commented on the same thing. Next, we sort the resulting list by the summed score. The aspects with the highest scores will then be the aspects which students were most frequently positive about and vice versa. It would have also been a possibility just to take the average of each aspect. The approach we used has the advantage of taking both into account the average polarity score, and the number of times an aspect was mentioned.

Clustering the aspects with an algorithm proved to be complicated. This could be because some aspects refer to multiple different things and the aspects are all quite closely linked to teaching. Thus, we clustered the top 20 aspects of each list manually with the option to assign multiple categories. This ensures that we find the exact reasons.

Due to the nature of the text we are analyzing, our approach tends to be more positive than negative. Sentences like "Getting sufficient information is essential!" will be labeled with a positive score even though the writer's intention was the opposite. Therefore, the scores on the positive side are higher.

Methodology Evaluation

To iteratively improve our approach, after building and implementing our method, we conducted semi-structured interviews to evaluate the method. We conducted three interviews which lasted for 30 up to 60 minutes depending on the interviewees need to communicate. We interviewed the teaching evaluation unit from our university that handles evaluation results and has done most of the overall evaluation of students' assessment of Covid-19 teaching. Additionally, we interviewed the university's largest study board (which uses the results for study improvement) and the head of a department (who focuses on teacher developments based on the students' comments). The interviews took place online and were recorded due to quality assurance in accordance with the interviewee's consent. We talked about current issues when working with survey data, the meaningfulness of our method for their daily work, the implications of our findings for their work, and the potential for improvement. We took their insights into consideration for the future process of this project.

[8] TextBlob documentation, https://textblob.readthedocs.io/en/dev/, accessed: May 23, 2021.
[9] NLTK documentation, https://www.nltk.org/, accessed: May 23, 2021.
[10] flair documentation, https://github.com/flairNLP/flair, accessed: May 23, 2021.

4 Results

This section outlines positive and negative factors about online teaching. We also look at aspects students liked most about the best-performing teachers and least about the worst-performing teachers.

4.1 Aspects students liked most and least about online teaching

For what students liked most about online teaching, we show the 20 aspects with the highest polarity scores of the column "Please explain what you liked the most about online teaching". Similarly, for what students liked least about online teaching, we show the 20 aspects with the lowest scores of the column "Please explain what you liked the least about online teaching". The results are shown in Tables 3 & 4.

Table 3 Aspects that students liked most

Lemmatized Aspect	Category	Polarity
time	Flexibility	22.7
lecture	Flexibility	21.9
flexibility	Flexibility	19.9
teaching	Teaching	17.3
video	Pace	9.4
pace	Pace	6.9
opportunity	Pace	6.2
nothing	Misc	6.0
teacher	Teaching	5.2
schedule	Flexibility	4.6
class	Misc	4.3
freedom	Flexibility	4.2
life	Misc	3.9
note	Pace	3.2
understanding	Pace	3.1
help	Pace	2.9
format	Misc	2.9
hour	Misc	2.9
session	Misc	2.8
tool	Misc	2.5

Table 4 Aspects that students liked least

Lemmatized Aspect	Category	Polarity
quality	Teaching, Technology	-6.8
connection	Technology	-5.1
Communication	Teaching, Information	-5.0
motivation	Motivation	-4.2
Teacher	Teaching	-3.5
teaching	Teaching	-2.9
structure	Teaching	-2.5
opportunity	Teaching	-1.9
Video	Technology	-1.8
coordination	Information	-1.7
question	Teaching	-1.6
Grade	Misc	-1.5
Fence	Misc	-1.4
group	Teaching	-1.3
benefit	Misc	-1.3
professor	Teaching	-1.3
Feedback	Teaching	-1.2
concentration	Motivation	-1.2
explanation	Teaching	-1.1
Invite	Misc	-1.0

From the top 20 aspects, we manually identified seven categories. Some appear only on one side of the table, while others are on both sides. The categories and the corresponding summed scores over the top 20 are *Flexibility* (73.3), *Pace* (31.7), *Teaching* (-0.7), *Information* (-4.2), *Motivation* (-5.4), *Technology* (-10.3), and *Misc* (20.1). The *Misc* category is for aspects that did not have a clear pattern in the comments. Furthermore, sometimes aspects referred to two different categories (quality, communication). If aspects had two categories, the score was split between them.

