# Text-mining student comments: insights gained from two instruments

**Dr Elena Zaitseva**

Teaching & Learning Academy

Liverpool John Moores University

# Liverpool John Moores University

- North-West England

- Five Faculties, 17 Schools.

- 27,199 students (21,628 undergraduates).

- Using Blue for module evaluations (ME) since 2016

# Institutional arrangements for ME

- Overall responsibility for ME lies with the Teaching and Learning Academy (focus on enhancement; research expertise)

- Collaboration with Registry (Blue Admin side)

- The ME data is also used for bespoke, enhancement focused research and evaluation projects

- ME comments (institutional level) are analysed on ad hoc basis, using semantic analysis software Leximancer

- Introduced to BlueML in 2021.

# Background to this research exercise

- Challenges of 2020-21 academic year (virtually all teaching online)

- HEIs in the North West of England were particularly affected

- Low engagement with institutional surveys, ME including, following by unexpectedly low satisfaction scores in the National Student Survey (NSS)

- Undertook institutional analysis of ME comments with Leximancer to get additional insights

- Took up Explorance offer for the comments to be analysed with BlueML.

- To explore what each instrument offers in terms of understanding student experience during pandemic

- To reflect on the strengths and limitations of each type of analysis

- To see how both types of analysis can contribute to informing institutional actions.

**Data set**

- Combined data set – responses to 2 open questions

*Please comment on the most interesting aspect of this module.*

*Please comment on how this module could be improved.*

- 33000 comments.
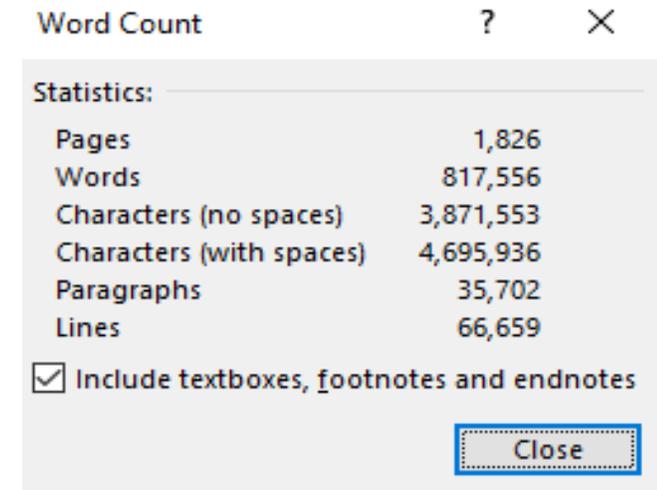
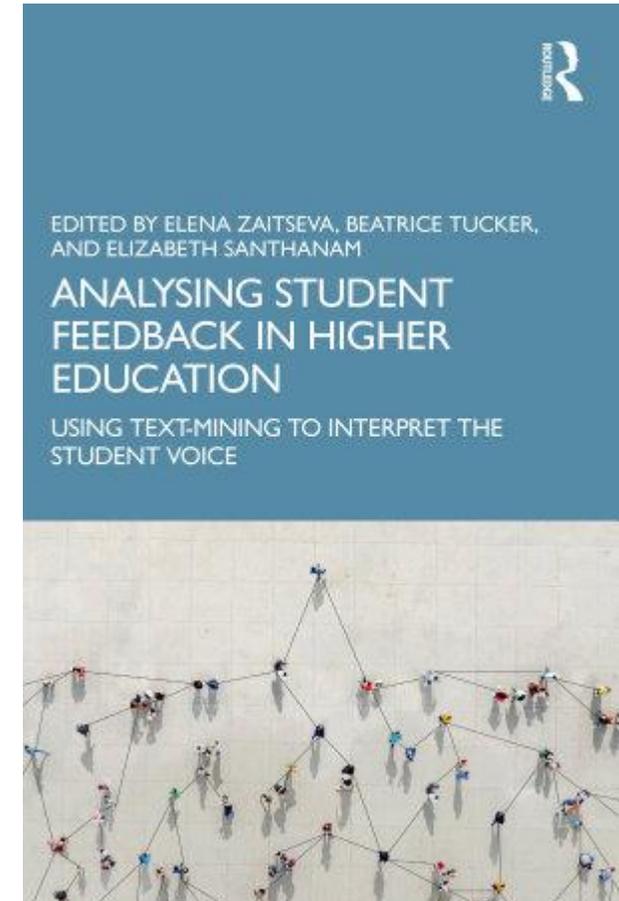| Word Count | ? | × |
|---|---|---|
| **Statistics:** | | |
| Pages | | 1,826 |
| Words | | 817,556 |
| Characters (no spaces) | | 3,871,553 |
| Characters (with spaces) | | 4,695,936 |
| Paragraphs | | 35,702 |
| Lines | | 66,659 |

