Work in Progress: Quantification of Learning through Learning Statements and Text Mining

Mr. Jackson Lyall Autrey, University of Oklahoma

Jackson L. Autrey is a Master of Science student in Mechanical Engineering at the University of Oklahoma from Tulsa, Oklahoma. He holds a Bachelor of Science in Mechanical Engineering from the University of Oklahoma and currently is involved with research into design-based engineering education. After completion of his Master’s degree, Jackson plans to pursue a Ph.D. in Mechanical Engineering.

Jennifer M. Sieber, University of Oklahoma

Jennifer M. Sieber recently graduated with a Master of Science degree in Data Science and Analytics from the University of Oklahoma. She is currently employed full time as a Data Scientist. Her research interests include text mining, machine learning, and data analytics.

Prof. Zahed Siddique, University of Oklahoma

Zahed Siddique is a Professor of Mechanical Engineering at the School of Aerospace and Mechanical Engineering of University of Oklahoma. His research interest include product family design, advanced material and engineering education. He is interested in motivation of engineering students, peer-to-peer learning, flat learning environments, technology assisted engineering education and experiential learning. He is the coordinator of the industry sponsored capstone from at his school and is the advisor of OU’s FSAE team.

Prof. Farrokh Mistree, University of Oklahoma

Farrokh’s passion is to have fun in providing an opportunity for highly motivated and talented people to learn how to define and achieve their dreams.

Farrokh Mistree holds the L. A. Comp Chair in the School of Aerospace and Mechanical Engineering at the University of Oklahoma in Norman, Oklahoma. Prior to this position, he was the Associate Chair of the Woodruff School of Mechanical Engineering at Georgia Tech – Savannah. He was also the Founding Director of the Systems Realization Laboratory at Georgia Tech.

Farrokh’s current research focus is model-based realization of complex systems by managing uncertainty and complexity. The key question he is investigating is what are the principles underlying rapid and robust concept exploration when the analysis models are incomplete and possibly inaccurate? His quest for answers to the key question are anchored in three projects, namely,

Integrated Realization of Robust, Resilient and Flexible Networks

Integrated Realization of Engineered Materials and Products

Managing Organized and Disorganized Complexity: Exploration of the Solution Space

His current education focus is on creating and implementing, in partnership with industry, a curriculum for educating strategic engineers—those who have developed the competencies to create value through the realization of complex engineered systems.

Email <farrokh.mistree@ou.edu> URL http://www.ou.edu/content/coe/ame/people/amefaculty/mistree.html
LinkedIn http://www.linkedin.com/pub/farrokh-mistree/9/838/8ba

©American Society for Engineering Education, 2017
Abstract

How can instructors leverage assessment instruments in design, build, and test courses to simultaneously improve student outcomes and assess student learning well enough to improve courses for future students? We recognize the need for a framework for shaping and assessing student learning. In that spirit, we contend that in design, build, and test courses students learn when they are required to reflect on their experiences and identify their learning explicitly. Further, we posit that utilization of an assessment instrument, the learning statement (LS), can be used to both enable and assess student learning.

In our course, AME4163: Principles of Engineering Design, a senior-level, pre-capstone, engineering design course, students learn by reflecting on doing by writing statements anchored in Kolb’s experiential learning cycle. In Fall 2016 we collected over 11,000 learning statements from over 150 students. To address the challenge of analyzing and gleaning knowledge from the large number of learning statements we resorted to text mining to analyze student-submitted learning statements.

We assert that text mining empowers instructors to better assess student learning in design, build, and test courses. Utilizing current algorithms for identifying writing patterns, we form a picture of learning over the course of the semester. Employing this tool, we can quantify student learning through reflection on doing and improve the course for future offerings. We find that student insight is characterized by focusing on the future utility of learning, particularly in areas such as planning a design process and evaluating design concepts. In this paper, we cover the salient features of the course, the learning statements, text mining and initial findings.
1. Frame of reference

In our pre-capstone course, AME4163 Principles of Engineering Design, our primary goals are to enable students to internalize five Principles of Engineering Design (POED) that are relevant to a course that involves the designing, building and testing of a mechanism and hit the ground running as Junior Engineers upon graduation. We distinguish Junior Engineers from students by their ability to learn by reflecting on doing and thereby continuously learn from new experiences. Our goals arise from an emerging need for engineers able to manage and propose solutions to complex problems using tools which do not yet exist. We contend that current evaluation methods used by instructors in design, build, and test courses, which often focus on project output, do not adequately allow instructors to assess student internalization of learning.