The positive-only category *Flexibility* refers to the increased flexibility students had. They did not need to come to the university to attend a lecture and could watch the lecture at a time that suited them. This allowed them to manage their work and university schedules better. One exemplary comment is:

"The flexibility due to teachers recording lectures gave us the opportunity to watch it at a later time if we had work, rather than simply missing it completely like we would've had with normal teaching."

Another positive-only category is *Pace*. This describes the opportunity that students could watch lectures at their own pace. If they did not understand something, it was possible to rewatch the corresponding part. One student wrote:

"I can take it at my own pace, take breaks when I need it, rewind if I don't understand it."

On the positive side of the category *Teaching* students appreciated the efforts made by teachers to adapt to online teaching in such a short period and were satisfied with online teaching. One comment was:

"Some of the teachers really did their best in making it interactive, as if we are still in a classroom, which I really liked and appreciated."

However, *Teaching* also appears on the negative side. Some students found the teaching *quality*, in general, was worse. Furthermore, many complained about a lack of *communication* or one-sided *communication* in class. Moreover, the feedback was described as much lower than in regular classes, which was partly caused by a lack of opportunity to ask questions. Interestingly, the aspect *fence* also refers to *Teaching*. This is due to the Danish expression "skipping where the fence is lowest", which refers to teachers not putting much effort into their teaching.

"Lack of discussion in class - and thereby other people's input and questions."

Information is another negative-only category. It is also the other category for the aspect of *communication*. Students described that they did not receive sufficient information about online teaching. Issues were that the information came too late, was scattered over different platforms, and many complained that the calendar function within the Learning Management Systems was not working anymore. A comment addressing that issue was:

"There was basically too little communication from the lecturers to us, and when information came, it was put in different places, so it got a little confusing."

Another problem for students was the lack of *Motivation*. Students did not have the same motivation as for regular lectures in online teaching and were struggling to concentrate:

"It can be very difficult to look into a screen for several hours in a row, and simultaneously the motivation has been sharply declining."

Technology refers to problems with the technical practicalities of online teaching. First, some teachers were struggling with the technical aspects of online teaching and did not know how the tools work. As a result, some of the teachers were not able to adapt their teaching to online teaching. Many students complained about bad sound or video quality. Furthermore, connection problems were common. Students described that either they or the teacher had internet connection problems. These disruptions extended lectures or students missed certain parts of the class:

"Internet downtime, disconnections during workshops."

4.2 Aspects on a Teacher Level

To obtain insights into the factors students appreciated most (least) about the best-performing (worst-performing) teachers, we looked at the students' text comments to the question, "What would you like to tell this teacher about the teaching".

Tables 5 and 6 show the identified aspects. The corresponding categories are *General* (203.6), *Method* (49.5), *Knowledge* (22.3), *Behavior* (7.0), and *Technology* (-1.6). Finding categories for the aspects was more difficult than in the previous section, as many identified aspects are not very specific. They are often just an overall assessment of the teacher, so the *General* category dominates both tables.

The *General* category on the positive side describes that the teacher was, for example, "super" and "excellent". Many students were satisfied with their teachers. However, there were also many negative aspects for the *General* category. The following quote is from a student that was satisfied with his teacher.

"Best teacher at our university, it would be great if all teachers were like this, so we could actually learn a lot."

