2021 ASEE ANNUAL CONFERENCE
Virtual Meeting | July 26–29, 2021 | Pacific Daylight Time
Paper ID #34693

Work in Progress: Assessment of Reflective Thinking in Graduate Engineering Students: Human and Machine Methods

Dr. Roman Taraban, Texas Tech University

Roman Taraban is Professor in the Department of Psychological Sciences at Texas Tech University. He received his Ph.D. in cognitive psychology from Carnegie Mellon University. His interests are in how undergraduate students learn, and especially, in critical thinking and how students draw meaningful connections in traditional college content materials.

Dr. Micah Iserman, Texas Tech University

Interested in language use, mental representation, and social processes.

Ms. Jessica C. Pittman, Texas Tech University

Jessica C Pittman is a third year graduate student in the Experimental Psychology doctoral program at Texas Tech University. Current interests are in the overlap of cognition and education with a direct focus on note taking.

Mr. Nigel Yeo, Texas Tech University

Nigel Yeo is a senior student at Texas Tech University, pursuing a degree in psychology. He hopes to attend graduate school and become an occupational therapist in the future. His interests include travelling to new places and exploring the United States. His hobbies are playing tennis and piano.

Dr. Ryan C. Campbell, Texas Tech University

Having completed his Ph.D. through the University of Washington's interdisciplinary Individual Ph.D. Program (see bit.ly/uwiphd), Ryan is now a Postdoctoral Research Associate at Texas Tech University. He currently facilitates an interdisciplinary project entitled "Developing Reflective Engineers through Artful Methods." His scholarly interests include both teaching and research in engineering education, art in engineering, social justice in engineering, care ethics in engineering, humanitarian engineering, engineering ethics, and computer modeling of electric power and renewable energy systems.

Dr. Jeong-Hee Kim, Texas Tech University

Jeong-Hee Kim is Chairperson and Professor of Curriculum Studies and Teacher Education in the Department of Curriculum and Instruction at Texas Tech University. Kim is a curriculum theorist, teacher educator, and narrative inquiry methodologist. Her research centers on various epistemological underpinnings of curriculum studies, particularly engaging in hermeneutical excavation of the stories of students and teachers around the notion of Bildung, a human way of developing or cultivating one's capacity. She received the Faculty Outstanding Researcher Award in 2018 from Texas Tech University, and the Outstanding Publication Award from the American Education Research Association in 2017 for her book, Understanding Narrative Inquiry, published in 2016. She has published numerous articles in journals including Journal of Curriculum Studies, International Journal of Qualitative Studies in Education, and Educational Philosophy and Theory.

Dr. Danny D. Reible, Texas Tech University

Dr. Danny D. Reible is the Donovan Maddox Distinguished Engineering Chair at Texas Tech University. He was previously the Bettie Margaret Smith Chair of Environmental Health Engineering in the Department of Civil, Architectural and Environmental Engineering and the Director of the Center for Research in Water Resources at the University of Texas in Austin. Dr. Reible holds a Ph.D. in Chemical Engineering from the California Institute of Technology, and is a Board Certified Environmental Engineer, a Professional Engineer (Louisiana), and was elected to the National Academy of Engineering in 2005 for the "development of widely used approaches for the management of contaminated sediments". His research is focused on the fate, transport, and management of contaminants in the environment and the sustainable management of water resources.

Work in Progress: Assessment of Reflective Thinking in Graduate Engineering Students: Human and Machine Methods

Abstract

Engineering education is increasingly looking to the liberal arts to broaden and diversify preparation of students for professional careers. The present study involves an elective graduate environmental engineering course that incorporated the arts and humanities. The goal of the course was to develop engineers and technical professionals who would become both more appreciative of and better equipped to address technical, ethical, social, and cultural challenges in engineering through the development of critical and reflective thinking skills and reflective practice in their professional work. A reflective writing assignment was submitted by students following each of fourteen course topics in response to the following question: *Reflect on how you might want to apply what you learned to your development as a professional and/or to your daily life.* Student responses were classified by human coders using qualitative text analytic methods and their classifications were attempted to be learned by a simple machine classifier. The goal of this analysis was to identify and quantify students' reflections on prospective behaviors that emerged through participation in the course. The analysis indicated that the primary focus of students' responses was self-improvement, with additional themes involving reflection, teamwork, and improving the world. The results provide a glimpse into how broadening and diversifying the curriculum might shape students' thinking in directions that are more considerate of their contributions to their profession and society. In the discussion, we consider the findings from the human and machine assessments and suggest how incorporating AI machine methods into engineering provides new possibilities for engineering pedagogy.