1.1 AME4163 course structure and features

We require senior-level undergraduate mechanical engineering students to undergo an authentic, immersive, design, build, and test experience in the semester before their senior design capstone. During this process, we treat them as Junior Engineers and therefore expect them to be active participants in their learning. In AME4163, students form teams to complete course assignments which map to the steps of a structured design project. In this paper, we focus on data collected from the first five assignments in the course. We have designed the assignments in such a way that each represents a step of the design process and thus targets particular POED. This course structure is anchored in the experiential learning cycle of David Kolb [1] by the learning statement, a reflective learning exercise. We provide our course map of the relationship of the POED to each assignment addressed in this paper in Figure 1.

Through course lectures we provide the information and context required for students to complete assignments and, through reflection, identify competencies needed as Junior Engineers. Lecture topics range from discussion of the assignments and the POED to design process strategies and tools. For example, we give a lecture on ‘attention-directing tools,’ which enable students to make informed decisions based on qualitative data. As we observe in the course schedule in Figure 2, between assignments and course lectures we also require the students to undergo a Mid-term Design Review in early October and a Prototype Update in late October before the mechanism demonstration in early November. Finally, we provide students with the course booklet, which we have developed for the course as a central repository of essential information including goals for the course, scaffolded assignments and assignment evaluation rubrics.

Having provided the context and expectations, we challenged our Fall 2016 students to address “Project POP: Prospect or Perish,” an engineering design problem and contextual vignette from the work of Mistree, et al. [2]. In the story, citizens of the growing society of the fictional planet Vayu require an autonomous vehicle for navigating a hostile frontier, identifying subterranean resources, and beginning the drilling process. Our students must develop a vehicle to accomplish these goals on a simulated track of sand, marsh, gravel, humps, grease, and a 90° turn before deploying a piercing mechanism to pop a balloon covered by a layer of Styrofoam. We also subject the vehicle to the following technical constraints: the vehicles must run autonomously (save one ‘initializing’ action), weigh under two pounds, fit within a 1’ x 1’ x 1’ box, cost less than $75, and meet certain standards for safe operation.
<table>
<thead>
<tr>
<th>POED</th>
<th>Assignment Description</th>
<th>Assignment 1</th>
<th>Assignment 2</th>
<th>Assignment 3</th>
<th>Assignment 4</th>
<th>Assignment 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>POED</td>
<td>Forming a team</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1a</td>
<td>Accepting and executing a team contract</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1b</td>
<td>Understanding the problem</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1c</td>
<td>Proposing a plan of action</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>1d</td>
<td>Ideation: generating concepts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>2a</td>
<td>Developing concepts (ensure feasibility and realizability)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>2b</td>
<td>Evaluating concepts; identifying most likely to succeed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2c</td>
<td>Refining/modifying most likely to succeed concept</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3a</td>
<td>Stipulating a Bill of Materials</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3b</td>
<td>Ensuring functional and technical feasibility, safety, etc.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3c</td>
<td>Bill of Materials as built; understand all components</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4a</td>
<td>Ensuring built device meets performance requirements</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4b</td>
<td>Critical analysis of device; causes of success and failure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4c</td>
<td>Critically evaluating the design, build, and test process</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5a</td>
<td>Articulating internalized POED via learning statements</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5b</td>
<td>Carrying lessons to future: capstone and other ventures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5c</td>
<td>Post-mortem report</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1: Description of all POED and mapping each POED to the five assignments addressed in this paper.
1.2 The learning statement as a course instrument

Given that the students face an authentic challenge with physical, technical, and qualitative constraints, students now have a firm foundation for experiential learning, that is, learn by reflecting on doing. We seek to enhance this process by encouraging deliberate reflection. One form that this takes in the course is the ‘learning statement,’ hereafter referred to as ‘LS.’ The LS instrument is anchored in the work of David Kolb and has been explored by us in prior work [4], Turns [6], and Allen et al. [7]. In the LS, students articulate a lesson learned in terms of a specific experience and identify some future value or utility that the lesson may provide. Since we are interested in assessing a lesson learned and not just a feeling (I learned a lot) we insist that all LSs are structured as shown in Figure 3. In addition, we emphasize the importance of using the LS written during lectures as a means of growing into a Junior Engineer versus the assignment LS, which enable the internalization of POED.

