Using qualitative student evaluation data to illuminate quantitative scale item ratings: seeing beyond the numbers

Stuart Palmer[a], Wayne Hall[b]
Melbourne Centre for the Study of Higher Education[a], Griffith School of Engineering and Built Environment[b]
Corresponding Author Email: stuart.palmer@unimelb.edu.au

Introduction

Strong and lightweight fibre reinforced polymeric composites now dominate the aerospace, marine and low-volume automotive sectors. These new materials are driving innovation in engineering and design, and there is a need for current engineering students to be exposed to composite design and manufacture courses in their undergraduate curriculum (Boyles et al., 2011; Hayhurst, Kedward, Soh, & Turner, 2011). In 2014, a new course (subject) in design and manufacturing with composite materials, '6505ENG Manufacturing with Composites' (referred to as '6505ENG' hereafter), was offered at the Griffith School of Engineering and Built Environment. A previous evaluation of the initial course offering found that the students generally perceived the course to be valuable, and the evaluation also offered insights into potential areas for enhancement of the course (Hall & Palmer, 2015). Over the subsequent three offerings of the course (in 2016, 2017 and 2018), a range of deliberate changes to the course learning design were made, with the aim of improving student learning and engagement. An inspection of the data from the university student experience of course (SEC) evaluation instrument for the first four course offerings showed essentially no change in mean ratings for the quantitative scale items, even though aspects of the course learning design had been deliberately changed.

Limitations with the ability of quantitative scale item student evaluation of teaching (SET) data to reveal meaningful variation or provide useful guidance in response to changes in learning designs are described in the literature (Huybers, 2014). In particular, five-point Likert-type scales, as typically used in university SET instruments, can exhibit a strong 'ceiling effect', due to there being a limited range of response points for students to use to record an improved perception of teaching, especially if the course/teacher already rates above average (Keeley, English, Irons, & Henslee, 2013; Menges & Brinko, 1986). Note that Griffith University also uses a student experience of teaching survey (referred to locally as SET), but this survey is not considered here, and 'SET' hereafter refers to the more general abbreviation 'student evaluation of teaching'.

The Griffith University SEC instrument also includes the option for students to provide open-ended text comments about the course. An investigation was undertaken to determine if computer-based analysis of the student comments (text analytics) could identify differences in the students' perceptions of the course that related to the changes in the course learning design over the first four course offerings, and hence whether this data source could supplement the quantitative scale item ratings in understanding student perceptions of the course. We present the details of the method used, a summary of the results, and a discussion of the findings. The text analytics method presented may be a useful tool for others undertaking similar analysis of the open-ended course evaluation comments provided by students.

6505ENG Manufacturing with Composites

6505ENG aims to provide students with a theoretical and practical understanding of fibre reinforced composite materials, and to introduce the fundamentals of composite design and manufacture. A design module focuses on the relationship between mechanical properties and fibre length, orientation and volume fraction. Short fibre and long fibre composites and their laminates are considered. A manufacturing module covers processes including basic

wet layup, spray up, vacuum bagging, resin transfer moulding, vacuum assisted resin transfer moulding, and compression moulding. An introduction to filament winding and pultrusion is also provided. The laboratory practical tasks practiced by the students are carefully selected and managed. There is also a composite design and manufacture project to underpin the theoretical concepts presented in the classroom within a project-based learning mode. The assessment for 6505ENG consists of four elements, the weighting of which were rebalanced after the initial course offering based on consideration of the relative effort required by students for successful completion. The four elements of assessment, their initial 2014 weightings, and their subsequent 2016-2018 weightings were:

- In-class test – 15 per cent initially, 12 per cent subsequently;
- Four laboratory practicals – ten per cent initially, eight per cent subsequently;
- Design project – 25 per cent initially, 30 per cent subsequently; and,
- Final examination – 50 per cent, unchanged.

The course content, format and assessment changed in only minor ways across all four offerings considered here. In contrast, the nature of the design project activity completed by students changed in more significant ways. In 2014, the design project was theoretical only, with students asked to design a composite bike seat post, and to describe how it could be manufactured. In 2016, a manufacturing aspect was introduced, with students asked to complete a creative design exercise with minimal constraints for a composite wine bottle holder, which they subsequently manufactured. In 2017, the design-and-manufacture project format was used again and supplemented with a product testing phase. Students were asked to complete an engineering design for a bike seat post, to manufacture the item and to then perform a static test on the item to assess its performance against the international standard ISO 4210 - Cycles -- Safety requirements for bicycles. In 2018, the engineering design, manufacture and test project format was retained, but the design artefact was changed to a set of bicycle handlebars.

