Automatic Text Analysis of Reflective Essays to Quantify the Impact of the Modification of a Mechanical Engineering Course

Aneet Narendranath and Jeffrey S. Allen
Michigan Technological University

1. Abstract

Students' reflective essays in engineering education provide insight and context for instructional modification and assessment. However, the assessment of reflective essays numbering in thousands can be time-consuming. This is notably important when trying to find specific changes in focus from one essay to another and measuring how strong those changes are across multiple corpora of essays.

In this paper we describe and demonstrate an automated text analysis method for the at-scale, corpus-normalized analysis of reflective essays. We apply it to quantitatively measure whether the modification of an undergraduate mechanical engineering course had the conjectured impact of a stronger emphasis on teamwork.

Our analytical method is a "pipeline" composed of Text Mining (TM), Natural Language Processing (NLP), and Recurrence Quantification Analysis (RQA). We use this method to measure the presence of a specific thematic element in reflective essays to confirm the impact of the modification of a team-driven, model-based engineering design course. The original course and its modification were visualized using Sandoval's conjecture mapping framework.

The novel innovation of this approach is that the input (text from hundreds of reflective essays, sourced one at a time) when passed through this pipeline quickly produces a quantitative indication of the presence of thematic elements and their recurrence normalized across a corpus of hundreds of essays. A comparison of this quantitative indicator across separate corpora (each corpus of essays is for a different year) of reflective essays signaled a change in student focus toward the conjectured outcome.

We conclude that the TM-NLP-RQA pipeline can be applied for quick and at-scale extraction of the relative magnitude of thematic statements from reflective essays. We observe that our conjectured redesign had the impact that we desired.

2. Background

In the Fall semester of 2014, the Department of Mechanical Engineering began a two-year roll out of four practice-based courses designed to prepare students for the increasing complexity of engineering systems and blurring of traditional disciplinary boundaries. These courses were developed, in part, on recommendations and findings from national studies of engineering education that included *The Engineer of 2020* prepared by the National Academy of Engineering (2004), and *ASME Vision 2030* by Kirkpatrick, Danielson, and Perry (2012). A common theme through these and other reports is that engineering students must be prepared for interdisciplinary systems-level analysis and design. An additional recommendation in these reports was that complex system-level perspectives, interdisciplinary team work, innovation, project management, and technical communications ought to approach the priority given to traditional technical topics in engineering science. To this end, the department faculty selected the following discipline threads to weave through these four practice-based courses.

1. application of thermal-fluids
2. application of design and manufacturing
3. application of solid mechanics
4. application of dynamic systems
5. programming, modeling, and simulation,
6. instrumentation, measurement, data acquisition, system control
7. application of structured design process
8. making and tinkering,
9. technical communication, and
10. interdisciplinary team work.

This article discusses a quantitative method to measure the impact of a course modification to "Mechanical Engineering Practice 3 (MEP-3)" which is the third course in this sequence. This is a two semester-credit course focused on team-driven, model-based design using advanced engineering software including AMESim™, MotionView™, and HyperMesh™. Teams of students simulate a complex payload delivery system to safely move people and goods from point-to-point along a specified route. In a team setting, students use a sequence of computer models and engineering calculations to guide design decisions. A significant effort is required by students to learn and effectively use these advanced engineering tools while developing team-level competence for complex engineering design. There are approximately 350 students enrolled in this course each academic year.

One of the course instructors (primary author) observed that students focused on the theme of their individual work and mastery of software with little attention to the theme of engagement with their team, team-created work products, and team-driven design decisions. This observation was based, in part, on a selective review of reflective essays included in a portfolio that each student submits at the end of the semester. A course modification (change in student assessment/grading) was implemented in an attempt to shift the focus of students towards "working in teams". This effort begged the following questions:

1. Was the course modification effective towards shifting student focus?
2. Would a shift in student focus be evident in the reflective essays, included in the student portfolios?
3. Could a quantitative assessment process pipeline be used to detect any such shift in student focus?

In the following sections, we first contextualize the challenge and our approach in prior re- search on two topics: the utility of reflective essays and the existence of an identity issue ("team" vs. "self") in collaborative engineering courses. Next, we articulate our conjectured course modification in the Sandoval's conjecture mapping framework (Sandoval, 2014). Finally, we describe and demonstrate our analytical technique's ability to quantitatively measure the desired impact of a course modification.

3. Challenges In the Current Reflective Essay Assessment Technique

We require students to submit a reflective essay of approximately 600 words as part of an end- of-semester portfolio. Reflective essays have been well regarded as an important element in student learning and their professional development (Boud, Keogh, & Walker, 2013; Boyer, Maher, & Kirkman, 2006; Campbell & Schmidt, 2005; Collins, Brown, & Newman, 2018; Driscoll & Wood, 2020; Estrem, 2015; Hatton & Smith, 1995; Rodgers, 2002; Ryan, 2011; Sattler, Kilgore, & Turns, 2010; Scott, Inoue, Adler-Kassner, & Wardle, 2015; Scouller, 1998; Thompson, Sattler, & Turns, 2011). The focal point of the cited papers is that reflective essays (i) can allow students to metacognitively probe their disciplinary identity and any tensions within (ii) serve as an indicator to support outcome-based course assessment, course design, and continuous improvement practices.

Each academic year, students in MEP-3 produce a corpus of approximately 350 reflective essays. This amounts to thousands of lines of text and hundreds of thousands of words. These words are domain-specific or are metadiscursive and communicate authorial stance. These essays are used to understand student perceptions of the course and drive decisions on course and curricular improvement. These essays are evaluated for completion, and full marks are awarded for any attempt to answer the following four questions.