Key terms: reflective writing, reflective thinking, qualitative data analysis, machine analysis

"*It is so Much Easier to Educate Students for Our Past than for Their Future"* Aldert Kamp

Introduction

The quote above from Kamp's 2020 [1] book[1] implies that engineering education should look forward, not backward, in its pedagogical principles and practices. According to Kamp, the world is rapidly changing, increasingly complex, often chaotic, and being re-built on the reality of globalization and connectedness. He suggests that "It might even be more important to found educational change on the things technology cannot do, the things that are strictly human!", and advises that "Engineering students have to learn that people policies, environmental aspects, politics, economics or cultural values often override disciplinary expertise" (p. 17).

The idea of building engineering curricula with attention to globalization, social and cultural factors, humanitarianism, and social justice, is not new to engineering educators. Attempts to broaden engineering curricula have included creating learning communities consisting of faculty from engineering and disciplines in the humanities, and developing new courses that integrate

[1] Quote originally appeared in Andreas Schleicher "Educating students for the fourth industrial revolution" https://thehill.com/opinion/education/390565-educating-students-for-the-fourth-industrial-revolution

liberal arts with engineering [2]. Course objectives in liberal-arts-infused courses have focused on critical thinking, reflection, and identity development [3].

Pedagogical methods underpinning the broadening of engineering education generally involve developing the reflective practitioner, as exemplified in the work of Donald Schön [4]. Schön proposed that reflective practice should accompany the traditional engineering curriculum involving technical expertise. Expanding on that idea, Bolton and Delderfield [5] asserted that professionals are responsible for not only their own actions and values, but also for the broader contexts involving the political, social, and cultural situations within which they live, learn, and work. The development of a reflective practitioner involves developing self-awareness and extracting meaning from experiences for personal and professional growth.

The present curriculum was designed to develop reflective, critical, and creative thinking about the broader social, environmental, and ethical contexts of engineering practice. The course was built around fourteen topics summarized in Table 1, with a selection of related in-class and outside-class activities.

Table 1. The 14 Topics of the Course Curriculum

Topics	**Brief Descriptions** (with sample readings/materials)
1. Course Foundations	Bolton and Delderfield's [5] notion of reflective practice; ethics case studies, e.g., VW emissions scandal
2. Reflective Practitioner	Schön's [4] reflective practitioner: crisis of confidence in professional knowledge, technical rationality, reflection-in-action
3. Introduction to Bildung	Bildung in the classroom: autobiographical writing as philosophical exercise (askēsis) for wholistic development of self [6]
4. Visual Thinking Strategies	Field trips to the university museum to practice Visual Thinking Strategies (see vtshome.org) using paintings
5. Relational Visual Reasoning	Semiotics (icon, index, and symbol) in the work of philosopher Charles Sanders Peirce; small group project applying semiotics to sculptures
6. Reflective Engineering	Definition and examples of moral dilemmas; engineering ethics [7]; individual character strengths (see viacharacter.org); ethical considerations in Flint water crisis [8]
7. Visual Storytelling	Basic elements of narrative [9]; developing a visual story outline and creating a group video story
8. Creative Art	Paper engineering [10] group art project
9. Narrative in Engineering	Engineering identity; human-centered vs tech-centered narratives; Engineers Without Borders “Dream Big” video; indigenous economies
10. Creative Engineering	Creative engineering of James Turrell (YouTube video); the Oculus in New York (YouTube video); engineering for sustainability, and solving for pattern [11]
11. VTS in Engineering	VTS review/practice with on-screen art images; apply VTS to diagrams and photos explicitly related to environmental engineering

12. Climate Change	Perspectives on climate change over time (National Geographic Magazine)
13. Design Your Own Ethics Case	Small group activity creating and analyzing an engineering ethics case
14. Bildung Presentations	In-class presentations of personal autobiographical stories and companion art creation; feedback from classmates and instructors

The goal of the present project was to identify and quantify students' reflections on prospective behaviors in their professional and personal lives that emerged through participation in the course.

Method

The curriculum was implemented in an elective graduate course in environmental engineering. The data reported here are from a course section that enrolled 20 engineering graduate students, with most majoring in environmental engineering in a masters-level program (see Table 2). The course was led by a professor in environmental engineering and a post-doc. Six faculty from non-engineering departments on campus led instruction and activities, as sole or co-instructors, on topics 3, 5, 7, 8, 9, 10, and 14 from Table 1.