Table 5 Aspects that students liked most about best-performing teachers

Lemmatized Aspect	Category	Polarity
teacher	General	94.2
teaching	General	30.2
course	General	19.8
job	General	18.2
way	Method	17.7
class	General	12.8
experience	Knowledge	12.6
energy	Behavior	12.5
example	Method	11.8
work	General	10.5
lecture	General	9.5
approach	Method	8.6
life	General	8.1
lecturer	General	7.7
structure	Method	7.4
explanation	Method	7.3
professor	General	7.0
feedback	Method	5.4
level	Knowledge	4.9
knowledge	Knowledge	4.8

Table 6 Aspects that students listed least *about worst-performing teachers*

Lemmatized Aspect	Category	Polarity
time	Method	-4.4
powerpoints	Method	-3.0
teaching	General	-2.6
curriculum	General	-1.7
subject	General	-1.7
people	General	-1.7
attitude	Behavior	-1.7
tone	Behavior	-1.6
fence	General	-1.4
teaching method	Method	-1.3
effort	Behavior	-1.3
thread	General	-1.0
area	General	-1.0
motivation	Behavior	-0.9
figure	General	-0.9
difficult	General	-0.8
deadline any	General	-0.8
lecture video	Technology	-0.8
module	General	-0.8
sound	Technology	-0.8

The *Method* category describes how the teachers designed their online teaching. Some teachers struggled with time management and had bad presentations. However, it was appreciated if the teacher used examples and provided detailed, constructive feedback. A good structure and ability to explain were also positive. One student said:

"The teacher gave a lot of good feedback, which made students want to ask questions. Even when I said something wrong, he would rephrase it to not offend me. It made me not afraid of being wrong, which facilitated long discussions."

The category *Knowledge* focuses on two things. First, it was appreciated if the teacher was experienced at teaching. Second, it was valued if the teacher linked the theory to their own experience. For example, one student mentioned:

"In his teaching, the teacher gives really good examples from his own experience and everyday life which puts the material into a good context so that it is better understood."

The *Behavior* of a teacher also had an influence on the students. If the teacher was energetic and motivated himself, it helped the students be as well. However, some teachers did not put much effort into a course which was demotivating for the students. The attitude of some teachers was even described as arrogant. One example is provided below:

"Arrogant attitude to teaching with the attitude' I know best - you must not believe you are anything."

Lastly, *Technology* was a problem again. As described above, it refers to problems with the technical practicalities of online teaching.

5 Discussion

One of the most important factors that students disliked about their situation in the spring semester was that they had insufficient information about online teaching. It was found that both teachers and the university did not provide sufficient information. This was caused by generally not having information available online; it being spread across different platforms or being delivered too late. Specifically for the university, the problem was that the calendar function was deactivated, which showed the times of the classes. Besides, insufficient information could have also been caused by the sudden change from regular to online teaching. Furthermore, some of the teaching methods were not appropriate for online teaching. The teaching must be adapted to the online setting and cannot just continue like it was on-campus. There was less in-class discussion than in regular classes. Students perceived the communication in classes as one-sided and not interactive. Increased interaction positively influenced the students' satisfaction of undergraduate students, as it can be seen in other studies in, e.g., South Korea and India (Baber, 2020). Students also complained that they were not getting enough feedback during the semester. Technical problems arose during online teaching. Bad internet connection, as well as video and sound problems, were other common issues.

However, there were also positive aspects associated with online teaching. Students enjoyed the flexibility. Online teaching is independent of the location and, in the case of pre-recorded videos, also independent of time. This allowed students to align their work and university schedules. Another positive aspect is that students can learn at their own pace. Recorded lectures allow rewatching parts that were hard to understand. Therefore, the learning process is adjusted to the individual student's pace.