Consequently, by having some experience, reflecting on it, articulating that learning explicitly, and then identifying the future utility of the lesson, students repeatedly revisit each step of Kolb’s learning cycle, aiding in the internalization of the POED. We require students to write LS both in class (individuals) and at the end of each assignment (both team and individual LS). It is our central contention in this paper that student LS serves both as an assessment tool for instructors and as a reflective exercise for the student to enable internalization of the POED.

<table>
<thead>
<tr>
<th>Experience x</th>
<th>Learning y</th>
<th>Value/Utility z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Through x (From x, By doing x, …)</td>
<td>I learned y</td>
<td></td>
</tr>
<tr>
<td>I did not consider x initially</td>
<td>I realized y</td>
<td>Value/Utility z in future of learning y</td>
</tr>
<tr>
<td>I thought (expected) x before/initially</td>
<td>I found out y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>I discovered y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>I became conscious of y</td>
<td></td>
</tr>
</tbody>
</table>

Value (Lectures) = Help you transition from a student to a junior engineer and gain insight into how to do the assignments
Value (Assignments) = Principles of Engineering Design

Figure 3: Structure of learning statements
In our prior work [4], we analyzed the LS submitted by students and teams using statistical methods anchored in a two-pronged evaluation method. First, each statement is categorized based on the LS subject matter (which POED the LS content is most closely associated with). Second, each statement is rated on a zero to three-point scale of ‘insightfulness.’ Thus, LS data are quantified by the number of statements in each POED category and each insightfulness rating. Insightfulness is quantified as follows:

1. **Zero points**: Statements earn a rating of zero if the LS is not written to conform to the structure illustrated in Figure 3.
   a. Example: “Projects tend to be extremely overwhelming when viewed in the holistic sense, but when a plan of attack is proposed that breaks down the project into smaller tasks, the project becomes more conceivable and therefore more manageable.” – AME4163 student, Fall 2016
   b. Regardless of a lesson expressed, the student fails to put the learning in the context of an experience and therefore is not a learning statement.

2. **One point**: Statements receive a rating of one point if the structure is present but the insight is trivial or obvious.
   a. Example: “Through Assignment 1, I have realized that communication protocols are crucial for a team to work together to complete a goal.” – AME4163 student, Fall 2016
   b. The student both states something obviously true and neglects to explore any deeper relevance that the learning might have.

3. **Two points**: Statements receive a rating of two points if in the LS the student demonstrates a connection between their learning and something not explicit to the experience such as a novel circumstance in which the lesson might be applied.
   a. Example: “Through considering the customer requirements in greater depth individually, this has taught us more about the entire breadth of the problem and what needs to be taken into account in producing a successful end product, this has a value of allowing us to tailor the device to the end customer more effectively.” – AME4163 student, Fall 2016
   b. The student expresses learning in terms of an experience and then connects that to a future scenario involving a later stage of the design process.

4. **Three points**: Statements merit a rating of three points if the student exhibits a deeper understanding of the lesson learned and relates its utility to a wider context. Additionally, statements which embody any of the Principles of Design merit this rating.
   a. Example: “By developing an assembly of our future device, I have learned that preparing a plan of action for assembling it piece by piece in a logical order by stepping through the functions will lead to a better resultant vehicle, more so than just assembling it without regard to the order in which it should be done, which will lead to fewer mistakes in the future when sizing and manufacturing parts and will save our team money and time by eliminating errors and allowing focus to be kept on the completion of the project.” – AME4163 student, Fall 2016
   b. The student draws connections while demonstrating a more generalizable lesson learned. The student takes the learning beyond the obvious and directly relates to the fourth POED, which involves manufacturability.
1.3 Study methods and paper outline

As stated in Section 1.2, to date our efforts to quantify learning via LS data have been primarily focused on identifying patterns in the ratings and categories of individual and team learning statements as they change between assignments. Through this approach, we have identified discrepancies between the learning embodied in the LS data and traditional measures of student success in design courses, such as project outputs. Specifically, we find no relationship between learning explored in the LS (subject matter or insight) and student performance in the course. In this paper, we make this discrepancy a central assumption and therefore do not compare LS data to project outputs.