Table 2. Distribution of Students

Engineering Major	Female	Male	Total
Environmental	11	5	**16**
Chemical	2	0	**2**
Bioengineering	2	0	**2**
Total	**15**	**5**	**20**

A reflective writing assignment after students completed each topic was based on Foucault's [12] notion of askēsis, a theoretical framework for care of self and others through reflective thinking [6]. At the conclusion of each of the fourteen topics in the course, students responded to the following (Foucault) question as part of a homework assignment: *Reflect on how you might want to apply what you learned to your development as a professional and/or to your daily life.* Students' responses to this portion of the weekly writing assignment are the data in the present study.

Analytic Methods by Human Coders

The analysis of students' responses to the writing assignment was carried out by human coders, who applied qualitative analytic methods [13]. The approach was bottom-up and inductive, meaning there were no prior hypotheses. The primary purpose was to uncover the general topics in the students' written responses. In the first phase of the analysis, three independent coders read through the responses to the Foucault-based question from each student for each of the fourteen topics. Because the responses were complex, the coders decided to parse each response into sentences, which resulted in 1,123 codable units. After reviewing the sentence units, the coders agreed to separate responses into three categories: those that addressed the "How" question directly, those that described an "Effect", and those that expressed a "Belief" or knowledge. The next phase followed a similar pattern, focusing on the "How" responses because these directly addressed the question that was posed to students. The coders agreed on classifying the How responses into five categories, proceeded to classify responses, and resolved any differences in classification through discussion and mutual agreement. The method is summarized in Table 3.

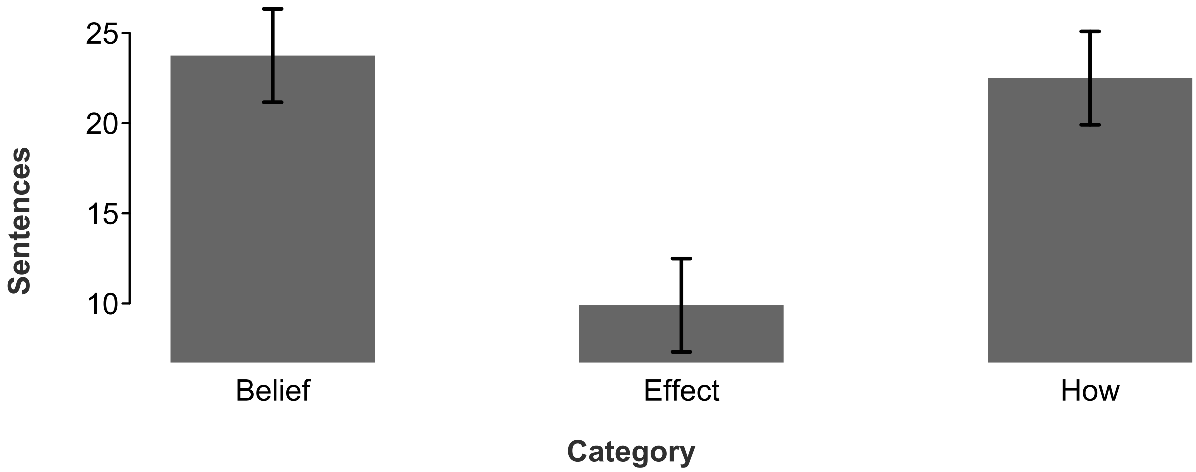
Table 3. Summary of Classification Procedure by Human Coders

Step	Description
1.	Three human coders read sentences from student responses and classified them into three categories: **1. How**—statements that directly addressed the question posed for the writing assignment: *Reflect on how you might want to apply what you learned to your development as a professional and/or to your daily life.* **2. Effect**—statements related to effects that the lesson might have, but not addressing "how." **3. Belief**—statements indicating personal beliefs or knowledge arising from within or outside a course lesson.
2.	Disagreements between raters were resolved via discussion, for a final classification of each sentence.
3.	The raters further classified the sentences classified as "How" into five subcategories: **1. Reflection**—statements related to self-consciousness, asking questions, keeping an open mind, thinking through issues, looking for the deeper picture, considering or predicting outcomes, reflecting on the implications of actions. **2. Self-Improvement**—statements related to curiosity, gaining skills, opportunities, adaptability, perseverance, attention to detail, happiness. **3. Teamwork**—statements related to leadership, humility, getting help, receiving help, speaking up, understanding all sides. **4. Improve the World**—statements related to environment, community, social issues, morality, ethics, empathy, perspective-taking, understanding, mindfulness. **5. Other**—statements that did not fit the other categories.
4.	Disagreements between raters were resolved via discussion, for a final classification of each "How" sentence.