1. *Explain what two assignments you selected and why you chose to include them in your port- folio. Consider what you learned from these assignments, what steps you took to retain that learning for future application, and how you incorporated your instructor/GTA's (Graduate Teaching Assistant's) feedback to improve the current work.*
2. *Which lesson or assignment in this course has been the hardest for you so far? What steps did you take to help you master the material presented? What courses outside this course helped you understand the concepts in this class?*
3. *How has your ability to communicate your ideas evolved since you began your career at this university? What specific activities or assignments have helped you develop your communication skills? What has not worked well? What would help you continue to improve your speaking and writing skills in the last few semesters before you graduate?*
4. *At this stage in your academic and professional career you have had the opportunity to work with many different people on group projects with varying levels of success. What has been the biggest challenge for you when you work in teams? What role do you typically perform in a group setting, e.g. project leader, documentation leader, person who takes on tasks no one else wants, or other? What advice would you give to an incoming student about working in groups?*

Instructor assessment of the content and themes in these reflective essays is time consuming even with well-structured rubrics and qualitative labels applied only to a subset of portfolios. In addition to time consumption, an instructor assessment will not reveal corpus-normalized subtleties such as the relative strength of a word or a phrase judged not only by its presence in a document within a corpus of documents, but its absence in the rest of a corpus.

In addition to the time-consuming nature of essay assessment, our current practice of collecting essays once a semester may not have sufficient granularity to probe and measure the emergence of content patterns, metacognition, or affect. We plan to collect essays more frequently and use quantitative assessment process to analyze them. This will help us under- stand emergence of content patterns, metacognition, and associated affect in collaborative courses. This would be tantamount to intensive longitudinal assessment that would consist of thousands of reflective essays per year requiring specialized and specific computational tools for analysis and extraction of insight in a short span of time. Therefore, it is necessary that we create an quantitative assessment process that may be applied for assessment and insight extraction at scale.

From a broader perspective, STEM education is embracing learning analytics (Clow, 2012; Siemens, 2013) as a means to creating personalized learning environments and to gain in-sight toward course design. "Learning analytics is the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs." Our automated analytical method is a natural step in this direction.

Our automated text analysis method can also complement future longitudinal studies that probe the emergence or damping of metadiscursive styles. Since the submission of this manuscript, a separate paper (Narendranath, 2023) on a framework to perform intensive longitudinal psychometrics through a metadiscursive analysis has been peer-reviewed and accepted at an education research methods venue. This metadiscursive analysis applied to students' discourse work products is not part of the current paper. However, in the future, this new metadiscursive analysis framework will be applied in conjunction with this manuscript's thematic quantification framework to probe emergence or damping of discursive styles and associated authorial stance. This framework could be relevant to measure a change in discursive styles before and after the COVID-19 pandemic or to add additional credence to changes in discursive styles post a course refinement. The psychometric indicators will include but not be limited to analytical capacity as measured through interactive metadiscourse, emotional tone as measured through sentiment valence, and psychological indications inferred from the distribution

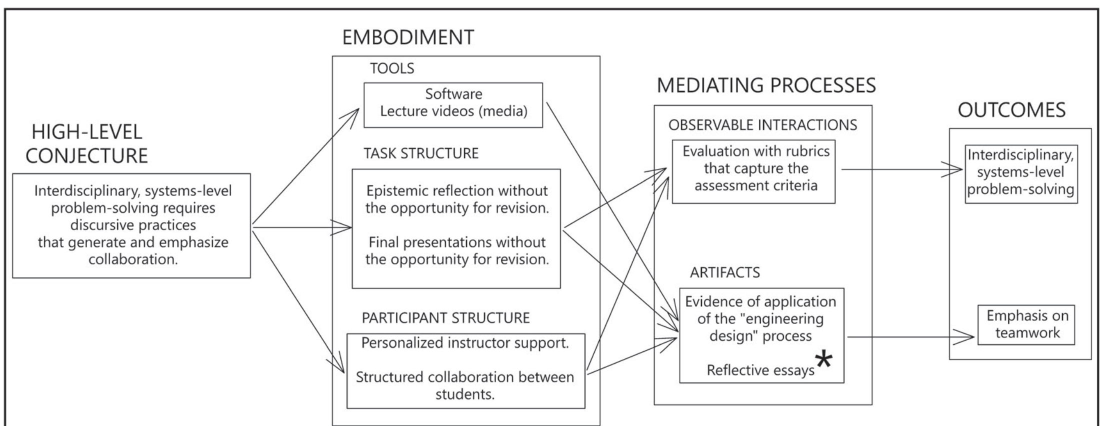

Figure 1. Original conjecture map. The asterisk denotes augmentation of the assessment of reflective essays with automated text analysis.

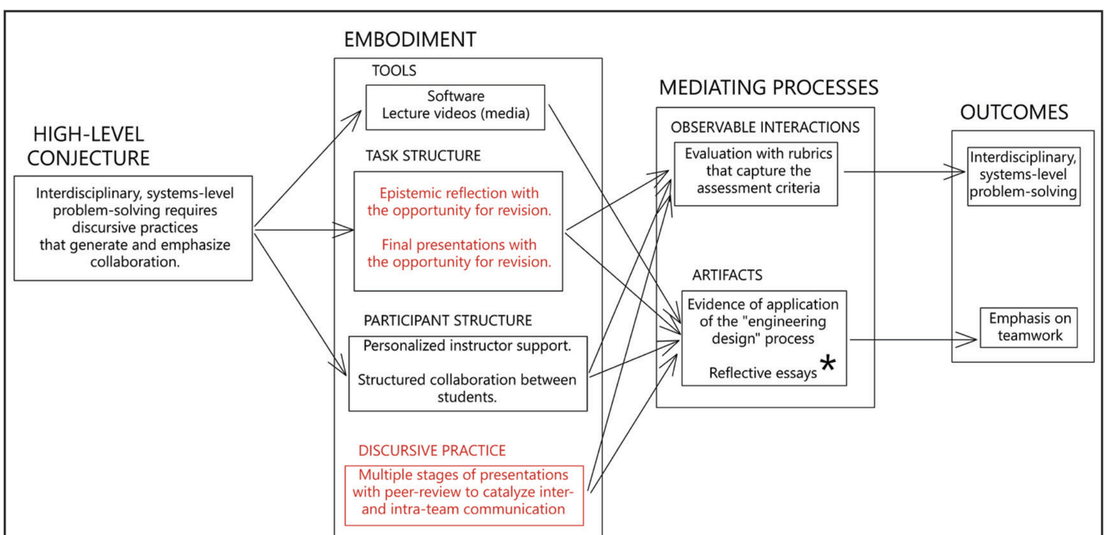

Figure 2. Conjecture map for the modified course. The modifications are highlighted in red. The asterisk denotes augmentation of the assessment of reflective essays with automated text analysis.

of pronoun usage as an indicator of shifts in identity.

