The Future of Engineering Education
2024 Annual Conference & Exposition
Oregon Convention Center
Portland, OR . June 23 - 26, 2024
Paper ID #42554

Use of Sentiment Analysis to Assess Student Reflections in Statics

Dr. Amie Baisley, University of Florida

I am an Instructional Assistant Professor at the University of Florida teaching primarily 2nd year mechanics courses. My teaching and research interests are alternative pedagogies, mastery-based learning and assessment, student persistence in their first two years, and faculty development.

Chiranjeevi Singh Marutla, University of Florida

Use of Sentiment Analysis to Assess Student Reflections in Statics

In a flipped, mastery-based Statics course, the students are tested on a single course problem every other week. Each problem is graded by the instructor using a rubric to score the different mastery objectives for the single problem. The mastery objectives are the key parts necessary to solve every statics problem and there are eight of them that require the students to include text, equations, or a drawing for that part. Following the in-class assessment, the students are asked to complete a self-assessment of their work where they grade themselves for each objective and comment on how they performed for each one. The qualitative data collected from the student comments have depth and richness of the students thoughts and reflection on their work in the course. Over the last four years, a massive amount of qualitative data has been collected through these self-assessments that includes roughly 780 student comments per assessment with seven assessments completed each semester. Prior to this study, it has not been feasible to analyze the student comments for meaningful results. However, the use of machine learning approaches has proven beneficial for understanding written comments, and natural language processing (NLP) is an area of machine learning that allows computers to process and begin to identify ideas in written text. This offers a systematic and efficient way to analyze the student reflections. For this study, the initial analysis of the student comments was completed using sentiment analysis. Sentiment analysis determines whether the text is positive, negative, or neutral. The results of the sentiment analysis are used to better understand the students attitudes towards the different parts of the statics problem. This paper will provide the results from the sentiment analysis for different assessments throughout a semester along with how the polarity of student comments compare to their assessment scores.

Introduction

In engineering, students are often assessed on their knowledge of theory and their problem-solving skills. These assessments are scored, and this quantitative information is considered the students' evaluation of their learning in the course. This system does not provide the students opportunity to close the loop on their learning through reflection. Reflection provides students with an opportunity to revisit their work, assign meaning to the experience, and guide their future actions [1]. The National Academies has called for more "opportunities for reflection to connect thinking and doing, and to [develop] students' metacognitive abilities to foster self-directed, lifelong learning skills [2]. Implementing opportunities to reflect promote the students to critically review their work and process the outcome to further encourage their learning. Reflective exercises also have a rich detail of the students understanding, experience, and their process used during the assessment exercise. This insight can complement the student scores and inform an instructor of student ability often better than the quantitative data, but a major challenge is the qualitative nature of student reflections.

The use of open-ended student reflections where students aren't guided on what to discuss but asked to reflect on their previous work can lead to massive amount of written text that can address all different aspects from the problem to the topic to the structure of the assignment. [1]. The reflection exercise itself is valuable but the desire to find meaningful information from it is a significant obstacle for many. The increasing use of artificial intelligence (AI) involves tools that can help analyze reflection data to digest and interpret it in a succinct way.

Natural language processing (NLP) is a subfield of AI that enables machines to understand, interpret, and respond to human language in both written and spoken forms. As a crucial area within AI, NLP facilitates the development of applications and systems capable of tasks such as spam detection, translation, powering chatbots, conducting sentiment analysis, and summarizing text [3-5]. These capabilities allow NLP to significantly enhance the interaction between computers and human language making it a key component in advancing AI technologies that are being used in a variety of ways in business and education spaces today. NLP offers tools to analyze the large amount of written information collected through student reflections and report it for instructors to use in a meaningful way [6].

Examples of NLP in education include identifying student certainty in answering conceptual questions for a signals and systems course. The student text was analyzed through lexical analysis to categorize written responses and determine a relationship between students that guessed and answered incorrectly [7]. A first-year engineering course was transitioned online halfway through the semester and the students were asked to reflect on the experience. An NLP program was used and compared to an instructor’s codes to identify common categories that were discussed by the students [8]. NLP was also used to identify conceptual or procedural discussions from recently graduated engineering students in the workforce and how those discussions change over time [9]. These examples represent the variety in the use of NLP, but the area that was of interest for the reflection data of this study is the use of sentiment analysis.