Results – Human Assessment of Students' Responses

The average length of students' responses across the fourteen lessons was 4.01 sentences (standard deviation = 1.06; range = 2.50–6.29 sentences). Human raters agreed on the initial classification of the 1,123 sentences in students' responses into How, Effect, and Belief categories 87% of the time (see Table 3, Step 2). The final classifications were reached through consensus among reviewers and were then aggregated for each student across the fourteen lessons to derive the mean frequency of each classification for each student. Across the twenty students, most classifications fell into the Belief (total = 475; mean per student = 23.75) and How (450; 22.50) categories, with fewer classifications falling into the Effect category (198; 9.90), as shown in Figure 1. A Linear Mixed-Effects Regression model (with an intercept estimated for each student; fit in R [14] with the lme4 package [15]) indicated differences between sentence categories summed across student responses [$F(2, 38) = 18.43$, $p < .001$] (using Kenward-Rodger approximated degrees of freedom [16]), with Belief sentences being used more than Effect sentences ($b = -13.85$, $p < .001$), but about the same amount as How sentences ($b = -1.25$, $p = .62$) . Examples from students' responses are shown in Table 4.

Figure 1. Average Number of Sentences Per Category Across Students



Error bars show standard error.

Table 4. Examples of *Belief*, *Effect*, and *How* Sentences in Students' Responses

Human Classification Examples		
Belief	Effect	How
Especially for an engineer they have a great amount of responsibility they can't just do things because they have a gut feeling that it would turn out the way they want it to.	I feel that this class and the way it encourages reflection in action makes me feel more thought out in the decisions that I am making, be they simple and seemingly insignificant decisions that affect my life.	We should understand all sides of a problem, and try to think beyond your view and see how others may view this same situation.
In a way, when you are writing about design processes in engineering, you are writing a piece of your story.	I think these skills will only help me to get better at the skills I already use in my daily life.	Putting empathy into engineering can be a good practice in reflective engineering.
With how fast paced everything usually is, I get wrapped up in routine without thinking about how I really feel about my classes, my peers, or myself.	In my personal life, I think being able to reflect on my behavior with my friends or family has helped me get to where I am in my relationships with those people.	Asking questions through each step in the process can also help in being more reflective and avoiding mistakes that can impact whole communities.

Usually, I take what is being taught to me and try to synthesize the information at face value.	Then, after joining a company or group of researchers, if there ever is a potentially unethical dispute, I can feel comfortable expressing the problem at hand.	I realized that there will probably be times in the work place when I should speak up, either to ensure the ethical option is being considered, or to add new ideas.
Having self-discipline is an amazing quality in people and is hard to top in the greater scheme of things.	Thinking of my story has allowed me to be more reflective over my actions.	I would like to apply the things i learned to help myself maintain a clear perspective.
I may sound like a broken record at this point, but obviously I am going to be going through a huge transition period over the next year.	Being a minority in my field has me always questioning myself and who I am to different people, but really reflecting on your actions can give you a sense of the common threads.	In my professional life i would like to focus on keeping people safe while proving a necessary service.
Even though I will technically be the same person, there will be many experiences that are to come that will likely very quickly mold and change the way I view certain things.	I think this helped me to realize areas in my life where I have a black-and-white mindset, to where in the future I will be able to stray from analysis and approach these situations with a mindset to learn, discuss, and grow.	A thought process change seems like the best method of applying what I've learned so far to my professional life and my daily life.

In the next step of the analyses (see Table 3, Step 3), human coders classified the 450 How sentences into five subcategories, with 96% agreement on the initial classifications. The final classifications were reached through consensus among reviewers and were then aggregated for each student across the fourteen lessons to derive the mean frequencies of each classification for each student. Across the twenty students, Self-Improvement sentences were the most frequent How subtype (total = 197; mean per student = 9.85), with Improve World (89; 4.45), Reflection (90; 4.50), and Teamwork (67; 3.35) sentences seeing less use, and Other (7; .35) seeing least use, as shown in Figure 2. A Linear Mixed-Effect Regression model indicated differences between these subcategories [$F(4, 76) = 27.76$, $p < .001$], with Improve World sentences being used about the same amount as Reflection ($b = .05$, $p = .96$) and Teamwork ($b = -1.10$, $p = .24$) sentences, but less than Self-Improvement sentences ($b = 5.40$, $p < .001$), and more than Other sentences ($b = -4.10$, $p < .001$). Examples from students' responses are shown in Table 5.