Methodology

The Griffith University Ethics Committee was consulted to confirm that research using the student evaluation data for the course 6505ENG was exempt from ethics approval. The Griffith University SEC instrument contains six scale items and two open-ended response items for students' feedback. Each scale item is presented as a statement, to which students indicate their level agreement on a five-point scale of the form: Strongly agree (5); Agree (4); Neutral (3); Disagree (2); and Strongly disagree (1). The two open-ended response items are presented as a question to which students can provide a text response. The SEC items are:

- SEC1 - This course was well-organised.
- SEC2 - The assessment was clear and fair.
- SEC3 - I received helpful feedback on my assessment work.
- SEC4 - This course engaged me in learning.
- SEC5 - The teaching (lecturers, tutors, online etc) on this course was effective in helping me to learn.
- SEC6 - Overall I am satisfied with the quality of this course.
- SEC7 - What did you find particularly good about this course?
- SEC8 - How could this course be improved?

Following the completion of the SEC survey period, a 'SEC Detail Report' which tabulates the student response data is provided to the course convenor. These Detail Reports contain no information capable of identifying any student respondent, and include the following data:

- the distribution of individual item rating scores;
- the mean item ratings;
- the standard deviation of the mean item ratings;
- the median item ratings;

- a set of benchmark comparison mean item ratings based on the 25 per cent, 50 per cent and 75 per cent quartile mean item ratings for the group of comparable courses (from the same Faculty group and of a similar sized enrolment); and
- a tabulation of all the student comments received for the open-ended response items.

For each of the six scale items, a one-way analysis of variance (ANOVA) was performed to identify any significant differences in the mean student ratings over the first four course offerings. Levene's test for homogeneity of variance was performed to identify the most appropriate ANOVA test.

In the previous evaluation addressing only the initial 2104 course offering (Hall & Palmer, 2015), a word cloud visualisation of the text in open-ended comments was produced. This method was useful to show the key themes reported by students and their relative frequency, but did not show relationships between the individual text terms, and could not effectively visualise inter-year relationships between terms in the multi-year comment data set here. Here we applied a more sophisticated text analytics visualisation method. Text analytics approaches offer several methods to analyse and visualise the text data. In the work presented here, we used the text analytics software package KH Coder (Higuchi, 2017) to analyse the content of the student comments. KH Coder was selected as it is free and provides a range of analysis and visualisation options. The student open-ended comments were tagged with the year of offer that they related to, and then pooled to form the input, on a common basis, to a multidimensional scaling (MDS) analysis (Palmer, 2019). The purpose of the MDS analysis was to visualise and identify any clustering of comment themes by year that might reveal differences in the students' perceptions of the first four offerings of the course.