☑ Include textboxes, footnotes and endnotes

Close

# Text-mining student feedback in HE

- Wealth of techniques, algorithms and platform for extracting meaning from the textual data available
  - A number of factors at play: researcher/analyst's expertise/skills and preferences
  - availability of licenced instruments or text-analytics blocks powered by/integrated into the survey platforms
- Still limited use withing the HE sector.

EDITED BY ELENA ZAITSEVA, BEATRICE TUCKER, AND ELIZABETH SANTHANAM

ANALYSING STUDENT FEEDBACK IN HIGHER EDUCATION

USING TEXT-MINING TO INTERPRET THE STUDENT VOICE

- The software allows instant interaction with large qualitative datasets to reveal semantic characteristics of the text and patterns in the data.
- Identifies concepts, themes and connections between them, including their relevance and proximity.
- Sentiment analysis can be enabled as well as concept 'seeding'.
- Visual, intuitive, shows a 'landscape' of your data, user-friendly.
- Need to read the text segments behind each concept to extract themes.
- Use fragments of comments (not full ones)
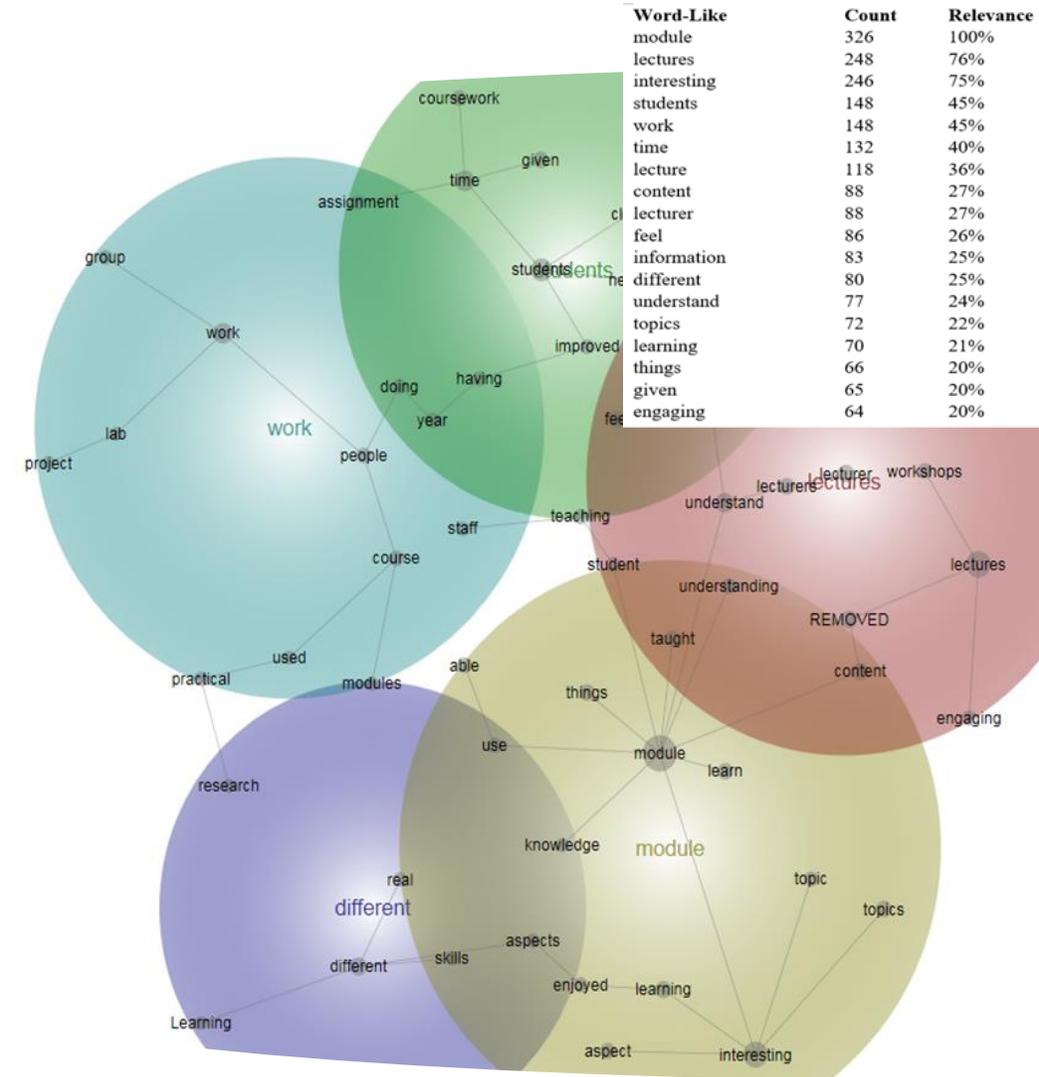
| Word-Like | Count | Relevance |
|---|---|---|
| module | 326 | 100% |
| lectures | 248 | 76% |
| interesting | 246 | 75% |
| students | 148 | 45% |
| work | 148 | 45% |
| time | 132 | 40% |
| lecture | 118 | 36% |
| content | 88 | 27% |
| lecturer | 88 | 27% |
| feel | 86 | 26% |
| information | 83 | 25% |
| different | 80 | 25% |
| understand | 77 | 24% |
| topics | 72 | 22% |
| learning | 70 | 21% |
| things | 66 | 20% |
| given | 65 | 20% |
| engaging | 64 | 20% |