Figure 2. Average Number of Sentences Per *How* Subcategory Across Students

Table 5. Examples of Improve World, Reflection, Self-Improvement, and Teamwork Sentences in *How* Sentences

Human Classification Examples			
Improve World	Reflection	Self-Improvement	Teamwork
In my professional life I would like to focus on keeping people safe while providing a necessary service.	Thinking creatively can help us think of other solutions not traditionally thought of before.	I will refocus my efforts on others, as society and life in general is a dangerous and brutal nightmare for many.	Being a good listener as an engineer can help in working with a team.
Considering others before myself is a healthy practice within my daily life and my life as a professional which naturally involves the process of reflection.	We can approach different problems through multiple ways to ensure we are looking at all aspects of the problem or project and ensure we have the best solution.	It helped me learn how to create an outline and then use that outline to create a piece of work.	When working as a team you must consider everyone's point of view because you might be missing out on a good opportunity.
I'm going to try to be more cognizant of the situations of the communities and people affected by an issue, while I work as a professional.	This is where reflective engineering can be utilized to see the bigger picture and find the best solution for everyone involved.	These lectures are an eye opener that a professional that has a responsibility for the general public that I need to review, review, and review again that all the paperwork or work I turn in is close to 100% right.	If you are able to paraphrase what everyone else is saying and to hold a questioning mindset while leading you foster open communication and the progression of the whole to see every angle of a problem.

By acting with concern for others you will not perform actions that affect other people and animals, for self-pleasure or greed.	This also helps in keeping an open mind which helps every engineer be more open to other solutions.	I want to apply what i learned from visual story telling in my professional life, specifically when i need to present information about a project.	I've learned that people have very different perspectives while looking at the exact same thing, and that would be a great thing to add into the work place.
In the professional setting, especially as an engineer, it is crucial to think of everything you're doing and how it is going to affect not only you, but everyone who will be affected by whatever project you are working on.	Art and Engineering have alot to offer each other, but to meaningfully incorporate one into the other, you need to work from the foundations of each and incorporate the foundations of the two disciplines together.	I'm going to be a field engineer construction wastewater treatment facilities and water conveyance infrastructure when I get out of college and I feel that I have to be quick to recognize work that is inadequate or where corners have been cut.	I will apply this technique while conducting a group discussion to collect all related aspects or views about my planned infrastructure or programme or solution.

Results – Machine Assessment of Student Responses

With advances in technology and the availability of computing devices, automated assessment generated by machine algorithms is becoming increasingly tenable as a tool for assessment and supporting timely student learning. One downside of the human assessment process is that it is slow, which could be addressed by automation. Machine methods have been applied to teaching writing [17], scoring of essays [18] [19], and science education [20], as some examples. Since we have a human-classified set of examples, this task can be formulated as a supervised learning problem. A Bernoulli Naïve Bayes classifier is a particularly simple means of learning from examples—it simply calculates a smoothed probability of occurrence of each feature in each category, then uses these probabilities as weights to classify new examples. In this case, examples are sentences, and features are the individual words used in those sentences.

As in Table 3 (minus the consensus phases), we trained two classifiers: one for the Belief, Effect, and How categories, and another for the Improve World, Reflection, Self-Improvement, and Teamwork subcategories (dropping the Other subcategory because it was so rare). We trained these classifiers on 70% of the students such that we could assess the classifiers' performance on the remaining 30%. For more stable results, we repeated this splitting, training, and testing process 100 times, and averaged the testing accuracies. To extract features from text to be fed into the classifiers, we simply broke all texts into words (unigrams; using the lingmatch package [21]), dropping function (stop) words and those appearing in fewer than 9 sentences within each training set, then coded sentences as containing each word or not, resulting in a binary sentence by word matrix.

The overall accuracies (average sentence-classifications matching humans in the testing set) of these simple classifiers were 53% for the Belief, Effect, and How categories, and 46% for the How subcategories. Tables 6 and 7 break these results down by category in the form of a confusion matrix, where rows (labeled "Actual") present the human classifications, and columns (labeled "Predicted") present the machine classifications. The cells on the diagonal of the table show the proportion of match between humans and machine for each of the categories. Cells that are off the diagonal show the proportions of mismatches between humans and machine.