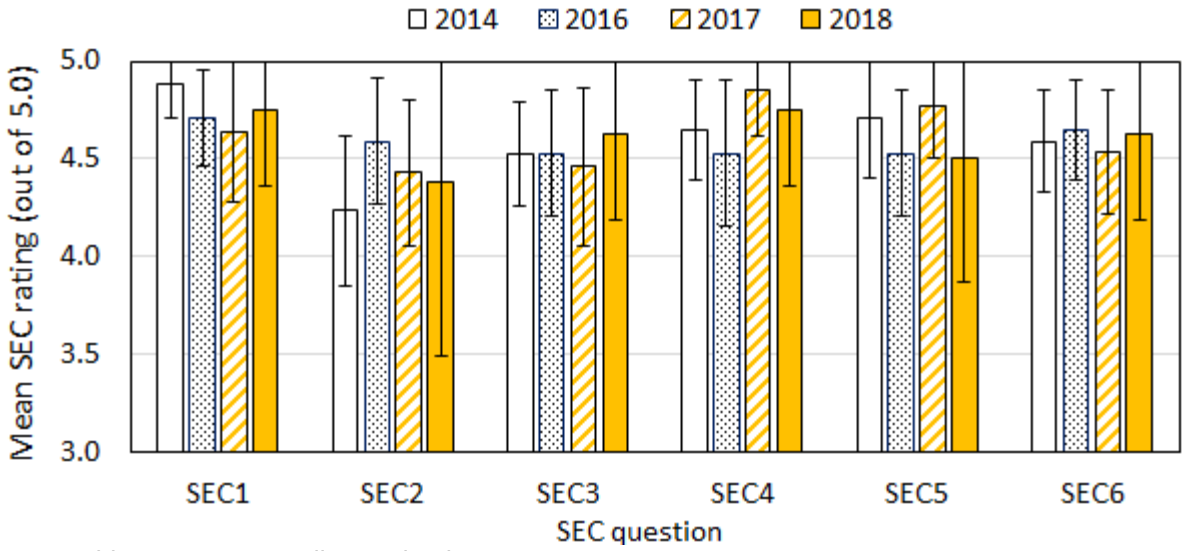
MDS computes a statistical measure of 'distance' between all pairs of unique text terms based on the frequency with which they co-occur in the student comments relative to the total of number of occurrences of both terms combined. This results in a matrix of distances with dimensionality equal to the number of unique text terms. A dimensional reduction technique analogous to principal component analysis is used to re-map the inter-term distances onto a new set of dimensions where the majority of the distance variation is concentrated in the first few dimensions (principal components), such that the overall loss of distance information is minimised. The MDS visualisations used here map the terms onto the first two principal component dimensions, providing a two-dimensional visualisation.

Results

Table 1 presents the response rates for the components of the SEC survey over the four course offerings, as well as the overall pooled data set. Figure 1 presents mean ratings and 95 per cent confidence interval estimates for the SEC scale items for the first four course offerings included in this work. Note that the vertical axis is compressed. Figure 2 presents the resultant KH Coder MDS plot of the text of the student comments received for SEC7 - What did you find particularly good about this course? Figure 3 presents the MDS plot of the text of the student comments received for SEC8 - How could this course be improved?

Table 1: Response rates for components of the SEC survey

Year	Enrolment	Scale items	SEC7	SEC8
2014	23	73.9 %	56.5 %	65.2 %
2016	24	70.8 %	58.3 %	58.3 %
2017	23	60.9 %	56.5 %	43.5 %
2018	27	29.6 %	22.2 %	18.5 %
Pooled	97	57.7 %	47.4 %	45.5 %
Total words	-	-	942	1252



SEC1 - This course was well-organised.
SEC2 - The assessment was clear and fair.
SEC3 - I received helpful feedback on my assessment work.
SEC4 - This course engaged me in learning.
SEC5 - The teaching (lecturers, tutors, online etc) on this course was effective in helping me to learn.
SEC6 - Overall I am satisfied with the quality of this course.

Figure 1: Mean ratings and 95 per cent confidence interval estimates for SEC scale items

Discussion

There are some limitations to the research presented here. The SEC ratings were derived from response scales that are ordinal in nature, and the use of ordinal data in many parametric statistical procedures is not universally accepted as valid. However, there is a significant body of research that has demonstrated the practical utility of analysis of ordinal data, based on the robustness of many parametric statistical methods (de Winter & Dodou, 2010; Norman, 2010). In each case presented in Figure 1, the confidence intervals shown take into account the respondent sample size and the variance in the mean rating, but in all cases the distribution of ratings departed from normality with negative skewness. However, all the distributions were similar and the basic statistical test from which confidence intervals are derived has a degree of robustness to departures from normality (Hubbard, 1978; Lumley, Diehr, Emerson, & Chen, 2002). Table 1 shows that the data set of student open-ended comments was relatively small (6505ENG is a final-year elective), however the individual year and overall pooled response rates were generally good, and the MDS analysis method employed here should still produce useful results. The MDS analysis method is algorithmic, and applies quantitative methods to data that are essentially qualitative, so the results require interpretation. For example, the MDS visualisations here include an indication of clustering of terms using colouring - this clustering is based on the adjacency of terms when mapped to the two-dimensional plot space, and is indicative only. With these limitations in mind, we can consider the results in more detail.