Sentiment analysis, also known as emotion analysis, involves leveraging NLP to detect and assess the affective qualities of human language such as the sentiment polarity of a sentence, paragraph, or document [10]. This approach facilitates the categorization of the texts sentiment into ranges of positive, negative, or neutral effectively quantifying the emotional resonance of the language [13]. Sentiment analysis has been used in many fields and different sentiments are captured in each discipline. A medical example identified student emotions including surprise, fear, anger, joy, and other emotions in their reflections about different body regions [10]. The use of sentiment analysis was performed on psychology students to help categorize their level of reflection [11]. It has been used to determine how emotions and views affect teacher lesson plans for their elementary students on new material relating to computer science [12]. In a study done with first year engineering students the emotion of the student reflections was determined using sentiment analysis to find that gender and race played a role in the emotion of the response recorded [13]. Another example includes student reflections on a mechanical engineering assignment that are scored with sentiment analysis to determine the overall student feeling on the assignment and guide further course improvement [14]. A final example categorizes first year student emotions as they transition into college over the course of their first semester [15].

The many uses of sentiment analysis have been used to identify many different parameters. This study will use sentiment analysis to identify the polarity of student responses on their reflections following the course assessments. The identification of polarity is the first step to analyze the large amount of reflection text collected from the second-year undergraduate engineering statics course.

Student Reflections

In their second year many engineering students, particularly aerospace, civil, and mechanical students, are required to take courses focused on mechanics. Statics is the first course in the mechanics series where students learn the foundations of equilibrium that will be applied in many later courses. At a large R1 university in the southeast students take the course in a flipped, mastery-based classroom environment. The mastery-based approach is employed for the assessments where students are evaluated for how they demonstrated the course mastery objectives on each assessment in the course. The students solve a single assessment problem every other week where they are asked to organize their solution following the mastery objectives. The mastery objectives are the key pieces of the solution solving process for every statics problem. The students are required to write, draw, or include equation(s) for each objective for each problem, but the work is unique to the type of problem being assessed each week. This allows the students to show mastery on the pieces needed to solve statics problems for different problems on different days throughout the entire course [16]. The mastery objectives and a description of their requirements are shown in Table 1.

Table 1. Mastery Objective for Statics including an abbreviated version of the requirements.

Objective	Student requirements for the objective
A) Modeling	List the constraints of the problem and any assumptions used to model the system
B) Solution strategy	List how force and moment equilibrium will be used to solve for the unknowns in the problem
C) Problem geometry	Set up all necessary direction vectors and load functions
D) Free body diagram (FBD)	Draw a free body diagram for each body required to solve the problem
E) Force equilibrium	$\sum \mathbf{F} = \mathbf{0}$ for each FBD
F) Moment equilibrium	$\sum \mathbf{M} = \mathbf{0}$ for each FBD
G) Distributed effects	Evaluate all integrals from the force and moment equations
H) Solve	Solve each force and moment equation in a logical order to find the unknowns

Every student is graded by the instructor for each individual objective using the rubric shown in Table 2. The students are also asked to grade themselves following each assessment and prior to receiving instructor feedback using the same rubric. This self-grading activity is known as the self-assessment. The self-assessment is not a required assignment, but the students receive participation points for doing it, so most students complete it. The self-assessment requires the students to grade themselves for each objective using the same rubric as the instructor and to comment on their work for each objective. The students are asked to do the self-assessment prior to instructor feedback so their scores are not impacted by the instructor scores. This is an example of authentic self-assessment since the students grade themselves in the same way that the instructor grades them

[17]. Requiring the self-assessment to be completed immediately after the assessment and before instructor feedback also provides incentive for the students to review the solution and reflect on how they did without their actual grades impacting their thoughts. The self-assessment captures the students raw thoughts. The quantitative results for student accuracy on the self-assessments compared to instructor scores was previously documented [18].

Table 2. Grading rubric used to evaluate mastery for each objective.

Label	Description
a	Understanding is correct and complete
b	Significant understanding but with minor errors
c	Some understanding but with minor logic errors
d	Some recognition shown, but with significant conceptual errors
e	No evidence shown

The comment section of the self-assessment is an open-ended reflective exercise. For each objective the students are given the prompt to "Comment on your response for Objective [X]". This leaves it open for the students to comment on how they did on the objective, what they did correct/incorrect, if they realized something about the problem, or even celebrate what they did. The responses vary for each student, for each objective, for each test, and responses can range from a single word to a small paragraph. The amount of text data collected in the self-assessments is too much for an individual instructor to review. However, it is within these comments that contain valuable information about student learning and how they are reflecting. The data included in this study is from 7 semesters that include 7 assessments each. The time it would take a person to read the comments and identify useful information is not plausible. Table 3 quantifies the number of reflection comments collected for each semester.

Table 3. Number of comments recorded for each assessment by semester.