Our collection of 10,000 LS in Fall 2016 has prompted us to look for a ‘big-data’ approach to analyze them. In this paper, we focus specifically on roughly 3,000 LS from one section of AME4163 and the method that we employ to analyze them is text mining. Specifically, we focus on algorithmic text clustering using the bisecting K-means method. Thus, we can see how students relate close ideas to one another and simultaneously identify in which areas students are most frequently describing learning. We seek in this paper to validate this method by comparing it to the established, two-pronged method of LS evaluation.

To facilitate text mining, we require a means of digitizing, storing, and accessing student LS data in a way that has not been necessary in our earlier work. Consequently, we have built a custom database to organize the data in a form readily analyzed by our text mining program. For this analysis, individual and team LS data from Assignments 1-5 (written by 76 senior-level, pre-capstone, mechanical engineering students) in Word were transferred to tab-separated variable (TSV) sheets along with associated information (anonymized author, insight rating, POED category, date submitted, et cetera), uploaded to the custom database, queried as data subsets, and analyzed using the custom-built text mining algorithm; we expand on this in Section 3. Making the text machine interpretable has allowed to tailor our analysis to address the following questions:

Primary: In the context of an engineering design, build, and test course anchored in Kolb’s experiential learning framework, how can text mining of learning statements be used to assess student learning over the course of a structured design project?

Question 1: What “doings” do students preferentially reflect on? Do we note differences between the LS submitted by teams versus individuals?

Question 2: Do students adequately internalize POED targeted by instructors on each assignment? If not, what learning, expressed in the student reflections (LS), can instructors assess from individual assignments?

Question 3: Do students whose reflections on doing are assessed to be more insightful focus on different “doings” than their peers? What can we infer by text mining the data about the ways that ‘insightful’ students write their LS?

We anticipate certain patterns to emerge, given our intentional effort to tie each assignment to successive POED over the course of the design process; see Figure 1. We are also interested in the degree to which the evolving focus in the course design is reflected in the student learning. Based on our prior work, we expect both team formation, concept generation,
and refining concepts (POED 1a, 2a, and 3a) to be the most well-explored areas in individual student LS and POED 1a and 1d (forming a plan of action) largely the focus for team LS.

In Section 2, we provide support for our pedagogical approach to the course and the analytical methods employed in our analysis of the LS data. In this section, we review literature relevant to experiential learning in engineering design courses, the framework for experiential learning embodied in the LS format, the clustering text mining approach used to analyze the data, and finally our conclusions regarding student learning in design, build, and test courses. In Section 3, we outline and describe the methods used to analyze the LS. Specifically, we describe how LS data are acquired and compiled, how the database is designed and how queries are performed, and finally how the text mining algorithm works to analyze the queried data subsets.

Having completed the data processing in Section 3, in Section 4 we summarize the work performed and analyze the results. We review the questions posed in this section and respond to them using the data presented. Finally, in Section 5 we offer some closing remarks. Specifically, we comment on whether the objectives outlined Section 1 are met and propose improvements that could be made. Finally, we identify a few contributions to research in engineering education.

2. Survey of relevant literature

In this section, we outline the literature used to support our pedagogical approach, the assessment instruments employed, the data analytics methods used, and the need for the analysis.

2.1 Pedagogical foundations of AME4163

Our course is anchored in the twin constructs of experiential learning as outlined by Kolb [1] and problem-based learning as outlined by Dym and co-authors [2]. In addition, we build on the work of others. For example, Mistree et al. [3] have provide a guide for project-based learning courses (including the problems and problem contexts themselves). In our previous work [4], we demonstrate that a discrepancy exists between the learning expressed by students in LS and the performance of students in the course as evaluated using conventional metrics such as design project output and course grades. Specifically, they recommend that instructors interested in assessing student learning in design, build, and test courses de-emphasize the importance of project output (design and construction of the device) and instead focus on learning explored in reflective exercises such as the LS. From Balmer [5] we adopt a model for iterative course improvement. Turns [6] illustrates how the incorporation of a learning essay enables instructors to provide direct feedback to engineering design students while also being used to modify the course for the future. Similarly, Allen et al. [7] demonstrate how the utilization of learning essays, particularly in a digital environment, enables students to become more “reflective learners” while also providing instructors the opportunity to understand student learning in an engineering design course.

We frame the course around competency-based education. Mistree [8] indicates that the primary competency needed by engineering graduates today is the ability to adapt; this is largely due to the way that engineering practice is changing rapidly due to technological innovation and globalization. Further, we have, in prior iterations of the course, structured our course around
competencies identified by ABET [9], Eggert [10], Lahidji [11], and others [12-15]. We consider the challenges in incorporating the project-based learning construct and possible solutions identified by Todd [16] and Etlinger [17].