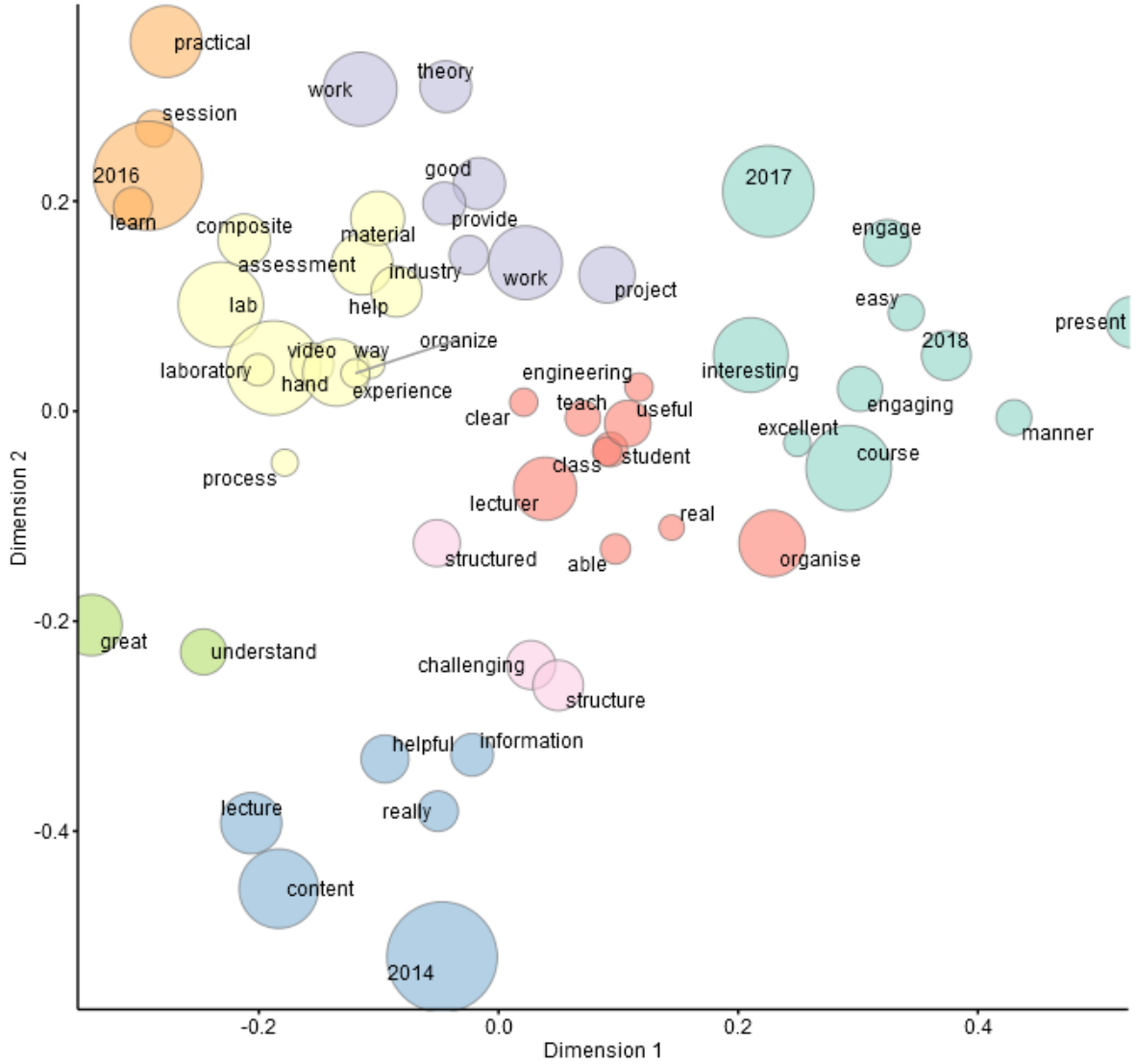


Figure 2: Multidimensional scaling plot of 'particularly good' student comments (SEC7)