	Assessment Number						
Semester	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Spring 2020	750	796	743	544	665	766	736
Fall 2020	496	547	613	547	616	507	595
Fall 2021	859	896	969	750	861	885	865
Spring 2022	696	734	669	579	664	624	602
Fall 2022	779	864	816	729	830	785	761
Spring 2023	805	881	849	650	816	819	723
Fall 2023	934	1060	1012	862	975	784	914

The number of recorded comments varies with each assessment due to the number of students that completed the self-assessment and the number of objectives on the assessment. For assessments 1 and 4 only seven of the eight objectives were tested, but all eight objectives were tested on the other five assessments. The length and intent of the comments varied, but below are two examples of comments including how they were categorized for the polarity analysis:

> "I made the free body diagram with all the forces included, but I did not include the couple moment. I was confused on whether or not to include it as it itself isn't a force but a moment. Otherwise, I had all my forces with their direction vectors attached." – Student comment from Assessment 1 on Objective D) Free body diagram. Polarity categorization: *Negative*
>
> "All of the forces are summed up correctly, as they are added with respect to their direction vector. Those that go opposite to that vector are subtractions to the equation, while the rest are added in the problem. Each of them is defined by both a force component and their direction, which are accurate to what was shown in the diagram that it is connected to, and they are labelled for easy reference to those diagrams. All of them are set equal to 0, which is necessary as that problem is for a static object. Finally, the bounds of integration are done correct for each of the different sections and diagrams for the problem." – Student comment from Assessment 6 on Objective E) Force equilibrium. Polarity categorization: *Positive*

Natural language processing offers a route to begin identifying useful information from this large data set. The first step of the study was to identify the polarity of the comments through sentiment analysis. This would determine whether the students were commenting in a positive sense, negative sense, or neutral on the self-assessments. The trends in polarity would help determine overall student feeling for the specific problem and objective. Polarity was chosen as the first analysis to do on the reflection data to understand a general breakdown for student responses and determine the accuracy of sentiment analysis on this type of date. The polarity does not identify students that did the objective correctly/incorrectly but identifies how they commented about their work. In the future analysis will be done on the emotional response of the comment which will identify explicitly what type of positive or negative response was used.

Sentiment Analysis Procedure

To perform the analysis the data first had to be deidentified, organized, and cleaned for processing. The original data is pulled from a Canvas quiz into a Microsoft Excel sheet. This Excel sheet is processed through a Python program to complete the sentiment analysis. There were four pre-established open-source sentiment analysis programs tested to find the one that best recognized the comment polarity. The four programs were TextBlob, Natural Language Toolkit, Deep Learning, and BART. An explanation of each program is provided next.

- **TextBlob:** TextBlob [19] is a Python library that provides simple natural language processing (NLP) capabilities, including sentiment analysis. TextBlob uses a pre-trained machine learning model to perform sentiment analysis. TextBlob's sentiment analysis is relatively simple and may not be as accurate or nuanced as more advanced sentiment analysis models. It does not consider context or sarcasm in the text, which can lead to misclassification. For more advanced sentiment analysis tasks, more sophisticated models and techniques, such as deep learning-based models, are often used.

- **Natural Language ToolKit (NLTK):** NLTK [20] is a Python library commonly used for natural language processing tasks, including sentiment analysis. NLTK provides various tools and resources for processing and analyzing text data. For this project the *SentimentIntensityAnalyzer* was used which is a part of the NLTK library. It provides a simple way to assess the sentiment of a piece of text by assigning a sentiment score to it, ranging from -1 (most negative) to 1 (most positive), with 0 indicating a neutral sentiment. NLTK is a versatile library that allows more control over the sentiment analysis process, but it requires more manual effort and expertise compared to libraries like TextBlob.

- **Deep Learning:** Deep learning models [21] for sentiment analysis are more sophisticated and capable of capturing complex patterns in text data, making them potentially more accurate than shallow models like TextBlob and NLTK; however, it must be trained to be used on data. To do this, the comments from our data had to be labeled with their associated sentiment labels (positive or negative). The data was preprocessed involving tokenization, lowercasing, removing punctuation, and split into a training set and a testing set. The training for the deep learning model requires the model to adjust its internal parameters to minimize the prediction error. The outcome is for the program to learn the relationships between the words and sentiments associated with it. After training, the model is used to classify new comments into positive/negative sentiments.

- **BART-Transformer LLM approach:** Large Language Models (LLMs) are advanced machine learning models built on transformers. These models are created to do specific tasks and are trained with huge amounts of data. Training these models to do their tasks well requires a lot of computer power. To make things easier, there are pre-trained models available. These are models that have already been trained to do certain tasks. For example, in our study, we use a pre-trained model named *facebook/bart-large-mnli*. This model was developed by Facebook and is based on the BART technology. It has been trained on a large dataset known as MNLI, which helps it understand and process language effectively.