2.2 Assessment in engineering design courses

In our prior work [4], we highlight two points, namely, they highlight the disparity between how instructors evaluate engineering design students and they outline how this disparity has created a need for improved assessment instruments. Smith et al. [18] highlight how student outcomes improve when design students are required to perform critical self-assessment. They report that increased confidence on the part of the students (expressed via surveys) reflects improved performance in a design, build and test course. Others, such as Besterfield-Sacre et al. [19] and Segers [20] highlight both the need for improved instruments to assess design students and the importance of student self-assessment in attainment of course competencies. However, neither utilize the LS as an instrument to assess student learning. Finally, Olds et al. [21] catalogue a variety of assessment instruments, including some examples of self-assessment for use by engineering educators.

2.3 Data gathering and analysis

Text mining is a subset of Data Mining and requires several standard preprocessing techniques to be implemented before analysis can occur. To begin, each statement needs to have the punctuation removed and all characters must be converted to lowercase. Then the stop words must be removed to eliminate the occurrence of words that are not useful to the analysis of this subject, but are merely frequently occurring connector words (such as “a”, “the” and “and”) in the English language. The choice of which stop words to remove has been widely studied and it is reported, to determine the key factors contributing to variance in the documents, base stop words must be removed together with subject matter words that occur frequently. Researchers have found sets of stop words that are widely used for specific text mining tasks, such as Choy who has determined which stop words are removable for Twitter analysis [22]. In this study, we use the English language stop word package ‘tm’ in the initial analysis and later a set of key words provided to the students that are specific to the learning statements are also removed to provide a more focused analysis on the factors contributing to student learning.

K-means clustering has been used to determine the words that are most similar to each other in terms of frequency based on comparative studies that have assessed K-means algorithm to outperform standard hierarchical clustering in determining the similarity of text. Steinbach et al. [23] show that a specific type of K-means, bisecting K-means, outperforms other clustering methods for text cluster analysis. Further, principal component analysis (PCA) is used to explore variance in the data, allowing us to identify variance in the individual data subsets. In this study eigenvalue decomposition is the chosen PCA method [24].

2.4 Text mining in education research

In the past, our desire to analyze the text of the LS resulted in the development of a two-pronged method for evaluation of the LS using POED categorization and the zero-three point ‘insightfulness’ scale. In this paper, we report on our use text mining to analyze the student writings. Within engineering education, Frasciello [25] outlines a model for text mining of
student writing samples in engineering courses. Outside the field of engineering Wu and Chen [26] and Kokensparger [27] highlight the growing suitability of text mining algorithms for exploring the subject matter of a set of student writing samples.

3. Data processing

In this section, we outline our approach to the database construction and data analysis approach used to address our research questions. Our database records are anonymized and stored locally (offline) to prevent possible issues regarding student data privacy laws such as FERPA. The data we describe in this section is collected from 77 senior-level mechanical engineering students comprising 19 teams who took our pre-capstone course AME4163 at the University of Oklahoma in Fall 2016.

3.1 Data conversion and compilation

First, we copy each LS, submitted in DOCX files, into a TSV file with the format of the student’s last name, first name, their school ID number, the LS itself, the POED rating, the grade rating, the assignment number, the date it was written, and the student’s team number. The TSV files are then anonymized by replacing student names and school ID numbers with a random string (unique to each student) and then loaded into the database where the information is sorted into tables corresponding to the information provided. Once the data is in the database, queries to subset the learning statements by assignment number, rating score, or other characteristic are used and the results are output to a directory storing text files containing only statements from those subsets. This directory is input into R for analysis. We illustrate this process in the activity diagram shown in Figure 4.

![Figure 4: UML Activity Diagram for file conversions involved in processing and compiling the learning statements to be used in analysis.](image)

3.2 Design of the database and query processing

The database is written in Java and SQLite. Using Java and html, we interact with the database using a locally hosted web interface that serves as the GUI. For LS written by individual students, the tables include one for the assignment, one for the student, and one table for the LS. We show the attributes that are stored in each of these tables in Figure 5. Each statement that is written must be unique to be placed in the database. This was done to ensure that students were writing a new statement for each assignment to reflect on their learning. For

---

1 We are keen to share this software with faculty interested in collaborating with us on writing papers and pursuing funded research opportunities.
LS written by the teams, only one table was used to hold that information, as there were not many statements written by the teams for each assignment and the queries would not benefit from indexing on the individual teams. We have designed the queries to subset the information into different assignments and with different rating values for the statements we intend to analyze. These queries are written in SQLite and provide only the plain text versions of each statement in that subset.