Moreover, many students were satisfied with some of the teachers. However, there were also some who did not put in much effort. To find more insights on this, we identified factors that students liked most/least about the best/worst-performing teachers. By looking at the most negative aspects of the worst teachers, we were able to identify the teaching practices students did not like and teachers should therefore avoid. The most successful teachers were appreciated due to their ability to provide detailed and constructive feedback and to express themselves in clear language. Furthermore, it seems like their lectures were structured, they were successful in engaging their students, used many examples to support their teaching, and explained the curriculum well. These are probably factors with universal validity, and they do not only hold for online teaching. Overall, factors previously investigated in the literature review regarding good and bad online teaching practices were also found with the help of our method based on the students' comments. Furthermore, it seems that good and bad online teaching practices during the emergency Covid-19 pandemic are in accordance with general online teaching findings. None of the investigated factors was new or surprising.

Regarding the applied method - ABSA - we found out that this method can be helpful for the application in an educational context, but also beyond in other contexts. The identified aspects were not deductively and a-priori defined; instead, they were derived from the corpus itself. As our identified aspects about good and bad online teaching practices were in line with findings from the literature, it seems that the aspect identification and polarity rankings give insights into the students' teaching assessment. This is not true for the subsequent manual identification of the categories, but this was only done for illustrative reasons. In future studies, aspect topic modeling approaches, such as Latent Dirichlet Allocation (LDA), can be explored to a larger extent. In our research, LDA did not work well because the identified aspects were too similar as they all relate to learning and teaching. Therefore, the modeling algorithms could not semantically distinguish between the different aspects, and the cluster coherence was low. Additionally, the indicated polarity score is an objective quantification of the intensity and frequency of the occurred aspect, which is also an advantage over traditional coding methods.

Furthermore, applying ABSA not only on one, but on two different text corpora with similar results in good and bad online teaching practices showed that our method works well and is robust and reliable. In the semi-structured interviews, we found that it is hard for the university to access the evaluation comments since resources for reading through every comment, splitting them up, and categorizing them do not exist yet. While the evaluation unit of our university currently distributes them to every department, each study board needs to extract those comments regarding the teaching and split the aspects manually. This was reported to be very time-consuming. Additionally, the Heads of Departments need the assessments of the teachers, both to get a general overview of the employees of the department and to help each teacher develop in teaching in higher education. Opposingly to traditional approaches, the larger the corpus, the more insightful and better are the results.

Furthermore, the interview partners at the university reported that the created aspect tables give them a quick high-level overview about multiple text comments and factors which already went well and factors that can further be improved during online teaching. The method we created also makes it possible to quickly show exemplary comments in the python notebook for the mentioned aspects. This further gives more meaning to them. In a subsequent step, the insights could be visualized in automatic reports or even visually appealing dashboards. The surveys we had access to had low completion rates. We cannot falsify nor verify whether the share of students who answered the surveys are representative of the whole student population (participation/non-response bias). It might be possible that really dissatisfied students might have completed the surveys to express their opinions more frequently. This can underrepresent other groups (e.g., students who were neutral/positive towards courses in online teaching but might not feel the necessity to complete a survey, and consequently, answers from this group are not submitted with the same frequency).

Furthermore, our data only regards one university. Therefore, we cannot draw conclusions about the situation at other universities or even in other countries. Technologically, we are located in one of the most advanced areas in the world. Therefore, the infrastructural conditions are different compared to other countries. For example, Adnan and Anwar (2020) conclude that in underdeveloped countries, online classes cannot produce the desired academic effect because many students cannot access the internet or do not have the devices needed for online learning. At the university in our study, almost every student has his own device, and therefore most do not face problems to that extent.

Another peculiarity about our university is that it is a business school. Compared to other universities, which offer a more comprehensive range of degrees, this one only offers management and business-related degrees. Compared to courses of study in, for example, medicine or chemistry, exercise sessions for business students do not take place in a lab and are discussion-based, making it easier to convert them into an online format. Future research could look at similar text comments, but with survey data from different universities within our country or even from universities in other less digitally advanced nations. This would make it possible to reveal university- and or countrywide similarities and differences, reducing the subjectivity which is always incorporated in surveys.