| Word-Like | Count | Relevance | |
|---|---|---|---|
| module | 7496 | 100% | |
| lectures | 4756 | 63% | |
| interesting | 4625 | 62% | |
| favourable | 4220 | 56% | |
| different | 2632 | 35% | |
| learning | 2628 | 35% | |
| work | 2491 | 33% | |
| time | 2444 | 33% | |
| understand | 2132 | 28% | |
| feel | 2125 | 28% | |
| students | 2097 | 28% | |
| information | 2020 | 27% | |
| online | 2016 | 27% | |
| enjoyed | 1913 | 26% | |
| unfavourable | 1782 | 24% | |
| lecture | 1741 | 23% | |
| teaching | 1715 | 23% | |
| questions | 1680 | 22% | |
| taught | 1565 | 21% | |
| content | 1550 | 21% | |
| sessions | 1537 | 21% | |
| assignment | 1503 | 20% | |
| difficult | 1482 | 20% | |
| better | 1469 | 20% | |
| helpful | 1439 | 19% | |
| real | 1316 | 18% | |
| topic | 1295 | 17% | |
| research | 1276 | 17% | |
| week | 1270 | 17% | |
| zoom | 1243 | 17% | |

# Leximancer:  quick insights

- 68 concepts extracted
- Areas needed attention are identified based on relevance and strong negative sentiment expressed
- Institutional narrative was dominated by **lectures, learning, work, time, understand**
- **Difficult** has a relatively high relevance as well as **research**
- Favourable sentiment is more prominent
- **Better**  - calling for actions