It is apparent from Figure 1 that the mean ratings for all six SEC scale items were similar in all years of offer. The ANOVA tests revealed no significant differences in the mean ratings for all SEC scale items over the first four course offerings, and post hoc pairwise comparisons confirmed no significant differences in the mean ratings for all pairs of years. The mean ratings for all SEC scale items in all years were comparatively high. As noted above, the SEC scale items are reported to course convenors with quartile benchmark mean ratings for similar courses. For the six SEC items across the four course offerings considered, in 23 of the 24 cases presented in Figure 1, the mean rating was above the relevant 75 per cent quartile benchmark, with the remaining case just under the 75 per cent quartile benchmark. In this situation, any positive changes in SET ratings are likely to be masked by the 'ceiling effect' (Keeley et al., 2013). That is, on a five-point rating scale, "there is little room to report a wide range of improvement for the already average or above average instructor" (Menges & Brinko, 1986, p. 9). An indicator of the presence of the ceiling effect in SET ratings is a negatively skewed ratings distribution (Hake, 1998), and this was apparent in all 24 cases presented here. The results presented in Figure 1 provide reassurance in as much that the SET ratings were high for all scale items in all years, but they offer limited practical information as to whether students perceived the deliberate changes that were made to the course learning design over the first four course offerings.

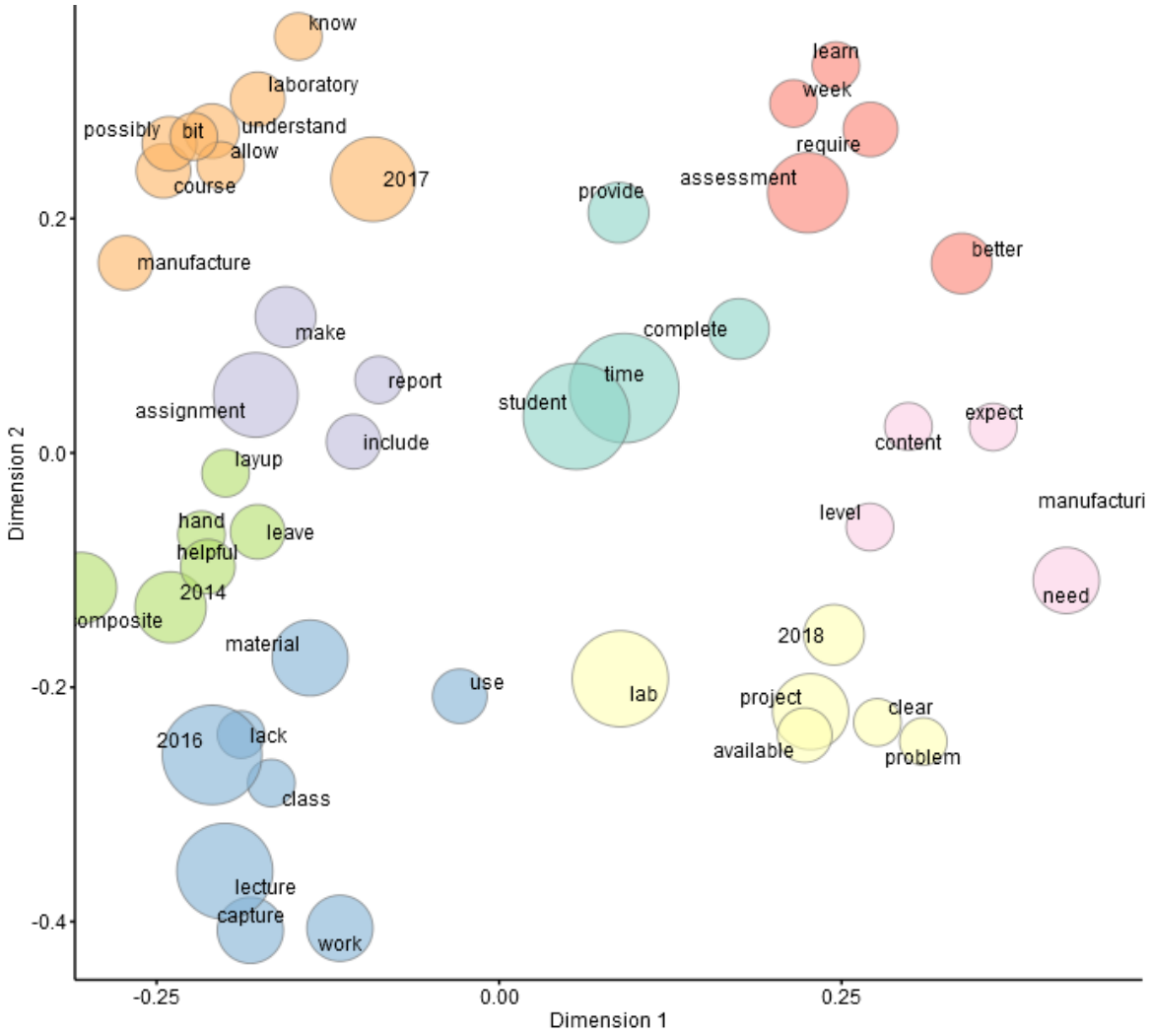


Figure 3: Multidimensional scaling plot of 'be improved' student comments (SEC8)