Looking along the diagonal in Table 6, the machine classifier was best at matching the Belief (57% accurate) and How (56%) categories, and worst at matching the Effect category (25%). Looking along the diagonal in Table 7, the machine classifier was best at matching the Self-Improvement category (55%), followed by the Reflection (42%), Improve World (34%), and Teamwork (24%) categories. As stated earlier, the machine classifications are based on smoothed probabilities to the words in student responses, and the words function as predictors for the classifications. Table 8 shows the twenty highest-weighted words for the Belief, Effect, and How categories, and Table 9 shows the twenty highest-weighted words for the How subcategories.

Table 6. Confusion Matrix Showing Proportions of Category Match for All Sentences

		Predicted		
Actual		Belief	Effect	How
	Belief	.5707	.2276	.2572
	Effect	.1554	.2493	.1812
	How	.2740	.5231	.5616
	Note. Values are averaged over 100 resamples.			

Table 7. Confusion Matrix Showing Proportions of Category Match for *How* Sentences

		Predicted			
Actual		Improve World	Reflection	Self-Improvement	Teamwork
	Improve World	.3385	.1847	.1630	.1985
	Reflection	.1999	.4182	.1577	.2165
	Self-Improvement	.2653	.2409	.5516	.3429
	Teamwork	.1962	.1563	.1277	.2421
	Note. Values are averaged over 100 resamples.				

Table 8. Weights for Twenty Highest-Weighted Words Used by Bernoulli Naïve Bayes Classifier for All Sentences

Belief		Effect		How	
think	.193	life	.210	think	.192
life	.132	want	.200	life	.188
like	.092	think	.155	professional	.153
professional	.080	like	.135	apply	.142
people	.078	apply	.110	help	.142
work	.071	learned	.105	want	.119
help	.069	professional	.100	like	.104
way	.069	better	.090	learned	.100
engineering	.063	help	.090	able	.091
important	.061	future	.085	use	.088
feel	.059	feel	.080	people	.084
see	.055	make	.075	work	.073
engineer	.052	people	.075	make	.071
know	.050	work	.075	try	.071
different	.048	going	.070	see	.069
believe	.046	thinking	.070	better	.066
art	.044	way	.070	engineer	.064
engineers	.040	get	.065	engineering	.058
time	.040	personal	.065	personal	.055
personal	.038	use	.060	practice	.055

Table 9. Weights for Twenty Highest-Weighted Words Used by Bernoulli Naïve Bayes Classifier for *How* Sentences

Improve World		Reflection		Self-Improvement		Teamwork	
think	.176	think	.250	life	.296	work	.159
want	.176	help	.141	think	.216	people	.145
people	.165	life	.141	professional	.211	see	.145
professional	.154	able	.130	apply	.201	apply	.130
apply	.143	look	.109	help	.171	like	.130
learned	.132	thinking	.109	want	.151	make	.130
engineer	.121	try	.109	learned	.141	project	.130
ethical	.121	actions	.098	like	.141	able	.116
life	.121	problem	.098	use	.141	discussion	.116
like	.121	problems	.098	able	.090	help	.116
help	.110	project	.098	personal	.085	professional	.116
better	.099	solutions	.098	art	.075	think	.116
environmental	.099	way	.098	better	.075	try	.101

engineering	.088	engineer	.087	different	.070	communication	.087
going	.088	professional	.087	future	.070	get	.087
working	.088	solution	.087	see	.070	perspectives	.087
world	.088	decision	.076	get	.065	better	.072
consider	.077	engineering	.076	make	.065	change	.072
make	.077	picture	.076	engineering	.060	feel	.072
order	.077	reflection	.076	new	.060	going	.072

Discussion

Academic preparation in engineering has traditionally had a strong focus on technical skills. In recent years there has been a unified and coherent effort to broaden and diversify the professional preparation of engineering students, in part by incorporating the arts and humanities into the engineering curriculum. Assessing intellectual, affective, and behavioral gains in these integrated courses may be particularly challenging, given the qualitative and open-ended responses that may be the primary academic products in and outside the classroom. The present study involved the assessment of students' responses to one of several prompts assigned at the conclusion of each of the fourteen course topics: *Reflect on how you might want to apply what you learned to your development as a professional and/or to your daily life*.