3.3 Exportation of words

Once the queries are performed to obtain all LS from a given assignment or with a specific rating value (or some other relevant query), the data subset are exported as a text file that contains only that cross section of the corpus of LS. In R, these sentences are parsed to remove all punctuation and stop words included in the R package ‘tm.’ The words are also converted to lowercase and are reduced to their stem words. Further, we are also able at this stage to filter out additional words which we recognize to be ubiquitous (or at least common) to the student LS (see Figure 3). These words include many of those built in to the structure of the LS described in Section 1.2.

3.4 Data analysis

Minimum cutoff values are set for the words to be considered frequent. To determine the cutoff values, the mean quantity across all assignments of the top one percent of frequent terms was taken as the number of terms we would select of the top frequent terms in each assignment. Within this top one percent, the quantity of the least frequent term was then rounded to the nearest factor of 5 and used as the cutoff for that subset. Though this is not as good as sensitivity, it does allow for standardization of the cutoff values across the assignments. Only one team statement is rated as zero, so to assess whether words which that statement contained overlap with higher-rated statements, the threshold is set to words occurring more than once.

For LS written by individuals, the minimum is set to 50. For the team LS, the minimum is set to 20, as there are fewer team statements than individual statements during this portion of the analysis where ratings are not taken into consideration while we analyze the differences between each assignment. For the analysis of the statements where ratings are taken into consideration, the cutoffs vary per rating score for the team statements due to the variance in the number of statements that were in each score group. Only one team statement is rated as zero, so to assess whether words which that statement contained overlap with higher rated statements, the threshold is set to words occurring more than once. Similarly, due to the low number of team statements rated as one and three, the cutoff value is set to 10. The cutoff for statements rated as two is set to 25. For individual statements, those rated as one and three have a cutoff of 30 and those rated as two have a cutoff of 100. There are no individual statements rated as zero. These differences in the cutoff values are taken to normalize for the large size difference in the different rating groups.

Another concern in the analysis of the data are which words we allow the algorithm to analyze; words such as ‘a,’ ‘the,’ or ‘and’ might needlessly cloud the results. The removal of the standard English stop words reduces the data considerably. The specific set of English stop words that are removed are defined in the R package ‘tm.’ In addition to these stop words, we
also remove the words that are found in the LS structure, except for the term ‘learned.’ These words are found in Figure 3. We choose to keep the words ‘learned’ and ‘concept’ due to their fundamental relevance to the learning in the course.

![Entity-Relationship Diagram](image)

**Figure 5:** Entity-Relationship diagram for individual learning statement data stored in the database.

Visualization plays a key role in understanding which words are relevant and which words are important for students to have reflected upon. We observe in the bar charts which terms are occurring most frequently in each cross section of the assignments. Additionally, the word clouds allow the most frequently occurring terms to stand out immediately. Finally, we use plots of the K-means clustering to determine the extent of how separated these terms are from one another in terms of their relative ‘importance’ to the students writing LS. During analysis, we find that using only two principal components allows us to consider more than ninety-five percent of the variability, so only two are used. Principal component analysis uses the eigenvectors and eigenvalues of the dataset of word frequencies to determine where the majority of the variance in our documents occurs [27]. For the purposes of this study, the words themselves become our dimensions and our goal is to reduce the number of dimensions to identify which words are not only most frequent in our whole corpus of learning statements, but also are weighted significantly given their relative frequency across individual documents. By using principal component analysis, we can find out where the majority of the variability occurs in the statements and can identify key words that correspond to the different ratings or assignments.

### 4. Results and discussion

These results include both a POED breakdown of the team and individual LS using our conventional approach, text mining results for all LS from all five assignments, text mining results for LS in each assignment, and text mining results for LS from the five assignments broken up by each statement’s rating.