  The *facebook/bart-large-mnli* model [22] is a variant of the BART (Bidirectional and Auto-Regressive Transformers) model fine-tuned for the MNLI (MultiNLI) dataset, which is primarily used for natural language inference tasks, including sentence classification into positive or negative categories. BART is a transformer-based model, and the f*acebook/bart-large-mnli* model utilizes its capabilities for sentiment classification. It leverages its understanding of language and relationships between sentences to classify input sentences into positive or negative sentiments using an additional classification layer.

There were four programs tested because each one increased in complexity and had different results. The first three programs (TextBlob, Nltk, and Deep Learning) were shallow open-source programs or had to be trained, so they all had shorter run times but were not as accurate. The multiple programs were initially selected to offer multiple sources of output that could be used to validate the results. It was quickly realized that the results were not matching amongst these three programs and the researchers needed to find an alternative program. The BART model has been used on much larger scale data and bigger applications making it highly trained and resulting in better results. The run time for a data set this size was anywhere from 2-8 hours for each assessment depending on the model used.

The output from each program was a score for the comment and based on that score the comment was initially categorized into one of five polarity scales: negative, slight negative, neutral, slight positive, positive. Each program was tested on a small sample of the data. The sample included 50 students from a single semester for two different assessments. This sample size was adequate for initial testing which produced over 700 polarity ratings from each of the four programs. The instructor of the course separately analyzed the comments from that sample set and provided their polarity rating. Each programs polarity was compared with the instructors polarity ratings and the program with the best percent match was chosen to use for the final analysis on all the data. The percentage match between the program and instructor ratings is given in Table 4.

Table 4. Percent match between sentiment analysis done by the instructor and a program

Program comparison	Assessment 1 % match	Assessment 7 % match	Overall % match
TextBlob	16.7%	38.8%	30.4%
Nltk	20.8%	0%	11.6%
Deep Learning	16.7%	0%	9.3%
BART 5-scale	46.8%	42.9%	44.9%
BART 3-scale	80.8%	81.0%	80.9%

The different programs varied in how accurate the percent match was with the instructor rating, but using the five categories scale had no strong results. The BART analysis was the most promising program with the highest percent match, so that program was compared to the instructor's ratings using a scale with only three categories: negative, neutral, and positive. Using the three-category scale the percent match was just above 80% so this program was chosen to complete the sentiment analysis for all the data. The BART program had many more results identified as slight-positive or slight-negative while the instructor ranked those comments as positive or negative, respectively, so reducing to a three-category scale did not change the overall interpretation of each comment's sentiment.

Results

The initial results from the sentiment analysis are computed for all seven semesters combined. The results are organized for each individual assessment. Figure 1 includes the sentiment analysis of the percent of comments that were negative, neutral, or positive for each objective on each assessment.

The results in Figure 1 broken down by assessment show that student reflection comments start out evenly split between positive and negative comments but the comments trend more positive for each assessment. This is an encouraging sign that students are reflecting more positively as the semester goes on. That could be due to a better understanding of the material or being more comfortable with the objectives. However, the most positive assessment was Assessment 4, Figure 1-d. The problem type for Assessment 4 is truss analysis which tends to be favored by the students and is likely the reason for the more positive comments.