The present analyses provide two methods to assess the topics expressed in students' reflections across the span of the course. By breaking assignments down into sentences, and classifying those sentences into Belief, Effect, and How categories, we see that students mostly focused on the How and Belief aspects of what they learned (see Figure 1 and Table 4). This suggests that students were inclined to motivate prospective actions by drawing on background beliefs and knowledge, rather than simply proposing an action. By further dividing the How sentences into Improve World, Reflection, Self-Improvement, and Teamwork subcategories, we see that students mostly focused on the Self-Improvement aspect of how they might apply what they learned. Applications of Reflection and Teamwork provided additional ways in which students proposed to apply what they learned in the course, as well as through an increased consideration of social, cultural, environmental, and ethical issues (i.e., Improve World) (see Figure 2 and Table 5). Importantly, this method of assessment allows one to gauge the nature and scope of personal and professional gains afforded by the course.

Applying AI machine methods to open-ended responses represents a novel exploration of assessment methods for the engineering classroom. The modest capacity of the machine method to match the human classifications of How, Effect, and Belief responses is encouraging, but leaves ample room for improvement. An examination of the table of the highest-weighted words (Table 8) provides some insight. Although weights differ somewhat, many of the same predictors are associated with the three classifications (How, Effect, Belief). This suggests that students relied on many of the same concepts to address the three types of reflections extracted by the human reviewers. It may be that the outcome does not reflect a weakness in the machine method, but rather involves a somewhat complex relationship between a) the rhetorical demands of addressing the question, b) the classifications imposed by the human reviewers, and c) the machine's attempt to match the human classifications against the students' data.

Though the automated version of the human assessment does not provide the same opportunity to get a feel for how students might be incorporating the course topics into their thinking, a machine classifier can give an overview of the topics covered by student, and it can do so very quickly, making such free-response questions tractable for smaller teaching teams and larger classrooms. Inspecting the highest-weighted words for each category (Tables 8 and 9) can also provide some insight into what the human classifiers may have meant by them. Some of the words in the How subcategories (Table 9) are particularly intuitive, such as *ethical*, *people*, and *environmental* in the Improve World category, and *discussion*, *work*, and *perspectives* in the Teamwork category.

The machine classifiers applied here achieved only modest accuracies, but there is a great deal of room for improvement, either by refining the numerical representations of the texts, or by tuning the classifiers or applying more complex classification algorithms. This improvement process is likely to be sensitive to particularities of the context in which the classifier is to be applied, so it would be best done as part of the development of a course and its assignments.

Identifying and quantifying the topics in students' reflections through the assessment methods in the present paper contributes to the engineering education literature in several ways: 1) It provides insight into the content of those reflections and allows one to gauge the extent to which the course prospectively prepares students to integrate technical, ethical, social, and cultural considerations into their professional practice. 2) It models probing students' reflections over the course of the semester in a consistent manner. Although not implemented here, this method of continuous, consistent evaluation has the potential of identifying change in students' thinking over the course of the semester, particularly if (unlike in this course) topics progressively build off of one another. 3) This work implements a standard method of human analysis of qualitative data and demonstrates how a machine classifier could be used to automate that process.

This study provides findings of interest and value to engineering educators in several ways. Incorporating liberal arts into engineering provides new possibilities for engineering pedagogy. The analyses here provide a glimpse into how these pedagogical innovations can shape students' thinking in directions that are more considerate of their contributions to their profession and society. If machine-based methods of text analysis can be further improved in terms of accessibility and flexibility, they would be even better-suited to provide instructors with additional resources for assessing instruction and providing students with timely feedback.

Conclusions

The human information-processing system has limited resources [22], both for immediate attention [23] and longer-term storage [24]. From that perspective, students will not process and retain all information from a course. Importantly, students' course experiences, knowledge gain, and behavioral change may only dimly reflect intended course objectives. This point is made to emphasize the relevance and importance of the type of extensive data gathering and qualitative analyses we conducted in this course that crosses over the typical technical content of engineering courses.

The work of Kamp [1] and others [2] [3] set the stage for a new direction in engineering education—one that looks to a globalized, technological, and highly interconnected workplace that respects culture, environment, economies, and diversity. The present project attempts to move in that direction by developing and assessing a graduate environmental engineering course that draws on the arts and humanities.

Engineering students have the power to change the world. According to Kamp [1], engineering education has a role to play in this: *"Engineering education must empower the students to be leaders of innovation and to be able not only to adapt to the uncertain world and changing work environment, but also to change that world"* (p. 89). The present project takes a step in that direction.

Acknowledgement

This material is based on work partially supported by the National Science Foundation under Grant No. 1806889. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] A Kamp, "*Navigating the landscape of higher engineering education: Coping with decades of accelerating change ahead,"* Delft University of Technology and 4TU.Centre for Engineering Education, 2020. [Online]. Available: www.4tu.nl/cee/publications/navigating-the-landscape-of-higher-engineering-education-4tu.cee-web-def.pdf.