Literature review reveals no freely available quantitative tools that measure patterns in dis- cursive practices to validate course design. We constructed this quantitative assessment pipeline to measure the impact of a conjectured course modification towards a desired out- come. This pipeline can be further augmented with metadiscursive analysis and the application of large language models that power artificial intelligence. We plan to publish our method as an open-source, web browser-based tool for instructors and curriculum design researchers.

4. Modification Of Course Assessment Structure Situated In The Conjecture Mapping Framework

Prior to Fall 2021, instructor review of the reflective essays in MEP-3 portfolios indicated that students were focused more on the theme of "individual" effort and independent mastery of software tools rather than the theme of "working in teams". In fact, there is evidence in literature (Trevelyan, 2011) that students develop a misconception that engineering practice is purely technical in nature (instead of it being *sociotechnical*), thus precipitating an identity issue when working in teams. This is concerning because such a misconception can find the students at odds with the reality of working on professional engineering projects, where teamwork is an important facet. Faulkner (2007) mentions that "engineering is about 'nuts and bolts **... and people**' ". More evidence of students focusing on "individual" effort rather than on the main theme of "working in teams" in practice-based design course settings has been reported in literature (Blumenfeld et al., 1991; Newstetter & Kolodner, 1995; Ng & Bereiter, 1991; Stone, 1996), and summarized by Turns, Newstetter, Allen, and Mistree (1997) as precipitated due to:

1. Students being under pressure to complete their work may focus on individual effort than focus on team-challenges.
2. A combination of individual and team-based activities and related assessment can lead to students focusing on one type of assessment activity over the other, i.e., focusing on those assessment activities that measure individual effort versus those that measure team-created work products.

In an attempt to shift the primary theme of student focus, the assessment/grading structure in MEP-3 was modified in the Spring 2021 semester. The changes are framed via the conjecture mapping technique from Sandoval (2014). The conjecture mapping framework is a transparent, system's approach to course design, redesign, modification, or refinement. It starts with an input node of a *high-level conjecture* that leads to an output node of a *desired outcome*. The *embodiment* of the conjecture (tasks performed by students, discursive activities that students

5. Reader Orientation And Glossary Of Terms And References To Orient A Reader

Term	Definition and Citations
Centroidal lexical unit	This is a phrase coined by the authors. This is the LU of interest. For example, for word-level analysis, the centroidal LU is "team", while for sentence-level analysis, the centroidal LU is "working in teams.
Corpus	A corpus is a collection of documents. Its plural form is "Corpora."
Cut-off radius, ϵ	This is the nearness threshold parameter for a RP. When two data points at temporal or spatial positions *i* and *k* are within ϵ of each other, the RP at the intersection of the *i*-th row and *j*-th column and the *j*-th row and the i-th column is populated with '1'.
Distance matrix	It is a square matrix that holds pair-wise distance (Euclidean or other) between elements in a lexical unit. For a word, a distance matrix is treated as an LU, while for sentence-level analysis; selected sentences are treated as LU. These LU are parsed through our quantitative pipeline to extract "features" through a process known as "embedding."
Document	It is an ordered collection of lexical units. In common usage, a document is a collection of sentences.
Embedding	Embedding is the process of representing words as numeric vectors. A vector is an ordered collection of numbers.
Euclidean distance	This is the distance between points in space. For text analysis, this is the distance between features or the distance between the centroidal lexical unit and other lexical units.
Feature	The numeric vector that is the outcome of embedding is called a feature.
Lexical unit (LU)	This is a phrase coined by the authors. This could be a character, a single word, part of a word, or a sentence that is the focus of analysis and comparison with a central theme (centroidal lexical unit). For a word-level analysis, each word in a reflective essay is treated as an LU, while for sentence-level analysis; selected sentences are treated as LU. These LU are parsed through our quantitative pipeline to extract "features" through a process known as "embedding."
Natural Language Processing (NLP)	This is the creation or study of computer programs that take lexical units as input and perform Text Mining operations that can enable text comprehension. NLP is widely applied (Bertoni, Fontana, Gabrielli, Signorelli, & Vespe, 2023, Clark, Fox, & Lappin, 2012; Mitkov, 2022) in political science, economics, sociology, psychology, etc.
Recurrence Plot (RP)	RP is an RQA tool. It is a two-dimensional plot to visually examine recurrence or repetition of patterns in higher-dimensional space (like text which is embedded in vector space). The RP is a matrix plot of a sparse array that is a square matrix of '1's and '0's. The '1's represent all the instances in time (or space) with a recurrence within a nearness threshold called cut-off radius.
Recurrence Quantification Analysis (RQA)	Recurrence Quantification Analysis (RQA) is a method for analyzing the sequential or repetitive structures in complex data including text (Allen, Likens, & McNamara, 2017; Angus, Smith, & Wiles, 2012; Danvers, Sbarra, & Mehl, 2020; Lyby et al., 2019; Orsucci, Walter, Giuliani, Webber Jr, & Zbilut, 1997; Wallot, 2017, (Eckmann, Kam- phorst, & Ruelle, 1987; Marwan, Romano, Thiel, & Kurths, 2007)
Recurrence rate	This is one of many RQA measures and is the percentage of recurrence points ('1's) in an RP. This is the only RQA measure that we use in this paper.
Text Mining (TM)	This is the process of normalizing text and extracting information or knowledge from it. NLP is a text mining process.
TF-IDF	This is an acronym for "Term Frequency times Inverse Document Frequency" score (Harman, 2005; Jones, 1972; Robertson, 2004; Salton, 1989). It uses a specific algorithm that quantitatively measures the importance of a term is in a specific document that belongs to a corpus of many documents. Unlike the raw term-frequency, the IDF multiplier of TF-IDF multiplicatively increases the score of a document, in case the term of interest is rare in the corpus.
Word stemming	This is the process of reducing a word to its form before suffixes, prefixes, or other modifiers are added. For example, the word stem of "teams" and "teaming" is "team." Stemming allows different words that share a stem to be grouped together and understood by a computer program as a single concept.

engage in) and the *mediating processes* (interventions and student artifacts) are the intermediate nodes through which the outcomes may be realized.