In Figure 6(a), we plot the most frequently used words (at least 150 uses) across all assignments for LS written by individuals. The word ‘design’ is the most common word used in reflection throughout the course, followed by ‘learned’ and ‘team.’ The strong representation of
‘design’ and ‘learned’ are unsurprising given that ‘design’ is the focal point of all subjects in the course and ‘learned’ is one of the main words built into the LS structure. The high frequency of ‘team’ suggests students are cognizant of the importance of teaming to design. Additionally, in Figure 6(b), we show the cluster analysis of those same terms. We observe that ‘design’ is not only the most frequently used term, but its representation among the statements is high enough to make it the centroid term for its cluster, pulling in only the next nearest term ‘learned’ and diagramming that those two terms are closer in frequency to each other than the next nearest cluster’s terms. In Figure 6(c) we plot all the individual LS submitted in Assignments 1-5, organized by category. We observe that POED 1 (planning a design process) and POED 4 (prototyping, testing, and post-mortem analysis) represent the two largest areas of student focus, but with POED 2a (concept generation) the second largest individual category.

By comparing Figure 6(c) and the terms we highlight in Figure 6(a), we observe that the words most frequently cited overall seem to approximately map to the POED most frequently written about by individuals. We note the frequent use of the word ‘team,’ and the fact that POED 1, which deals with several issues pertinent to team formation, organization, and management, both represent areas frequently explored by individuals in the LS. Similarly, we observe that ‘concept(s)’ were extremely well represented in the writing samples. Those words could refer to several POED, but given that we observe in Figure 6(c) the prominence of POED 2a, which deals with concept generation, and POED 3a, which deals with concept modification, we find the frequency of ‘concept(s)’ unsurprising. These connections between the frequently used words and the LS POED breakdown lend credence to the text mining approach as we move forward with analysis of the LS broken down by assignment.

Figure 6: (a) Plot of terms taken from individual learning statements across all assignments that occurred more than 150 times and their relative frequency. (b) Cluster diagram of terms across all assignments for individual statements.
Whereas in Figure 6 we present an analysis of the individual LS from all five assignments together, we now seek to analyze those LS by individual assignment. We plot these results in Figure 7 and Figure 8, with Figure 7 used to represent the terms that occur a minimum of 50 times and their relative frequencies as a bar plot and Figure 8 showing the corresponding word clouds visually illustrating the relative frequencies of words occurring more than 50 times. From Figure 7, we observe that ‘team’ is the most frequently used word in Assignment 1 and ‘design’ is the most frequently occurring word in Assignments 2-5. As a progression throughout the semester, the focus for the individual students shifts away from their initial decisions of how they will work together as a team to complete their projects and towards the design of their project.

Throughout Assignments 2-5, we observe that design becomes more and more of a focus for the students. The key factors for Assignments 2-4, in which the students are focusing on their design, were the concepts that they are considering implementing, the ability for the student to identify what they are learning, the design process, and the materials they choose. Recalling our assignment-POED map in Figure 1, we observe that the text mining results appear to validate the assignment target POED. Though the text mining method does not enable us to identify word combinations particular to POED subcategories, we observe patterns in the word frequency which suggest that students are largely focusing on the main POED targets. In Assignment 5, the design is still a key factor, but we also observe that individual students begin reflecting on what is important, the analysis of their project, what they realize throughout, and the value of these experiences toward their future work.
Figure 7: Bar plots of most frequent terms and their relative frequencies for individual LS in Assignments 1-5 correspond to figures (a)-(e) respectively.
For the team learning statements, we analyze the terms across all assignments occurring with a minimum frequency of 50. We illustrate these findings in Figure 9 and Figure 10. From the K-means clustering in Figure 9(b), we observe that the key factors across all assignments for the teams are the design process and the teams (POED 3 and 1, respectively). All other factors for team learning are not as significant, as we note from the fact that only those two terms are in the first cluster. From Figure 10, we observe that the term ‘team’ is the most frequently occurring in Assignment 1 and that it remains a more prominent word as the semester progresses for the teams than for the individuals, as discussed before. During Assignments 2 and 3, the key factors for team learning are the team as a learning tool and the concepts that the teams used in the design process. In Assignment 5, the design process becomes the most important factor for team learning.
Having broken down the LS by assignment, we now seek to involve the LS ‘insight’ ratings described in Section 1. Separating LS based on their rating, we now employ the text mining approach used thus far for each set of team LS with ratings zero, one, two, and three. From Figure 11, we observe that for team statements rated as one, the words ‘team,’ ‘design,’
and ‘learned’ still occur more frequently than other terms, but now with addition of the word ‘concept.’ Interestingly, we observe that for statements rated as two, ‘team’ becomes an even more prominent word used in the LS while ‘learned’ has gone down substantially. In addition, we observe more words represented, a point true for statements rated three as well. This indicates that students who write about fewer subjects tend to be rated lower for LS insight. For statements rated as three, the term ‘future’ occurs frequently and this term is not seen in other team statements at lower ratings. We have omitted results for zero point LS due to a scarcity of data.