# Sentiment and themes

| Concept | Count | Relevance % | Fav % | Unfav % |
|---|---|---|---|---|
| module | 7496 | 100 | 35 | 32 |
| lectures | 4756 | 63 | 19 | 21 |
| work | 2491 | 33 | 9 | 11 |
| time | 2444 | 33 | 8 | 14 |
| understand | 2132 | 28 | 9 | 18 |
| feel | 2125 | 28 | 8 | 13 |
| students | 2097 | 28 | 8 | 11 |
| information | 2020 | 27 | 8 | 11 |
| online | 2016 | 27 | 8 | 15 |
| lecture | 1741 | 23 | 6 | 7 |
| teaching | 1715 | 23 | 8 | 9 |
| questions | 1680 | 22 | 9 | 10 |
| assignment | 1503 | 20 | 5 | 7 |
| difficult | 1482 | 20 | 4 | 50 |
| better | 1469 | 20 | 5 | 8 |
| helpful | 1439 | 19 | 24 | 5 |
| zoom | 1243 | 17 | 5 | 7 |
| group | 1189 | 16 | 4 | 8 |
| due | 1104 | 15 | 3 | 10 |
| assessment | 1092 | 15 | 3 | 5 |
| knowledge | 973 | 13 | 4 | 2 |
| live | 911 | 12 | 3 | 5 |
| year | 823 | 11 | 3 | 6 |
| reading | 817 | 11 | 3 | 4 |

# Leximancer insights

**Lectures** (live, lecture slides, zoom, online sessions, questions)

- *more live lectures instead of recorded*

- *pre-recorded lecture videos were quite long, break the lectures up a bit more*

- *slides used for lectures to be uploaded to canvas*

Selected Concept: **lectures (4756)**  |  Ranked View | Export▾

| Sentiment | Count | Likelihood | |
|---|---|---|---|
| 🔍 unfavourable | 380 | 21% | |
| 🔍 favourable | 801 | 19% | |

| Related Name-Like | Count | Likelihood | |
|---|---|---|---|
| 🔍 NAME | 305 | 25% | |
| 🔍 *Learning* | 28 | 02% | |

| Related Word-Like | Count | Likelihood | |
|---|---|---|---|
| 🔍 live | 821 | 90% | |
| 🔍 slides | 277 | 42% | |
| 🔍 lecture | 696 | 40% | |
| 🔍 zoom | 457 | 37% | |
| 🔍 engaging | 369 | 32% | |
| 🔍 sessions | 480 | 31% | |
| 🔍 questions | 511 | 30% | |
| 🔍 lecturer | 281 | 30% | |
| 🔍 week | 369 | 29% | |
| 🔍 difficult | 428 | 29% | |
| 🔍 online | 575 | 29% | |
| 🔍 lecturers | 290 | 28% | |
| 🔍 having | 284 | 27% | |
| 🔍 time | 648 | 27% | |
| 🔍 better | 381 | 26% | |
| 🔍 understand | 552 | 26% | |
| 🔍 modules | 268 | 26% | |
| 🔍 content | 398 | 26% | |
| 🔍 reading | 207 | 25% | |
| 🔍 information | 509 | 25% | |
| 🔍 feel | 531 | 25% | |
| 🔍 helpful | 359 | 25% | |
| 🔍 student | 146 | 24% | |
| 🔍 students | 500 | 24% | |

**Understand** (difficult, student, year, lecturers)

- *slow down, more interactive, make it easier to understand, more visuals less boring*

- *find it difficult to understand where to look for things on the canvas page*

- *it could ve been better if we would ve been able to go in the labs more, but I understand it s difficult under the current situation*

**Due (to)** (online, difficult, student, time) - reasons, explanations, reflections

- *practicals have been difficult to run due to the restrictions*

- *difficult to contact staff or peers due to not being at university*

- *due to pandemic, motivation levels for completing work are low*

- *group work has been a struggle, due to it being online*

**Reading** (slides, week, given)

- *not reading from the slides, more interactive during the lecture to keep people engaged and listening*

- *clear outlines on what each week's focus is (specific reading etc)*

- *it can be difficult to motivate yourself to engage in all the online material, I find it much easier when a task of reading/watching something is set before a session*

- *... recommended reading could only be accessed through library textbooks... a lot of us can't access the library / aren't living in Liverpool at the moment...*

**Better** (recommendations)

- *more examples and  practise papers*

- *better pace*

- *better communication about late or cancelled sessions*

- *better explanation on how to do the assignments*

- better Canvas page organisation

- *…would be much better face to face*

- Purpose built comments' analysis solution for users of the platform
- Uses own machine learning models/algorithms
- Trained on HE vocabulary – context specific categories