Some of the ways to improve our method could be along the following lines. The first improvement could be related to the dependency parsing approach we utilized. The polarity scores were too positive because of our approach, e.g., sentence structures not accounted for. Furthermore, the rule-based approach cannot extract cross-sentence dependencies. Since the results from our approach were already sufficient to give actionable recommendations, we did not explore more advanced NLP models. For example, another alternative is to use deep learning transformer models, which can learn grammar and context much better and thereby reveal more complex relations between the words. An additional way to improve the method is to use advanced machine learning approaches for dependency parsing, which would have led to even better and more accurate results. Similarly, we combined the information of three different modules to calculate the sentiment of a word. Using deep learning approaches to learn a model to automatically calculate the polarity scores would have probably led to even more precise polarity scores. Compared to deep learning methods, our approach has the advantages of being relatively simple, reproducible, and the outcomes do not lack explainability and transparency. Our approach also works better than human coding approaches - as outlined in the related work - which is laborious, error-prone, and subjective. With the help of ABSA, we identified the aspects objectively, and there is no subjective bias involved in the analysis. They were determined by the dependency rules we specified and the three modules we used to calculate the sentiment. Consequently, we developed code that helps to find positive and negative aspects as an individual Python module, which can easily be applied in the evaluation of other open text comments within the educational sector, but also beyond in all kinds of different settings and industries. Furthermore, future research could also quantitatively compare the results of the three libraries used, as the quantitative evaluation was beyond the scope of this work.

Our work is currently still only cross-sectional. Subsequent semesters also took place online. Thus, we could identify any potential trends or patterns by incorporating these surveys and applying our module to these open text comments. Longitudinal studies would increase our ability to say with confidence that the investigated impacts were caused by Covid-19 and not by other factors.Furthermore, our quantitative study could also be complemented with qualitative research. We plan to conduct interviews with teachers who have extremely good or bad online teaching scores. We identified these teachers and what aspects they did, particularly good or bad. Still, we do not know why for example, the worst-performing teachers put less effort into interacting, engaging, and giving feedback to their students. Revealing the root causes would make it possible to improve the overall teaching quality prospectively.

6 Conclusion

Our objective was to show how ABSA can be utilized to extract information from comments in students' evaluations in an easily interpretable manner to support teaching and learning. For this, we used online learning during Covid-19 as an illustrative context. We utilized different open text data from university evaluation surveys. We contribute to the research community on the implications of Covid-19 on university teaching. To the best of our knowledge, we are the first ones to analyze multiple data sources in northern Europe to provide a broader perspective on the topic. Apart from Covid-19 specific contributions, we are among the first to apply ABSA to unstructured evaluation text. This technique proved to be particularly useful for the open text fields in our surveys and can also be helpful in other settings. It gives employees working with survey data the opportunity to get a quick overview of the most positive and negative aspects without reading through all comments manually. We found insights such that students did not receive sufficient information, better teaching methods could have been used to improve engagement and feedback, and students had motivation or concentration problems. However, positive factors for the students were flexibility and the ability to study at their own pace.

As the pandemic is still present at the time of writing, the need for guidance in online teaching is high. Missing best practices lead to sub-optimal teaching, which leaves many students dissatisfied with their education. The positive and negative factors that we extracted help in improving online teaching. Our findings are subject to some limitations. First, as our analysis is based on survey data. We cannot verify whether our sample is representative and whether the answers provided are objectively correct. Second, our data is from a university focused on business and economics, based in a developed and well-digitized country. This limits the generalizability of our results. Future research can conduct longitudinal studies about the impact of Covid-19 on teaching. Some limitations of this work can be addressed by looking at similar questions across other universities and or countries.

References

Adnan, M. and Anwar, K. (2020), “Online learning amid the COVID-19 pandemic: Students perspectives”, *Journal of Pedagogical Sociology and Psychology*, Vol. 1 No. 2, pp. 45–51.

Aldowah, H., Al-Samarraie, H. and Fauzy, W.M. (2019a), “Educational data mining and learning analytics for 21st century higher education: A review and synthesis”, *Telematics and Informatics*, Elsevier, Vol. 37, pp. 13–49.