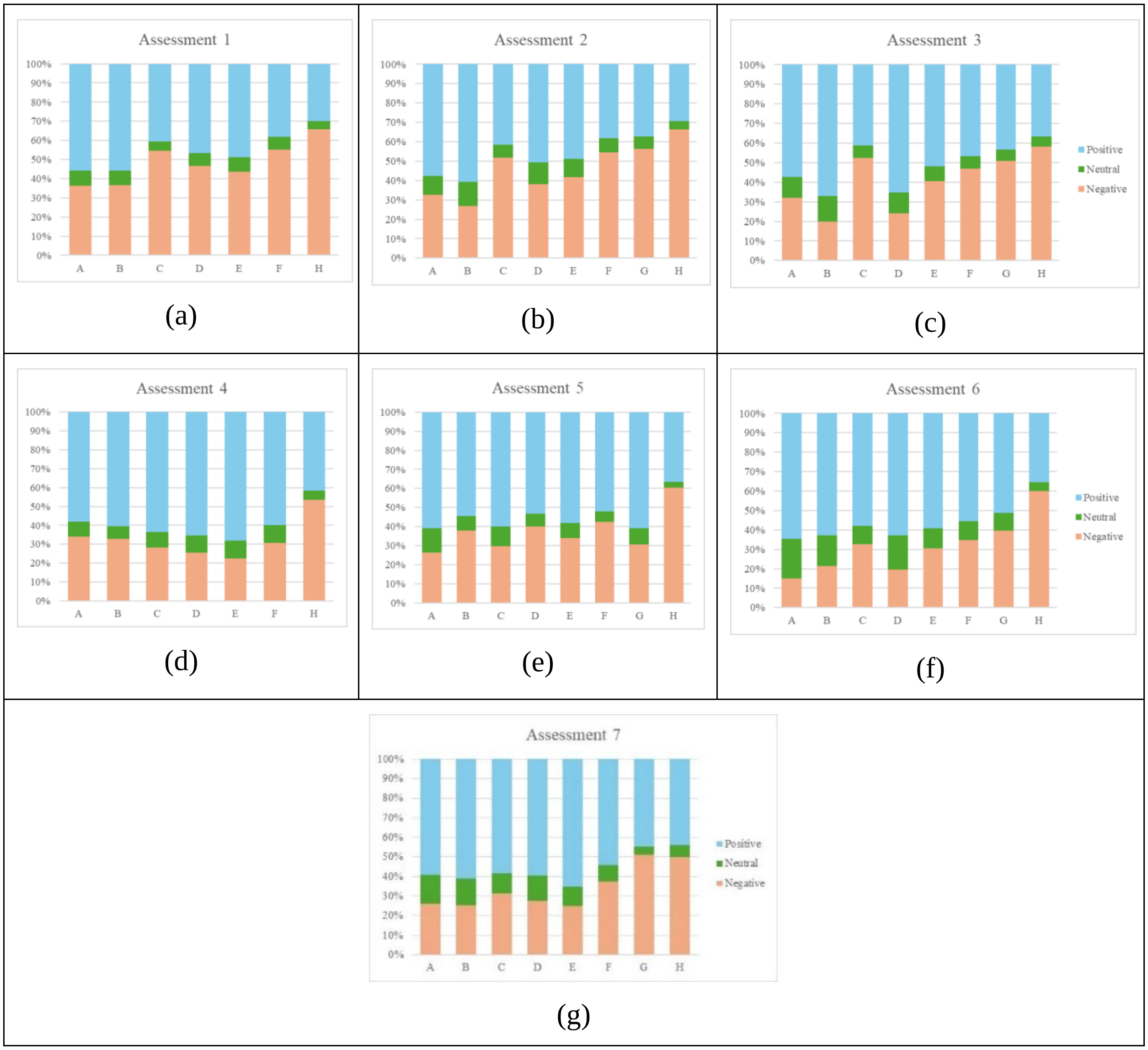

Figure 1: Sentiment analysis for each test by objective. The percentage of sentiments for each objective for each assessment that are negative comments are shown with the orange portion of the bar (the bottom color), and the positive comments are the blue portion of the bar (the top color).

The overall assessment trend increases in the percent of positive comments with each assessment, but each individual objective has a unique trend. Figure 2 shows the trends for the percentage of positive comments for each objective.

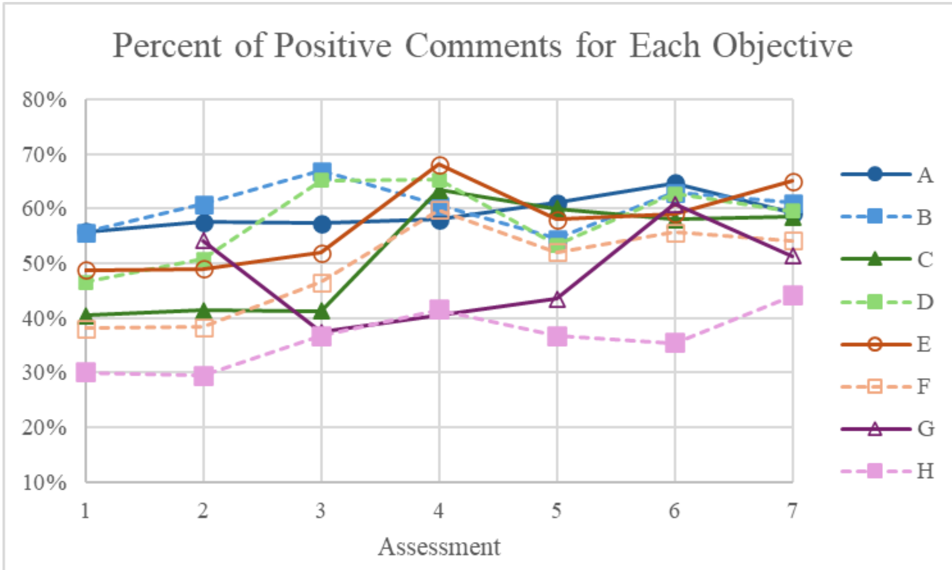

Figure 2: Sentiment analysis for the positive comments for each objective. The different colored and shaped points represent the eight different objectives.