[2] B. Li, R. G. Ryan, N. Warter-Perez, Y. Gan, H. Mustafa, H. Cox, and L. Ding, "Vertical integration of the liberal arts in engineering education," *Proceedings of the American Association of Engineering Education Annual Conference and Exposition,* 2016.

[3] A. Wood and R. Martello, "Work in progress: Liberal arts help engineering students change the world," *Proceedings of the American Association of Engineering Education Annual Conference and Exposition,* 2020.

[4] D. A. Schön, *The reflective practitioner: How professionals think in action*, New York, NY: Basic Books, 1983.

[5] G. Bolton and R. Delderfield, *Reflective practice: Writing and professional development*, Thousand Oaks, CA: Sage, 2018.

[6] J. H. Kim, J. A. Morrison, and E. Ramzinski, "Is Bildung possible in the classroom?: Autobiographical writing as philosophical exercise (askēsis) for developing one's Bildung, *Journal of Curriculum and Pedagogy*, vol. 16, no. 3, pp. 242-262, 2019.

[7] A. Cordner and P. Brown, "Moments of uncertainty: Ethical considerations and emerging contaminants," *Sociological Forum*, vol. 28, no. 3, pp. 469-494, 2013.

[8] M. Denchak, "Flint water crisis: Everything you need to know," *Natural Resource Defense Council,* 2018. [Online]. Available: www.nrdc.org/stories/flint-water-crisis-everything-you-need-know.

[9] D. Herman, *Basic elements of narrative*, John Wiley & Sons, 2009.

[10] M. Hiner, *Paper engineering for pop-up books and cards*. Tarquin Publications, 1985.

[11] G. Date and S. Chandrasekharan, “Beyond efficiency: Engineering for sustainability requires solving for pattern,” *Engineering Studies*, vol. 10, no. 1, pp. 12-37, 2018.
[12] M. Foucault, *The hermeneutics of the subject: Lectures at the College de France, 1981-1984* (G. Burchell Trans.), New York, NY: Macmillan, 2005.
[13] V. Braun and V. Clarke, “Using thematic analysis in psychology,” *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, 2006.
[14] R Core Team, *R: A language and environment for statistical computing*, Vienna, Austria: R Foundation for Statistical Computing, 2021. [Online]. Available: r-project.org.
[15] D. Bates, M. Maechler, B. Bolker, and S. Walker, “Fitting linear mixed effects models using lme4,” *Journal of Statistical Software,* vol. 67, pp. 1-48, 2015. R package version 1.1-27.
[16] M. G. Kenward and J. H. Roger, "An improved approximation to the precision of fixed effects from restricted maximum likelihood," *Computational Statistics & Data Analysis*, vol. 53, no. 7, pp. 2583-2595, 2009.
[17] M. D. Shermis and J. Burstein (Eds.), *Handbook of automated essay evaluation: Current applications and new directions*, New York: Routledge, 2013.
[18] D. Yan, A. A. Rupp, and P. W. Foltz (Eds.), *Handbook of automated scoring: Theory into practice*, CRC Press, 2020.
[19] E. Mayfield and C. P. Rosé, “LightSIDE: Open source machine learning for text accessible to non-experts,” in *Handbook of automated essay evaluation: Current applications and new directions*, M. D. Shermis and J. Burstein (Eds.), New York: Routledge, 2013, pp. 124-135.
[20] E. P. Beggrow, M. Ha, R. H. Nehm, D. Pearl, and W. J. Boone, "Assessing scientific practices using machine-learning methods: How closely do they match clinical interview performance?," *Journal of Science education and Technology*, vol. 23, no. 1, pp. 160-182, 2014.
[21] M. Iserman, “*lingmatch: Linguistic Matching and Accommodation*,” R package version 1.0.2., 2021. [Online]. Available: cran.r-project.org/package=lingmatch.
[22] H. A. Simon, "Invariants of human behavior," *Annual review of psychology*, vol. 41, no. 1, pp. 1-20, 1990.
[23] N. Cowan, "The magical mystery four: How is working memory capacity limited, and why?," *Current directions in psychological science*, vol. 19, no. 1, pp. 51-57, 2010.
[24] D. L. Schacter, "The seven sins of memory: insights from psychology and cognitive neuroscience," *American psychologist*, vol. 54, no. 3, pp. 182-203, 1999.