Allen, M., Bourhis, J., Burrell, N. and Mabry, E. (2002), “Comparing Student Satisfaction With Distance Education to Traditional Classrooms in Higher Education: A Meta-Analysis”, *American Journal of Distance Education*, Vol. 16 No. 2, pp. 83–97.

Aucejo, E.M., French, J., Ugalde Araya, M.P. and Zafar, B. (2020), “The impact of COVID-19 on student experiences and expectations: Evidence from a survey”, *Journal of Public Economics*, Vol. 191, pp. 1–15.

Baber, H. (2020), “Determinants of Students’ Perceived Learning Outcome and Satisfaction in Online Learning during the Pandemic of COVID19”, *Journal of Education and E-Learning Research*, Vol. 7 No. 3, pp. 285–292.

Bickle, M.C., Rucker, R.D. and Burnsed, K.A. (2019), “Online Learning: Examination of Attributes That Promote Student Satisfaction.”, *Online Journal of Distance Learning Administration*, ERIC, Vol. 22 No. 1, pp. 1–7.

Biggs, J. (1996), “Enhancing teaching through constructive alignment”, *Higher Education*, Vol. 32 No. 3, pp. 347–364.

Bolliger, D.U. (2004), “Key factors for determining student satisfaction in online courses”, *International Journal on E-Learning*, Association for the Advancement of Computing in Education (AACE), Vol. 3 No. 1, pp. 61–67.

Bozkurt, A., Jung, I., Xiao, J., Vladimirschi, V., Schuwer, R., Egorov, G., Lambert, S., *et al.* (2020), “A global outlook to the interruption of education due to COVID-19 Pandemic: Navigating in a time of uncertainty and crisis”, *Asian Journal of Distance Education*, Vol. 15 No. 1, pp. 1–126.

vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R. and Cleven, A. (2015), “Standing on the shoulders of giants: Challenges and recommendations of literature search in information systems research”, *Communications of the Association for Information Systems*, Vol. 37 No. 9, pp. 205–224.

Brynjolfsson, E. and McAfee, A. (2012), “Big Data: The Management Revolution”, *Harvard Business Review*, Vol. 90 No. 10, pp. 60–68.

Buenaño-Fernandez, D., González, M., Gil, D. and Luján-Mora, S. (2020), “Text Mining of Open-Ended Questions in Self-Assessment of University Teachers: An LDA Topic Modeling Approach”, *IEEE Access*, Vol. 8, pp. 35318–35330.

Calderon, T.G., Gabbin, A.L. and Green, B.P. (1996), *A Framework for Encouraging Effective Teaching*, American Accounting Association Center for Research in Accounting Education, James Madison University.

Cavanaugh, J. and Jacquemin, S.J. (2015), “A Large Sample Comparison of Grade Based Student Learning Outcomes in Online vs. Face-to-Face Courses”, *Online Learning*, Vol. 19 No. 2, pp. 1–8.

Chauhan, G.S., Agrawal, P. and Meena, Y.K. (2019), “Aspect-Based Sentiment Analysis of Students’ Feedback to Improve Teaching–Learning Process”, in Howlett, R.J. and Jain, L.C. (Eds.), *Information and Communication Technology for Intelligent Systems*, Springer, Singapore, pp. 259–266.

Chen, N.C., Drouhard, M., Kocielnik, R., Suh, J. and Aragon, C.R. (2018), "Using machine learning to support qualitative coding in social science: Shifting the focus to ambiguity", *ACM Transactions on Interactive Intelligent Systems*, Vol. 8 No. 2, pp.1–20.

Chen, Y. and Hoshower, L.B. (2003a), "Student Evaluation of Teaching Effectiveness: An assessment of student perception and motivation", *Assessment & Evaluation in Higher Education*, Vol. 28 No. 1, pp. 71–88.