The objective that always had the fewest positive comments was Objective H. Objective H is Solve, which is the last step of the solution and students do not always have enough time for this part of the problem. It is also an objective that is easy for a previous mistake to cause error in the final answer. Students tend to get the final answer wrong in this part from something they did prior to this objective, but if they do not get the correct numbers, it is marked as incorrect. As a result, it is a challenging objective for students to get full credit which could be reflected in the lower number of positive comments. The percentage of positive comments does increase over the course of the assessments. There is a drop for assessment 5 and 6 due to the problem type of those assessments which are more time consuming problems and harder for students to have the time they would like to solve the problem.

Objective C is Problem Geometry requiring students to create unit direction vectors and load functions for distributed loads. This objective had the largest increase in positive comments at Assessment 4 and continued to be more positive for the rest of the assessments. This trend is due to the increase in student comfort level with the objective and having 6 weeks of repeated practice with the objective prior to Assessment 4, so at that point most students felt more positive towards the geometry portion.

Objective G, which is Distributed effects, had the largest drop from Assessment 2 to Assessment 3 and had a large increase from Assessment 6 to Assessment 7. This is likely attributed to the problem types on those assessments and the comfort students felt with evaluating load functions. Assessment 3 is a hydrostatic pressure problem, which students do not do as well on compared to a problem with an applied distributed load like Assessment 2. The negative comments for this assessment are likely the result of not feeling comfortable with water pressure functions or disappointment in messing up the integration part of the problem. The large increase from Assessment 5 to Assessment 6 comes from the increase in practice that the students have with distributed loads in Module 6 compared to the previous modules. The students are likely commenting more positively because they feel more comfortable with this objective on that assessment.

Objective A and Objective B had the smallest amount of change in positive comments across the assessments. Objective A is Modeling and Objective B is Solution strategy, which are the two objectives that require the students to write down items about the problem. They list the constraints and assumptions used for Objective A and they have to describe how they will solve the problem for Objective B. Since these are the first two objectives the students complete for every problem in the class they gain the most comfort with these since they often can complete this part during the in-class recitation environment. Objective A requires students to state characteristics of the problem, which falls on the lower level of Bloom's Taxonomy, and becomes somewhat repetitive for many of the problems in Statics. Objective B has a small incline in positive comments at first and then drops down around assessment 4 and 5. The students start to get comfortable with writing a solution strategy early on through recognizing patterns in how to solve problems, but for Assessment 4 and 5 the strategy is more involved since those assessments deal with truss and frame problems.

The sentiment polarity of the student comments can be categorized into two types of comments, correctness or attitude. The comment could be marked as positive if the student mentioned something they did correctly, and identified as negative if they mentioned something that was incorrect. A positive comment could also be identified by a student that used a positive attitude for their comment about what they learned or could do better next time. The positive comments are not precisely correlated with the students that got that objective correct and the negative comments do not only connect with the students that were incorrect. The assessment scores of students for three of the semester assessments are provided in Figure 3. This compares the scores received to the percentage of positive comments for that objective. The student scores are based on the rubric from Table 1.

In many cases more students did the objective correctly than those that commented positively on it. This is a critical result for reflections demonstrating that students will reflect and self-assess themselves more harshly than the actual outcome. To complete the self-assessment, the students are provided with the correct solution and an image of their work to compare the two, but some students still have negative comments on their work that was correct. This was especially true for objective A and E. The fewer positive comments compared to correct scores could be due to both a lack in confidence on the concepts and a result of the students not understanding the variety of possible correct answers for an objective.