Although conjecture maps are used to explore or advance learning theories embedded in a course, we used it to decompose the initial conjecture or pedagogical intent into its embodiment of activities, resources, and support mechanisms. Instead of using a generic graphical representation such as a block diagram, we have chosen to use this learning design framework exemplar to communicate the pedagogical intent, its refinement, and flagging the student artifacts of interest. Conjecture maps have been used as a bridge between design-based research and analytical methods situated in the realm of learning analytics (Reimann et al., 2016). In addition, learning design has been used as a "form of documentation of pedagogical intent that can provide the context for making sense of diverse sets of analytic data" (Lockyer et al., 2013), and "design studies, particularly to the extent that they are hypothesis and framework generating may be viewed as contributing to model formulation" (this was in the context of creating generalizable models across learners or environments) (Kelly et al., 2004). Furthermore, the conjecture mapping framework will help us articulate and conduct research on our learning environment once we have established the analytical method in the current paper to measure subtle and nuanced changes in student artifacts driven by course redesign or refinement.

Our high-level conjecture is that "interdisciplinary systems-level problem-solving requires discursive practices that emphasize collaboration." The initial conjecture map and the conjecture map for the modified course are shown in Figures 1 and 2, respectively. We conjectured that a modification in the assessment structure of the course would create a previously non-existent discursive pattern that would lead to the intensification of teamwork. This intensification would be apparent in the students' end-of-semester reflective essays. The difference between the two iterations of the course is that the modified course includes unlimited revisions for the epistemic reflection tasks (individually completed homework assignments and mastery quizzes with unlimited attempts instead of timed examinations that allowed a single attempt) and multiple design reviews (team presentations) that are peer-reviewed to catalyze intra-team and inter-team discourse.

6. Analytical Approach And Its Application

We analyzed two corpora of reflective essays, one from Fall 2019 (original course) and the other from Fall 2021 (modified course). Our analytical technique had two separate analyses paths applied to each corpus. The first path is one that measures a numerical statistic called the TF-IDF for individual words as the fundamental lexical unit (Path A). The second path uses Recurrence Quantification to measure the recurrence of sentences close (Euclidean distance) to a theme (Path B). The two pathways are described in the flow diagram in figure 3. Path A is applied for a word-level analysis while Path-B is applied for a sentence-level analysis.

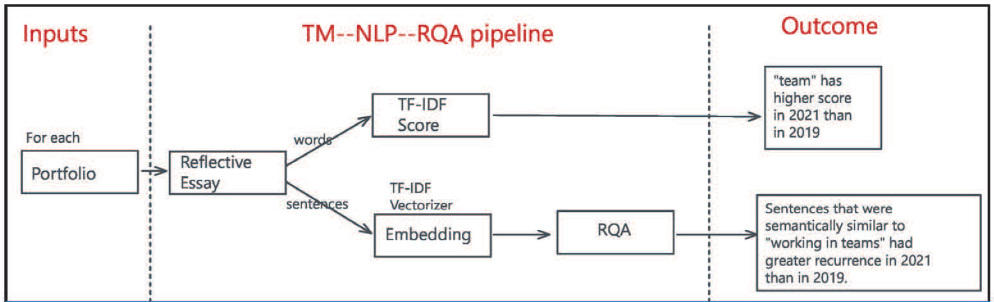

Figure 3. The process of transforming reflective essays for the purpose of TF-IDF calculation and recurrence quantification. Path (A) leads to the calculation of a TF-IDF score per document. Path (B) is for RQA.

In information retrieval, TF-IDF is a technique that measures the relative importance of a term in a document while not neglecting the rarity of terms. TF-IDF is an acronym for "term frequency 'times' inverse document frequency." It is called a "bag of words" technique since it does not account for the sequence or order in which words appear. It is a numerically computed measure that is used to quantify the importance of a term in a document while ensuring that highly frequent terms have their score damped. TF-IDF is calculated by multiplying the term frequency (TF) of a term in a document by the inverse document frequency (IDF) of the term. The term frequency is the number of times a term appears in a document. The inverse document frequency is a measure of how common a term is in a corpus of documents. A term that is common in a corpus will have a low IDF, while a term that is rare in a corpus will have a high IDF. In our corpora, an essay with a higher TF-IDF score for the lexical unit (LU) "team" has greater focus on team, teaming, teamwork, and group work than for an essay with a lower TF-IDF score for "team."

An illustrative example of computing the TF-IDF scores of four main characters in "Alice in Wonderland" by Lewis Carrol is provided in Table 1. The TF-IDF score for the following story-characters (multiple LUs) was evaluated: "alice", "queen", "hatter", "cat." It is clear from the TF-IDF scores that "alice" is the character who appears in all parts of the original document as suggested by non-zero TF-IDF scores. The other characters do not appear in some parts as indicated by a "0" TF-IDF score. It is also evident from the higher TF-IDF scores that "alice" is the primary character in all parts of the original document.

Recurrence quantification analysis (RQA) is a relatively recent technique for visual and numerical analysis of recurrence in time-series. Generally speaking, given one or multiple multidimensional numeric datasets, RQA can respectively visualize and quantify the pair-wise similarity between features within a dataset or similarity between pairs of datasets within a cut-off neighborhood radius (denoted as ϵ). Through RQA, an n-dimensional dataset is con- verted to a recurrence matrix, which is a two-dimensional matrix of '1's and '0's that serves as a visual indication of recurrence. Recurrence rate is one of the many measures that can be calculated from recurrence matrices. It is the percentage of '1's in the matrix. A higher recurrence rate denotes a high recurrence of a state within a cut-off neighborhood radius.