Cheng, F.-F., Chiu, C.-C., Wu, C.-S. and Tsaih, D.-C. (2017), "The influence of learning style on satisfaction and learning effectiveness in the asynchronous web-based learning system", *Library Hi Tech*, Vol. 35 No. 4, pp. 473–489.

Feldman, R. (2013), "Techniques and applications for sentiment analysis", *Communications of the ACM*, ACM New York, NY, USA, Vol. 56 No. 4, pp. 82–89.

Ferreira-Mello, R., André, M., Pinheiro, A., Costa, E. and Romero, C. (2019), "Text mining in education", *WIREs Data Mining and Knowledge Discovery*, Vol. 9 No. 6, pp. 1–49.

Gan, I. and Sun, R. (2021), "Digital Divide and Digital Barriers in Distance Education during COVID-19", *Hawaii International Conference on System Sciences 2021*, pp. 1–10.

Gonzalez, T., de la Rubia, M.A., Hincz, K.P., Comas-Lopez, M., Subirats, L., Fort, S. and Sacha, G.M. (2020), "Influence of COVID-19 confinement in students' performance in higher education", edited by Xie, H.*PLOS ONE*, Vol. 15 No. 10, pp. 1–23.

Händel, M., Stephan, M., Gläser-Zikuda, M., Kopp, B., Bedenlier, S. and Ziegler, A. (2020), "Digital readiness and its effects on higher education students' socio-emotional perceptions in the context of the COVID-19 pandemic", *Journal of Research on Technology in Education*, Routledge, pp. 1–15.

Harpe, S.E. (2015), "How to analyze Likert and other rating scale data", *Currents in Pharmacy Teaching and Learning*, Vol. 7 No. 6, pp. 836–850.

He, W. (2013), "Examining students' online interaction in a live video streaming environment using data mining and text mining", *Computers in Human Behavior*, Vol. 29 No. 1, pp. 90–102.

Hodges, C., Moore, S., Lockee, B., Trust, T. and Bond, A. (2020), "The difference between emergency remote teaching and online learning", *Educause Review*, Vol. 27, pp. 1–12.

Hu, M. and Liu, B. (2004), "Mining and summarizing customer reviews", *Proceedings of the 2004 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '04*, ACM Press, New York, New York, USA, pp. 168–177.

Johnson, N., Veletsianos, G. and Seaman, J. (2020), "U.S. faculty and administrators' experiences and approaches in the early weeks of the COVID-19 pandemic", *Online Learning Journal*, Vol. 24 No. 2, pp. 6–21.

Kang, B. (2021), "How the COVID-19 Pandemic Is Reshaping the Education Service", in Lee, J. and Han, S.H. (Eds.), *The Future of Service Post-COVID-19 Pandemic*, Springer, Singapore, pp. 15–36.

Klemenčič, M. and Chirikov, I. (2015), "How Do We Know How Students Experience Higher Education? On the Use of Student Surveys", *The European Higher Education Area*, Springer International Publishing, Cham, pp. 361–379.

Liew, J.S.Y., McCracken, N., Zhou, S. and Crowston, K. (2014), "Optimizing features in active machine learning for complex qualitative content analysis", *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science*, pp. 44–48.

Liu, B. (2010), "Sentiment Analysis and Subjectivity", in Indurkhya, N. and Damerau, F.J. (Eds.), *Handbook of Natural Language Processing*, Chapman & Hall, pp. 627–666.

Nazir, A., Rao, Y., Wu, L. and Sun, L. (2020), "Issues and Challenges of Aspect-based Sentiment Analysis: A Comprehensive Survey", *IEEE Transactions on Affective Computing*, pp. 1–20.

Ogden, J. and Lo, J. (2012), "How meaningful are data from Likert scales? An evaluation of how ratings are made and the role of the response shift in the socially disadvantaged", *Journal of Health Psychology*, Sage Publications Sage UK: London, England, Vol. 17 No. 3, pp. 350–361.