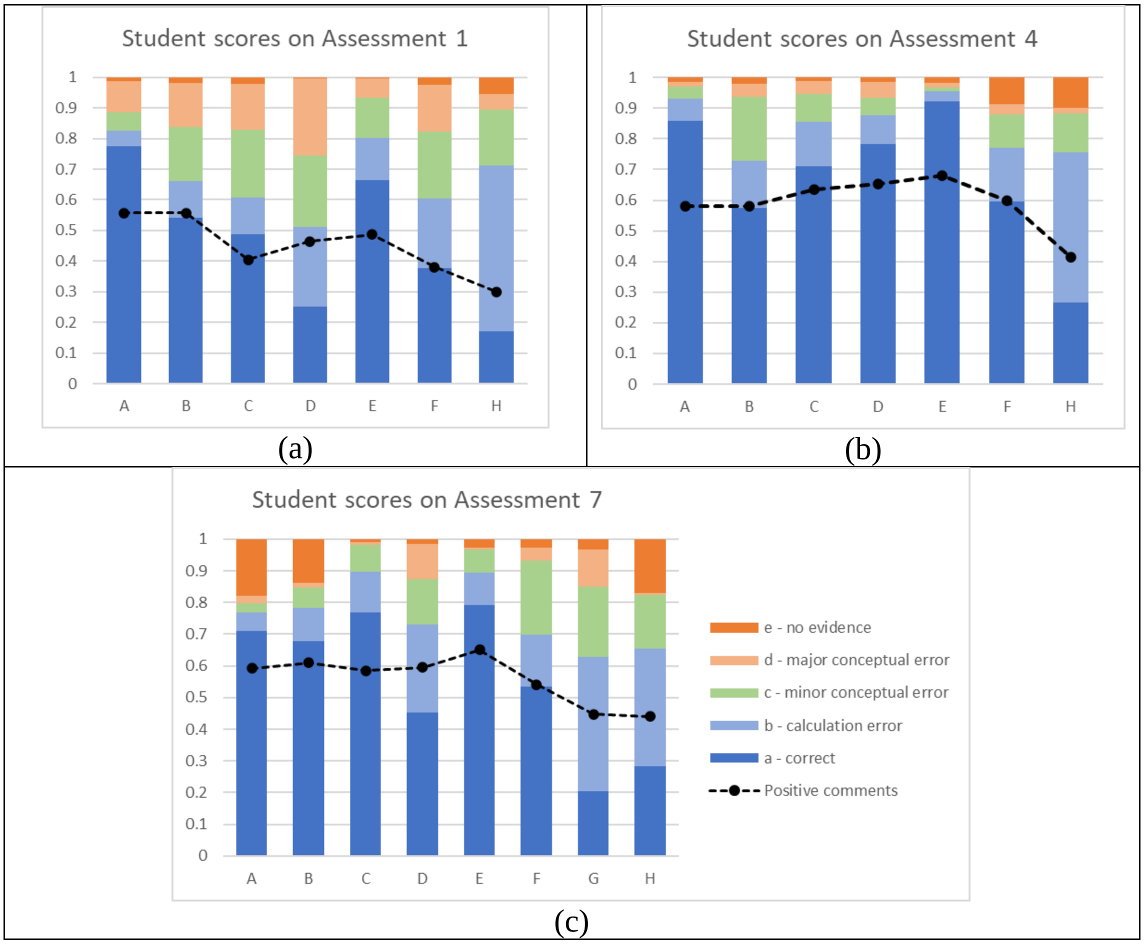

Figure 3: Student scores for Assessment 1, 4, and 7. The bar chart shows the percentage of students that received each score, and the black dots represent the percentage of positive comments for that objective.

Objective H, and a few other individual assessment objectives, are an example where students commented more positively than the number of students that did it correctly. The positive comments in this case would be positive attitude statements about knowing what to do but not having the time to do it or that they knew how to complete the problem correctly but had a previous math error that resulted in a wrong final answer.

Conclusion and Future Work

The use of NLP provides the sentiments of reflective comments to be analyzed in a more efficient and feasible way rather than reading each comment individually. The sentiment analysis on student self-reflections in Statics provides a route to classify student comments as positive, negative, or neutral but is not directly correlated to correctness. This is the first step in detecting meaningful outcomes from the student comments that can be used during a future semester to provide the instructor with real time analysis of the student's feelings for each assessment. The next step of this research is to continue using NLP methods to identify details from the comments that can be used to further improve the student learning experience. This will include identifying the types of emotions expressed in the comments and creating lists of common keywords associated with different sentiments and emotions. Some of the emotions to be identified will include anxiety, disappointment, frustration, happiness, hopefulness, realization, or neutral.

The results from this research will be used to create new resources and further the individualized feedback for each student to include guidance based on their needs identified in their reflection. Negative polarity was associated with doing the objective incorrectly or not understanding something about the objective, so identifying the students that respond primarily negative can allow the instructor to recognize that these students may need additional support. Resources can be provided in an individualized manner for the objectives the student responded negatively to. Further investigation of the polarity of each comment will be correlated with a student's self-efficacy in the course. Through looking at this trend of comment polarity over the course and comparing it to student performance, it is hypothesized that a student that comments more positively will have higher self-efficacy towards their abilities in the course and will likely improve more than students that are commenting more negatively. The use of natural language processing will provide a quick way to start identifying these trends and relationships between assessment scores and student reflections.

References

[1] J. A. Turns, B. Sattler, K. Yasuhara, J. L. Borgford-Parnell, and C. J. Atman, "Integrating reflection into engineering education," *ASEE Annu. Conf. Expo. Conf. Proc.*, 2014.