An illustrative example with the recurrence rate of words with pair-wise similarity at different ϵ values in "Alice in Wonderland" is included in Figure 4. Readers should note that our primary interest with the application of RQA is to measure the closeness of selected sentence LU to the centroidal LU of "working in teams."

6.1 Preprocessing the Corpora

Each semester of ME Practice 3 has an enrollment of 100-200 students who each submit portfolios that in-

Part	alice	queen	hatter	cat
1	0.122	0.0	0.0	0.009
2	0.116	0.0	0.0	0.009
3	0.139	0.009	0.0	0.0
4	0.200	0.059	0.034	0.029
5	0.165	0.072	0.021	0.010

Table 1. TF-IDF scores for four characters in "Alice in Wonderland" partitioned into five equal parts.

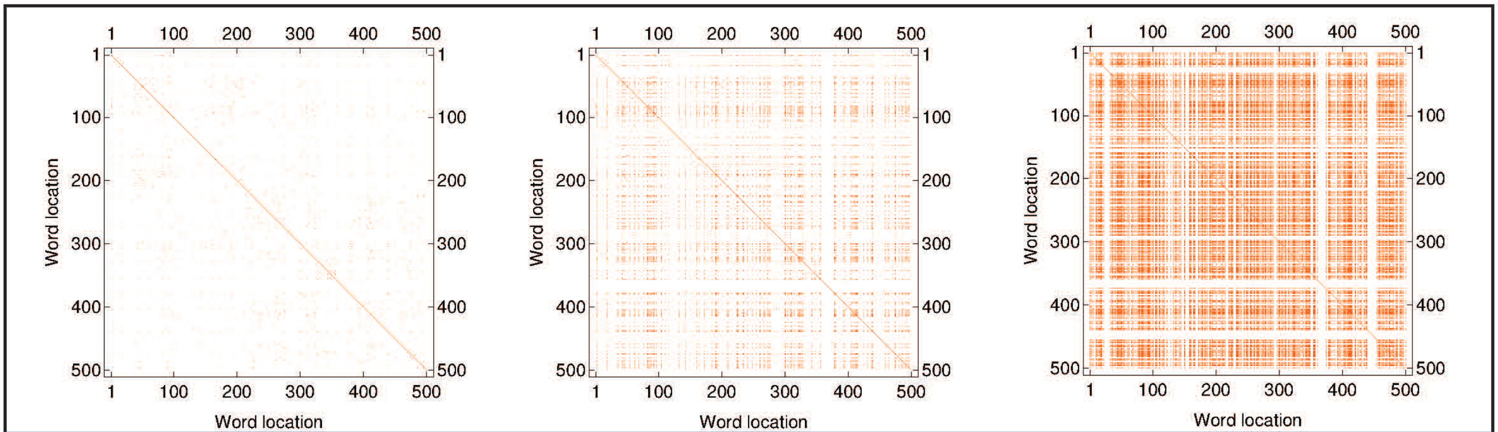

Figure 4. The recurrence plot of the vectorized (higher-dimensional representation) "Alice in Wonderland" document with individual words as lexical units. This is a two-dimensional representation of of pair-wise similarities between words. $\epsilon = 0.1$ (seeking pairs of words that are within 10% of each other) leads to a highly sparse recurrence plot that suggests that very few words are within 1% of each other in Euclidean space. $\epsilon = 0.3$ leads to a slightly more populated recurrence plot with a recurrence rate of 4%. $\epsilon = 0.6$ leads to a highly populated recurrence plot with a recurrence rate of 29%.

clude reflective essays. First, the reflective essays were extracted from each portfolio using a Linux script. Personally identifiable information (PII) were omitted as they only featured in the file name. This resulted in two corpora of PII-free reflective essays, one for Fall 2019 (the original course) and the other for Fall 2021 (the modified course). In addition, our university does not have an IRB requirement for the secondary data analysis for the purpose of quality improvement or quality assurance of the curriculum.

The Fall 2019 corpus of reflective essays held 4877 sentences with a total of 107,397 words while the Fall 2021 corpus of reflective essays held 7169 sentences with a total of 133,253 words. After the stop words are removed, these reflective essays have 4872 and 7161 words respectively for 2019 and 2021. Stop words are common words such as articles, conjunctions, prepositions, pronouns, punctuation marks, and symbols that do not carry content information. When performing content-specific thematic analysis, stop words lead to the reduction of the corpus-normalized strength of thematic lexical units. Although stop words were removed in the current study, they will be an important piece of follow-up metadiscursive and psychometric studies. In both corpora, references to "group", "group work", "team work" and their variants were are replaced with "team" to avoid omitting information due to synonymy. In addition, only word stems are retained.

6.2 Word-Level Analysis Through Tf-Idf

Reflective essays are decomposed into individual word LU and synonymy is managed. Stop- words are removed, word stemming is performed, and the TF-IDF scores for the LU "team" is calculated. The cumulative TF-IDF scores, depicted in Figure 5, are visual confirmation that reflective essays from 2021 have a greater focus on "team" than in 2019. In fact, there are some reflective essays from 2019 that do not feature a reference to "team."

6.3 Sentence-level analysis with recurrence quantification

The TF-IDF technique was applied to compute the relative strength of single words. However, we are also interested in computing the recurrence of short phrases focused on "working in teams." We applied a sentence embedding technique to individual sentences in each essays to achieve this latter objective. Sentence embedding was performed using the Wolfram Mathematica function *FeatureExtraction*. This is a machine learning function that can converts input data (which may be numeric, images, audio stream, or text) into a vector of numbers. It uses TF-IDF to decompose sentences into their component words, converts the TF-IDF scores to vectors, and reassembles the sentences as vectors of vectors. More details and examples of Wolfram Mathematica's *FeatureExtraction* function may be found via the world wide web at https://reference.wolfram.com/language/ref/FeatureExtraction.html.