**Attractive features:**
- Comments are linked to demographics (zoom in/out possible)
- Sentiment/polarity analysis
- Ability to filter comments  by responses provided for a 'key' question
- Data driven recommendations
- Work with full quotes/instances

# BlueML: quick insights

Received analysis ready for use/interpretation (user-friendly)

Algorithm is strong for identification of positive sentiment, less so for negative

Where comments are categorised – it is mostly accurate.

| Total Comments Provided | Categorized | Percentage |
|---|---|---|
| 32923 | 8754 | 27% |

| Categorization | # Predictions | % Predictions |
|---|---|---|
| Positive | 12592 | 92.94% |
| Negative | 756 | 5.58% |
| The sentiment is not explicit | 201 | 1.48% |
| Grand Total | 13549 | 100.00% |

| Source | Count of Predictions |
|---|---|
| Q4_Please comment on how this mod | 1836 |
| Q3_Please comment on the most inte | 11713 |
| Grand Total | 13549 |

| #Predictions | Sentiment | | | |
|---|---|---|---|---|
| Categorization | Negative | Positive | The sentim | Grand Total |
| Course component | 170 | 7781 | | 7951 |
| Persons | 62 | 2731 | 90 | 2883 |
| Teaching component | 53 | 1363 | 70 | 1486 |
| Miscellaneous | 459 | 284 | 41 | 784 |
| Learning environment | 3 | 361 | | 364 |
| Learning outcomes | | 72 | | 72 |
| Other element | 9 | | | 9 |
| Grand Total | 756 | 12592 | 201 | 13549 |

**Course component**

Practical learning

- strong positive sentiment about seminars, workshops, labs, practicals, group work, tutorials, case studies, portfolio, real life scenarios, simulated practice, statistical tests.

Course material and structure

- problems with reading lists, workload/balance, access to literature, resources (daunting, overwhelming, crammed, rushed, brushed over)

[Work is a large and fuzzy concept in the Leximancer – and workload is rarely picked up as a separate concept]

- too many comments with positive sentiment (4413) – no aid to break down

**Miscellaneous – large group with important categories**

- Remote learning

- Academic support

- Institution in general

- Objectives & expectations

- Feedback

- Expenses/fees

- Services

- Student wellbeing

**Assessment & evaluation** (only positive)

- *the assignments were set early on and the marking guidelines published which was good*

- *I think that the lecturers are great and the assignments very challenging*

**Objectives and expectations** (only positive)

- *very well structured with clear expectations for the outset*

**Expenses/fees** (only negative)

Value for money is associated with live, engaging teaching

**Recommendations**

- Demographic group insights are valuable!

Mature students

- *more manageable workload*

- *face to face teaching to escape constant disturbances at home*

-  *more support via dedicated online sessions*

Many **recommendations**  are not tagged/categorised

- *more clear instruction about the assignment*

- **Not categorised comments** – the number is very high at the moment

- *The lectures delivered by the tutors are interesting  However, the ILX dashboard is very dull to follow and too slide show heavy  wording is quite confusing making the questions very difficult*

- *Personally I don't think vivas are a good way of accessing our skills, I think osce assessment are a much more realistic and accurate evaluation of our skills  Talking about what to do is so much harder than actually physically doing what you need to, especially for me as I am a visual learner*

- *More live lectures*

Complexity of language is the most difficult part!

- Using two instruments provided different angles for exploration and understanding of the student experience

- Leximancer leaves more room for interpretation in terms of constructing the meaning

- It is a 'niche' (one researcher) instrument, relies on consistency of analysis (settings) and expertise available

- Top level institutional insights are identified  via concept relevance and sentiment, but 'unpacking' concepts is time consuming as some are fuzzy

- Longitudinal exploration of changes in the institutional landscape adds value.

- Blue ML is more streamlined and efficient in slicing the data (where the algorithm works – it does the categorisation very well)

-  Usability is higher  - programme leaders and senior manages will love it!

- Instant insight into the voice of different group of students is very valuable

-  Some categories are too big  for efficient interpretation at institutional level

- Others are too small and less reliable.

- Fine tuning sentiment analysis and recommendations algorithm will improve the BLueML usability and attractiveness to HEIs.