The MDS plot text visualisations presented in Figure 2 and Figure 3 revealed clusters of terms in the student open-ended comments related to the year of offer. Within each figure, the sizes of the 'year' bubbles provide a relative indication of the total number of student comments received in that year, as each comment was tagged with its corresponding year. Similarly, within each figure, the size of each term bubble provides an indication of the relative frequency of occurrence of each term in each comment data set – 'particularly good' and 'be improved'. The sizes of the term bubbles are not comparable between Figure 2 and Figure 3, as the total number of comments and terms/words in each data set was different. Depending on the nature of the source text, it may be possible to attribute an ordinal or other meaning to the principal component dimensions of the resultant MDS visualisation (Namey, Guest, Thairu, & Johnson, 2007). However, here we are primarily interested in relative positioning of the individual terms and term clusters.

In Figure 2 (the SEC7 'particularly good' comments), the following observations were made. At the lower range of dimension 2 there was a cluster of terms from 2014 relating to 'lecture content' and 'really helpful information', suggesting that students perceived the course in 2014 as being more theoretical in nature relative to other years. We note that the 2014 design project had no manufacturing component. The term clusters for 2016, 2017 and 2018 (when the design project had a practical component) were all positioned toward the other end of dimension 2, suggesting significant dissimilarity (on this dimension) in the terminology used by students in their 'particularly good' comments. Terms associated with 2016 include 'practical', 'hand[s]' and 'lab'/'laboratory', suggesting that students had perceived the introduction of the practical manufacturing element into the design project in 2016.

While the term cluster for 2014 was separate from other years on dimension 2, there was also a separation between the term cluster for 2016, which was at the lower range of dimension 1, compared to the term clusters for 2017 and 2018, which were both close together at the upper range of dimension 1. While the design project in these three years included a manufacturing element, the nature of the project evolved from a free-from/creative design-and-build in 2016, to a more tightly specified engineering design-build-and-test of a bike component in 2017 and 2018. The relative position of the term clusters for these three later years suggests that the students perceived some aspects of the course differently in 2017 and 2018, compared to 2016. The terms most closely associated with 2017 and 2018 included 'engaging', 'easy', 'manner' and 'present', perhaps representing a consolidation of the course teaching style, by the third offering of a previously new course, being perceived by students.

It is a feature of the MDS algorithm that a term associated with multiple other terms will tend to be positioned between them (shared) in the resultant visualisation. In the case of Figure 2, this would mean that terms appearing in the centre of the plot represented general student perceptions of the course across all years of offer considered in the analysis. In the central region of Figure 2 terms such as 'clear', 'useful', 'structured' and 'organise[d]' were found, suggesting that students generally perceived the course as well-structured and organised. This result is in accordance with the quantitative SEC results, where SEC1 ("This course was well-organised") received the highest mean-of-means SEC rating for all years considered.

The intention of this research was to look for evidence in the open-ended comments from students that they had perceived deliberate changes made in the course learning design to enhance the course, so the focus was principally on whether they noted as 'particularly good' anything related to the course changes made. However, it was considered useful to also examine Figure 3 for any prominent features related to specific years in the visualisation of the SEC8 'be improved' comments. In Figure 3, associated with 2016, terms related to 'lecture capture' were observed. At Griffith University, lecture capture was introduced in 2013, and by 2014 was available in many larger teaching spaces. However, 6505ENG was a final-year elective course with a relatively small enrolment, and was timetabled in a smaller teaching space without lecture capture in 2014 and 2016. By 2016 lecture capture was more widespread across the university and was generally expected by students, and its absence in 2016 appears to have been noted by students in the open-ended comments. In 2017 and 2018 lecture capture was available for 6505ENG classes. In the central region of Figure 3, the terms 'student', 'time' and 'complete' were observed. It has been reported that engineering students often have comparatively high study loads (Kyndt, Berghmans, Dochy, & Bulckens, 2014), and that some engineering students who aspire to high marks express a desire for additional time to improve their academic results (Guillaume & Khachikian, 2011). Hence, students expressing a desire for additional time to complete their engineering study is perhaps not unexpected here.