Sentence embedding is a technique through which sentences are represented as a vector of numbers. In other words, words are *embedded* in their equivalent numeric space. The embedding process is also corpus-normalized. Sentence embedding makes it possible to compare sentences by measuring the distance (such as the Euclidean distance, which is the distance between two points in three-dimensional space) between their vector representations. The smaller the distance between vector representations of sentences, the more similar they are.

This technique of sentence embedding is used to quantify the effect of a change in assessment strategy measured through reflective essays submitted by students. After converting sentences in documents to numeric vectors (features), RQA is used to measure the effect of change in assessment structure through a density of LU that possesses spatial/Euclidean nearness to the centroidal LU, "working in teams."

The reflective essays have synonymy managed, are decomposed into sentence LU, and word stemming is performed. After this, only those sentences are extracted that possess a refer- ence to the stem "team." These selected

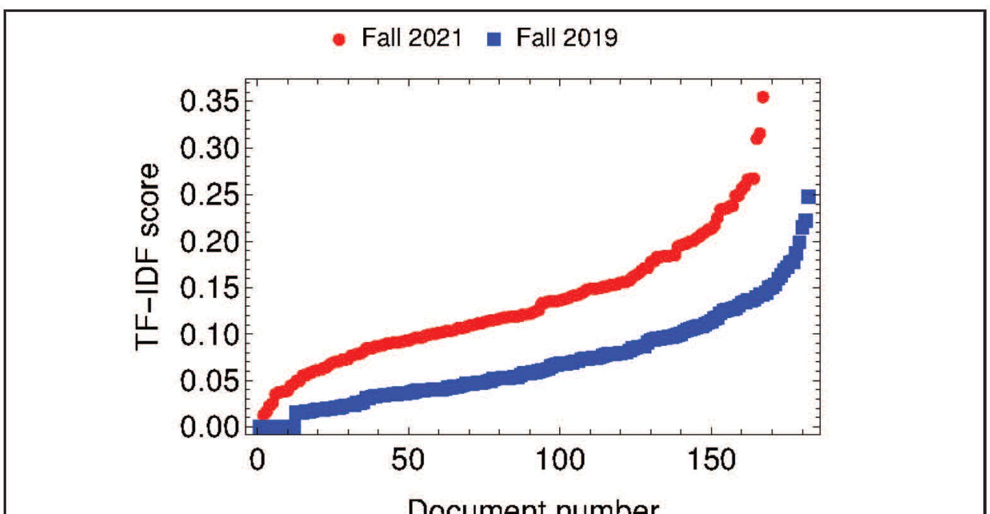

Figure 5. TF-IDF scores for the word "team" for the Fall 2019 and Fall 2021 reflective essays show greater importance given to teams, teamwork, groups, and group work in all essays from 2021. The scores are in ascending order of TF-IDF magnitude to allow for a visual comparison of the strength of the team-focused lexical units.