Pavlopoulos, J. and Androutsopoulos, I. (2014), "Aspect Term Extraction for Sentiment Analysis: New Datasets, New Evaluation Measures and an Improved Unsupervised Method", *Proceedings of the 5th Workshop on Language Analysis for Social Media (LASM) @ EACL 2014*, pp. 44–52.

Polikar, R. (2012), "Ensemble learning", *Ensemble Machine Learning*, Springer, pp. 1–34.

Popping, R. (2015), "Analyzing Open-ended Questions by Means of Text Analysis Procedures", *Bulletin of Sociological Methodology*, Vol. 128 No. 1, pp. 23–39.

Ramesh, A., Kumar, S.H., Foulds, J. and Getoor, L. (2015), "Weakly Supervised Models of Aspect-Sentimentfor Online Course Discussion Forums", *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pp. 74–83.

Redmond, P., Heffernan, A., Abawi, L., Brown, A. and Henderson, R. (2018), "An Online Engagement Framework for Higher Education", *Online Learning*, Vol. 22 No. 1, pp. 183–204.

Richardson, J.C., Maeda, Y., Lv, J. and Caskurlu, S. (2017), "Social presence in relation to students' satisfaction and learning in the online environment: A meta-analysis", *Computers in Human Behavior*, Vol. 71, pp. 402–417.

Sahu, P. (2020), "Closure of Universities Due to Coronavirus Disease 2019 (COVID-19): Impact on Education and Mental Health of Students and Academic Staff", *Cureus*, Vol. 12 No. 4, pp. 1–6.

Said, G.R. El. (2021), "How Did the COVID-19 Pandemic Affect Higher Education Learning Experience? An Empirical Investigation of Learners' Academic Performance at a University in a Developing Country", edited by Mandl, T.*Advances in Human-Computer Interaction*, pp. 1–10.

Schryen, G. (2015), "Writing qualitative is literature reviews—Guidelines for synthesis, interpretation, and guidance of research", *Communications of the Association for Information Systems*, Vol. 37 No. 12, pp. 286–325.

Spooren, P., Brockx, B. and Mortelmans, D. (2013), "On the Validity of Student Evaluation of Teaching", *Review of Educational Research*, Vol. 83 No. 4, pp. 598–642.

Strauss, A.L. (1987), *Qualitative Analysis for Social Scientists*, Cambridge university press.

Sun, A. and Chen, X. (2016), "Online education and its effective practice: A research review", *Journal of Information Technology Education*, Vol. 15, pp. 157–190.

Tratnik, A., Urh, M. and Jereb, E. (2019), "Student satisfaction with an online and a face-to-face Business English course in a higher education context", *Innovations in Education and Teaching International*, Vol. 56 No. 1, pp. 36–45.

Trolian, T.L., Jach, E.A., Hanson, J.M. and Pascarella, E.T. (2016), "Influencing Academic Motivation: The Effects of Student–Faculty Interaction", *Journal of College Student Development*, Vol. 57 No. 7, pp. 810–826.

Valakunde, N.D. and Patwardhan, M.S. (2013), "Multi-aspect and Multi-class Based Document Sentiment Analysis of Educational Data Catering Accreditation Process", *2013 International Conference on Cloud & Ubiquitous Computing & Emerging Technologies*, IEEE, pp. 188–192.

Wachtel, H.K. (1998), "Student Evaluation of College Teaching Effectiveness: a brief review", *Assessment & Evaluation in Higher Education*, Vol. 23 No. 2, pp. 191–212.

Webster, J. and Watson, R.T. (2011), "Analyzing the Past To Prepare for the Future : Writing a Review", *MIS Quartely*, Society for Information Management and The Management Information Systems Research Center, Minneapolis, MN, USA, Vol. 26 No. 2, pp. xiii–xxiii.

Williams, J. (2014), “Student feedback on the experience of higher education: a significant component of institutional research data”, *Using Data to Improve Higher Education*, Brill Sense, pp. 65–80.