Conclusion

The text analytics analysis revealed clusters of terms in the student open-ended comments by year. The term clusters that were observed did appear to capture aspects of the intentional changes to the learning design over the first four offerings of the course, providing some evidence that students did perceive these intentional course learning design changes. The text analytics method did reveal additional useful course evaluation information present in the open-ended comments provided by students. This qualitative data source, which many evaluation instruments routinely collect, offered a valuable complement to the quantitative scale item data which are traditionally the focus in institutional student evaluation of teaching processes. We have documented here a method for the analysis of the open-ended course evaluation comments provided by students that is likely to be valuable to others involved in engineering education and evaluation for course enhancement.

References

Boyles, C. T., Moynihan, A. R., Mulroy, J. D., Fitch, B., Watts, P. J., & Evans, J. L. (2011). Undergraduate student experiences with composite materials. *International Journal of Mechanical Engineering Education, 39*(4), 297-305. doi:10.7227/IJMEE.39.4.3

de Winter, J. C., & Dodou, D. (2010). Five-point Likert items: t test versus Mann-Whitney-Wilcoxon. *Practical Assessment, Research & Evaluation, 15*(11), 1-12.

Guillaume, D. W., & Khachikian, C. S. (2011). The effect of time-on-task on student grades and grade expectations. *Assessment & Evaluation in Higher Education, 36*(3), 251-261. doi:10.1080/02602930903311708

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*(1), 64-74. doi:10.1119/1.18809

Hall, W., & Palmer, S. (2015). Student opportunities in materials design and manufacture: Introducing a new manufacturing with composites course. *Journal of materials education, 37*(3-4), 155-168.

Hayhurst, D. R., Kedward, K. T., Soh, H. T., & Turner, K. L. (2011). Innovation-led multi-disciplinary undergraduate design teaching. *Journal of Engineering Design, 23*(3), 159-184. doi:10.1080/09544828.2010.544248

Higuchi, K. (2017). KH Coder (Version 3.Alpha.09b). Japan: Koichi Higuchi. Retrieved from http://khc.sourceforge.net/en/

Hubbard, R. (1978). The Probable Consequences of Violating the Normality Assumption in Parametric Statistical Analysis. *Area, 10*(5), 393-398.

Huybers, T. (2014). Student evaluation of teaching: the use of best–worst scaling. *Assessment & Evaluation in Higher Education, 39*(4), 496-513. doi:10.1080/02602938.2013.851782

Keeley, J. W., English, T., Irons, J., & Henslee, A. M. (2013). Investigating Halo and Ceiling Effects in Student Evaluations of Instruction. *Educational and Psychological Measurement, 73*(3), 440-457. doi:10.1177/0013164412475300

Kyndt, E., Berghmans, I., Dochy, F., & Bulckens, L. (2014). 'Time is not enough.' Workload in higher education: a student perspective. *Higher Education Research & Development, 33*(4), 684-698. doi:10.1080/07294360.2013.863839

Lumley, T., Diehr, P., Emerson, S., & Chen, L. (2002). The Importance of the Normality Assumption in Large Public Health Data Sets. *Annual Review of Public Health, 23*(1), 151-169. doi:10.1146/annurev.publhealth.23.100901.140546

Menges, R. J., & Brinko, K. T. (1986, 16-20 April). *Effects of Student Evaluation Feedback: A Meta-Analysis of Higher Education Research.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Namey, E., Guest, G., Thairu, L., & Johnson, L. (2007). Data reduction techniques for large qualitative data sets. In G. Guest & K. M. MacQueen (Eds.), *Handbook for team-based qualitative research* (pp. 137-162). Plymouth, UK: Altamira Press.

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education, 15*(5), 625-632. doi:10.1007/s10459-010-9222-y

Palmer, S. (2019). Visualising the Social Media Conversations of a National Information Technology Professional Association. *International Journal of Human Capital and Information Technology Professionals, 10*(1), 38-54. doi:10.4018/IJHCITP.2019010103

Copyright statement