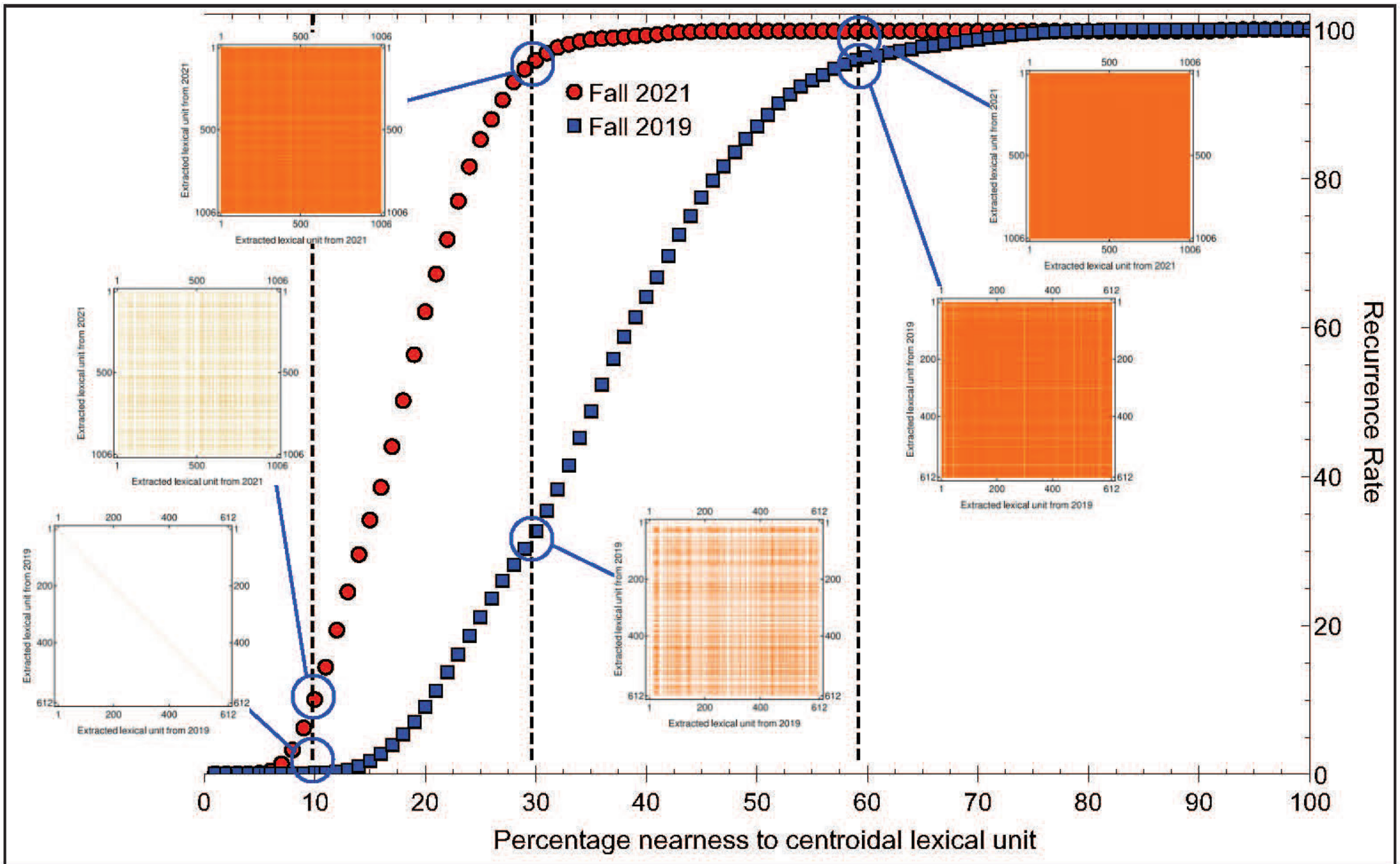

Figure 6. A greater density of comments (greater density of recurrence points) that are lexically similar to "working in teams" is observed in 2021 reflective essays as compared to 2019. As an illustration, only three nearness percentage values are depicted ($\epsilon = 0.1, 0.2, 0.3$.

sentence LU are embedded in numeric space the recurrence rate between them and the stemmed version of the centroidal LU "working in teams"' is computed at different nearness values of ϵ . This plot is shown in Figure 6. LU sorted by their distance from a centroid will be used in our future work to establish a threshold for a statistically relevant sample size from a large corpus of reflective essays or from multiple corpora across generations of students that could contain millions of sentences of text.

Three pairs of recurrence plots (RP) are shown in Figure 6 to visualize the pair-wise nearness of these extracted lexical units for $\epsilon = 0.1$ (10% nearness), $\epsilon = 0.3$ (30% nearness), and $\epsilon = 0.6$ (60% nearness). In each case, it is evident that reflective essays from 2021 yielded RPs with a greater density of lexical units that are closer to each other than those from 2019.

7. Conclusion

The assessment structure of a practice-based mechanical engineering course was modified. We conjectured that this modification would lead to a discursive process (which would be captured in end-of-semester reflective essays) among students that would intensify their focus on team work. We quantitatively assessed students' reflective essays before and after the course modification to detect a change in student focus. To perform this quantitative assessment, we created a pipeline of text mining (TM), natural language processing (NLP), and recurrence quantification analysis (RQA).

We performed two levels of analysis viz., at the word-level and the sentence-level analysis. The word-level analysis was conducted by computing the TF-IDF score of the lexical unit "team" in each reflective essay extracted from student portfolios from 2019 (original course) and 2021 (modified course). It was observed that the TF-IDF scores for the lexical unit "team" were higher and signaled a greater importance of this lexical unit in every instance for reflective essays from the modified course as compared to those from the original course.

The sentence-level analysis first required the representation of relevant sentence lexical units from 2019 and 2021 essays as higher-dimensional vectors. Next, these higher-dimensional vectors were passed through the numerical technique of recurrence quantification analysis to generate visual diagrams and to compute the quantitative measure of recurrence rate for varying nearness to centroidal LU. Recurrence plots and recurrence rates both indicated that reflective essays from the modified course had a higher incidence of statements that were lexically similar to "working in teams" than the essays from the original course.

At the beginning, we asked three questions:

1. Was the course modification effective towards shifting student focus?
2. Would a shift in student focus be evident in the reflective essays, included in the student portfolios?
3. Could a quantitative assessment pipeline be used to detect any such shift in student focus?

For the first question, we answer that a course modification led to a change in the shift in student focus. A deeper study can be applied where reflective essays or minute-papers could be used to track the evolution of this change on a weekly basis. For the second and third questions, we conclude that there is a quantifiable increase in student focus on 'team' and 'working in teams'.

The TM-NLP-RQA pipeline can be used to quickly capture the essence of and quantitatively contrast between hundreds of student submitted reflective essays in a matter of minutes. The process is faster than manually perusing essays and can be applied to multiple generations of student reflective essays to track students' academic evolution.

8. References

Allen, L. K., Likens, A. D., & McNamara, D. S. (2017). Recurrence quantification analysis: A technique for the dynamical analysis of student writing. *FLAIRS 2017 - Proceedings of the 30th International Florida Artificial Intelligence Research Society Conference, 240–245.*

Angus, D., Smith, A., & Wiles, J. (2012). Conceptual recurrence plots: Revealing patterns in human discourse. *IEEE Transactions on Visualization and Computer Graphics, 18* (6), 988–997. doi: 10.1109/TVCG.2011.100

Narendranath, A. D., Thelander, Z., & Hargrove-Leak, S. C. (2023, June). Measuring and Visualizing Metadiscursive Markers in Student Writing. In *2023 ASEE Annual Conference & Exposition.*

Bertoni, E., Fontana, M., Gabrielli, L., Signorelli, S., and Vespe, M. (2023). Handbook of Computational Social Science for Policy. Springer Nature.

Blumenfeld, P. C., Soloway, E., Marx, R. W., Krajcik, J. S., Guzdial, M., & Palincsar, A. (1991). Motivating project-based learning: Sustaining the doing, supporting the learning. *Educational Psychologist , 26*(3-4), 369–398.

Boud, D., Keogh, R., & Walker, D. (2013). Reflection: Turning Experience into Learning. Routledge.

Boyer, N. R., Maher, P. A., & Kirkman, S. (2006). Transformative learning in online settings: The use of self-direction, metacognition, and collaborative learning. *Journal of Transformative Education, 4*(4), 335–361.

Campbell, M. I., & Schmidt, K. J. (2005). Polaris: An undergraduate online portfolio system that encourages personal reflection and career planning. *International Journal of Engineering Education, 21*(5), 931.

Clark, A., Fox, C., & Lappin, S. (2012). The Handbook of Computational Linguistics and Natural Language Processing (Vol. 118). John Wiley & Sons.

Clow, D. (2012). The Learning Analytics cycle: Closing the Loop Effectively. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 134– 138).

Collins, A., Brown, J. S., & Newman, S. E. (2018). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. Knowing, Learning, and Instruction, 453–494. Routledge.