[2] S. A. Ambrose, "Undergraduate Engineering Curriculum: The Ultimate Design Challenge," in *The Bridge - Linking Engineering and Society*, vol. 43, no. 2, 2013, pp. 16–23.

[3] IBM, "What is natural language processing (NLP)?," 2024. [Online]. Available: https://www.ibm.com/topics/natural-language-processing.

[4] E. Cambria and B. White, "Jumping NLP curves: A review of natural language processing research," *IEEE Comput. Intell. Mag.*, vol. 9, no. 2, pp. 48–57, 2014.

[5] Chowdhary, K.R. (2020). Natural Language Processing. In: Fundamentals of Artificial Intelligence. Springer, New Delhi. https://doi.org/10.1007/978-81-322-3972-7_19

[6] M. Menekse, "The Reflection-Informed Learning and Instruction to Improve Students' Academic Success in Undergraduate Classrooms," *J. Exp. Educ.*, vol. 88, no. 2, pp. 183–199, 2020.

[7] A. M. Goncher, D. Jayalath, and W. Boles, "Insights into Students' Conceptual Understanding Using Textual Analysis: A Case Study in Signal Processing," *IEEE Trans. Educ.*, vol. 59, no. 3, pp. 216–223, 2016.

[8] A. Katz, M. Norris, A. M. Alsharif, M. D. Klopfer, D. B. Knight, and J. R. Grohs, "Using Natural Language Processing to Facilitate Student Feedback Analysis," *ASEE Annu. Conf. Expo. Conf. Proc.*, 2021.

[9] C. A. Arbogast and D. Montfort, "Applying natural language processing techniques to an assessment of student conceptual understanding," *ASEE Annu. Conf. Expo. Conf. Proc.*, vol. 2016-June, 2016.

[10] K. J. Rechowicz and C. A. Elzie, "The use of artificial intelligence to detect students' sentiments and emotions in gross anatomy reflections," *Anat. Sci. Educ.*, no. February, pp. 1–13, 2023.

[11] C. Chong, U. U. Sheikh, N. A. Samah, and A. Z. Sha'Ameri, "Analysis on Reflective Writing Using Natural Language Processing and Sentiment Analysis," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 884, no. 1, 2020.

[12] B. R. Belland, C. M. Kim, A. Y. Zhang, E. Lee, and E. Dinç, "Classifying the quality of robotics-enhanced lesson plans using motivation variables, word count, and sentiment analysis of reflections," *Contemp. Educ. Psychol.*, vol. 69, no. February, p. 102058, 2022.

[13] A. Roy and K. E. Rambo-Hernandez, “There’s So Much to Do and Not Enough Time to Do It! A Case for Sentiment Analysis to Derive Meaning From Open Text Using Student Reflections of Engineering Activities,” *Am. J. Eval.*, vol. 42, no. 4, pp. 559–576, 2021.

[14] Y. Wu, Z. Ming, J. K. Allen, and F. Mistree, “Evaluation of Students’ Learning Through Reflection on Doing Based on Sentiment Analysis,” *J. Mech. Des.*, vol. 145, no. 3, pp. 1–12, 2023.

[15] A. Satyanarayana, K. Goodlad, J. Sears, P. Kreniske, M. F. Diaz, and S. Cheng, “Using natural language processing tools on individual stories from first year students to summarize emotions, sentiments and concerns of transition from high school to college,” *ASEE Annu. Conf. Expo. Conf. Proc.*, 2019.

[16] Hjelmstad, K. D., & Baisley, A. (2020, June), *A Novel Approach to Mastery-based Assessment in Sophomore-level Mechanics Courses* Paper presented at 2020 ASEE Virtual Annual Conference Content Access, Virtual On line . 10.18260/1-2—34028

[17] Kearney, S., “Improving engagement: The use of ‘Authentic self-and peer-assessment for learning’ to enhance the student learning experience,” *Assess. Eval. High. Educ.*, vol. 38, no. 7, pp. 875–891, 2013.

[18] Baisley, A., & Hjelmstad, K.D. (2022 June) *The accuracy of student self-assessment in engineering mechanics*, Paper presented at 2022 ASEE Annual Conference & Exposition, Minneapolis, MN. https://peer.asee.org/40837

[19] TextBlob: Simplified Text Processing - https://textblob.readthedocs.io/en/dev/

[20] NLTK - Natural Language Toolkit - https://www.nltk.org/

[21] A. Mabrouk, R. P. D. Redondo and M. Kayed, "Deep Learning-Based Sentiment Classification: A Comparative Survey," in IEEE Access, vol. 8, pp. 85616-85638, 2020, doi: 10.1109/ACCESS.2020.2992013.

[22] facebook/bart-large-mnli - https://huggingface.co/facebook/bart-large-mnli