Danvers, A. F., Sbarra, D. A., & Mehl, M. R. (2020). Understanding personality through patterns of daily socializing: Applying recurrence quantification analysis to naturalistically observed intensive longitudinal social interaction data. *European Journal of Personality, 34*(5), 777–793.

Driscoll, A., & Wood, S. (2020). Developing outcomes-based assessment for learner-centered education: A faculty introduction. Stylus Publishing, LLC.

Eckmann, J., Kamphorst, O. S., & Ruelle, D. (1987). Recurrence plots of dynamical systems. *Europhysics Letters, 4*(9), 973.

Estrem, H. (2015). Disciplinary and professional identities are constructed through writing.

Faulkner, W. (2007). Nuts and bolts and peoples' gender-troubled engineering identities. *Social Studies of Science, 37*(3), 331–356.

Harman, D. (2005). The history of IDF and its influences on IR and other fields. *Charting a New Course: Natural Language Processing and Information Retrieval: Essays in Honour of Karen Spärck Jones*, 69-79.

Hatton, N., & Smith, D. (1995). Reflection in Teacher Education: Towards Definition and Implementation. *Teaching and Teacher Education, 11*(1), 33–49.

Sparck-Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation, 28*(5), 111-121.

Kelly, A. (2004). Design research in education: Yes, but is it methodological?. *The Journal Of The Learning Sciences, 13*(1), 115-128.

Kirkpatrick, A. T., Danielson, S., & Perry, T. (2012, June). ASME vision 2030's recommendations for mechanical engineering education. In *2012 ASEE Annual Conference & Exposition* (pp. 25-209).

Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing pedagogical action: Aligning learning analytics with learning design. *American Behavioral Scientist, 57*(10), 1439-1459.

Lyby, M. S., Mehlsen, M., Jensen, A. B., Bovbjerg, D. H., Philipsen, J. S., & Wallot, S. (2019). Use of recurrence quantification analysis to examine associations between changes in text structure across an expressive writing intervention and reductions in distress symptoms in women with breast cancer. *Frontiers in Applied Mathematics and Statistics, 5*, 37.

Marwan, N., Romano, M. C., Thiel, M., & Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics reports, 438*(5-6), 237-329.

Mitkov, R. (2022). *The Oxford Handbook of Computational Linguistics*. Oxford University Press.

National Academy of Engineering, (2004). The Engineer of 2020: Visions of Engineering in the New Century. National Academies Press Washington, DC.

Newstetter, W. C., & Kolodner, J. I. (1995, November). Learning to change the world: A case study of a mechanical engineering design course. In *Proceedings Frontiers in Education 1995 25th Annual Conference. Engineering Education for the 21st Century* (Vol. 2, pp. 4a3-10). IEEE.

Ng, E., & Bereiter, C. (1991). Three levels of goal orientation in learning. *Journal of the Learning Sciences, 1*(3-4), 243-271.

Orsucci, F., Walter, K., Giuliani, A., Webber Jr, C. L., & Zbilut, J. P. (1997). Orthographic structuring of human speech and texts: linguistic application of recurrence quantification analysis. *arXiv preprint cmp-lg/9712010.*

Reimann, P. (2016). Connecting learning analytics with learning research: The role of design-based research. *Learning: Research and Practice, 2*(2), 130-142.

Robertson, S. (2004). Understanding inverse document frequency: on theoretical arguments for IDF. *Journal Of Documentation.*

Rodgers, C. (2002). Defining reflection: Another look at John Dewey and reflective thinking. *Teachers College Record, 104*(4), 842-866.

Ryan, M. (2011). Improving reflective writing in higher education: A social semiotic perspective. *Teaching in Higher Education, 16*(1), 99–111.

Salton, G. (1989). Automatic text processing: The transformation, analysis, and retrieval of. *Reading: Addison-Wesley, 169.*

Sandoval, W. (2014). Conjecture mapping: An approach to systematic educational design research. *Journal of the Learning Sciences, 23*(1), 18–36.

Sattler, B., Kilgore, D., & Turns, J. (2010, October). "I have never spent time to think about what i have gained from my projects": Linking portfolio development and life-long learning. In *2010 IEEE Frontiers in Education Conference (FIE)* (pp. T3H-1). IEEE.

Scott, A., Inoue, A., Adler-Kassner, L., & Wardle, E. (2015). Assessing writing shapes contexts and instruction. In *Naming What We Know: Threshold Concepts Of Writing Studies* (pp. 29-31).

Scouller, K. (1998). The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Education, 35*(4), 453-472.

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist, 57*(10), 1380-1400.

Stone, N. J. (1996, October). Applying in the Classroom What We Know about Groups and Teams. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 40, No. 8, pp. 449-453). Sage CA: Los Angeles, CA: SAGE Publications.

Thompson, A., Sattler, B., & Turns, J. (2011, October). Understanding a studio environment: A complex system approach to a community of practice. In *2011 Frontiers in Education Conference (FIE)* (pp. F3H-1). IEEE.

Trevelyan, J. (2011). Are we accidentally misleading students about engineering practice?. In *Research in Engineering Education Symposium 2011* (pp. 268-278). Universidad Politecnica de Madrid.

Turns, J., Newstetter, W., Allen, J. K., & Mistree, F. (1997, June). Learning essays and the reflective learner: Supporting reflection in engineering design education. In *1997 Annual Conference* (pp. 2-274).

Wallot, S. (2017). Recurrence quantification analysis of processes and products of discourse: A tutorial in R. *Discourse Processes, 54*(5-6), 382-405.

Dr. Aneet Narendranath is an Associate Teaching Professor in Mechanical Engineering.¬† His eeucation research is focused on learning analytics with a specific emphasis on applying natural language processing and network analysis to close the achievement gaps in undergraduate STEM education. ¬†

Prof. Jeffrey S. Allen is the John F. and Joan M. Calder Professor in Mechanical Engineering. For the last six years he has served as the currently the Director of Undergraduate Studies in the Department of Mechanical Engineering - Engineering Mechanics where he has lead efforts to incorporate digital engineering tools such as natural language processing and machine learning into undergraduate coursework, curriculum assessment, and department operations.