![Figure 11](image)

*Figure 11: Bar plots of most frequent terms and their relative frequencies from team statements for all assignments rated as 1-3 correspond to figures (a)-(c) respectively.*

From *Figure 12*, we observe that, for individual statements at all ratings, the terms ‘design,’ ‘learned,’ and ‘team’ do occur frequently, but not always the most frequently. For statements rated as one and two, those three terms are the most frequent, but for statements rated as three, the terms ‘team’ and ‘learned’ are overshadowed by ‘future’ and ‘project.’ The key factors for student success are tied to design, but the ability of the students to consider the future
implications of their work and to focus on the project are also factors contributing to student success.

Figure 12: Bar plots of most frequent terms and their relative frequencies from individual statements for all assignments rated as 1-3 correspond to figures a-c respectively.

5. Closing remarks

In Section 4 we identify words that correspond to different ratings. Specific words, such as ‘design,’ ‘team,’ and ‘learned’ occur frequently at all rating levels, but the word ‘future’ is only found frequently in LS rated as three for both individual students and teams. This finding suggests that students and teams who write highly-rated statements are relatively more likely to frame the utility of their learning using the word ‘future.’ This is indicative that these students having recognized the importance of hitting the road running as Junior Engineers. Further, students writing statements in this category are relatively less likely to discuss learning related to design ‘concept(s).’ Additionally, throughout the course, the design process becomes more important to most students and the term ‘design’ does not occur as frequently as it does in Assignment 1 as it does in the later assignments. Though the word ‘design’ has myriad uses in a design course (used to describe concepts, the process, as an action, or many other uses), one possible reason that it does not appear as frequently in earlier assignments is that students
unaccustomed to a structured design process may not immediately classify the team formation, problem definition, or early brainstorming phases (among others) as being ‘design’ phases of the project.

In Section 1.3 we pose one primary question and three secondary questions for investigation. We ask: *In the context of an engineering design, build, and test course anchored in Kolb’s experiential learning framework, how can text mining of learning statements be used to assess student learning over the course of a structured design project?* To answer this question, we address three associated secondary questions.

5.1 Critical review of our approach and findings

Before addressing primary and secondary questions, we first seek to validate the text mining approach employed. That is, how well can the algorithmic analysis of individual words and their relations to other words be used to categorize the subject matter of the work in question? We demonstrate this by comparing the cluster analysis and word frequency histograms for Assignments 1-5 with the histogram of LS categorized by their relevant POED (see Figure 6 for individual and Figure 9 for team LS). We determine that the relative frequencies of words associated with specific POED found by the text mining algorithm map well to the breakdown produced by manually categorizing each statement.

In Question 1, we ask, *what “doings” do students preferentially reflect on? Do we note differences between the LS submitted by teams versus individuals?* Our response is anchored in Figure 6. In the histogram of the most frequently used words by individuals over the course of the entire design process, we observe that ‘design’ is by far the most common term used by the students. In addition, we observe in the cluster diagram that ‘design’ is so ubiquitous in the students’ LS that it pulled several words such as ‘will’ and ‘concept’ into its cluster even though their frequency is closer to the words constituting the next largest cluster. Of interest also is the fact that, of individual LS, the words ‘concept’ and ‘concepts’ are used so frequently that, combined, their frequency only trails ‘design.’ This is unsurprising given that Assignments 2-4 explicitly refer to the students’ design work in terms of their ‘concept.’ Of further note is that ‘team’ is the fourth most frequently used word by individuals and, unsurprisingly, the most frequently used word by teams in their LS. From this we infer that over the course of the design project, students report that they learn a great deal about team organization and management, results which are consistent with our prior work with student LS. We therefore feel confident that we have satisfactorily resolved Question 1.

In Question 2, we widen the query used to answer Question 1 by asking, *do students adequately internalize POED targeted by instructors on each assignment? If not, what learning, expressed in the student reflections (LS), can instructors assess from individual assignments?* In Figure 7(a), we note that individual LS in Assignment 1 largely focus on team formation and management, which are tied to POED 1. We observe here that ‘team’ is the most frequently word used, a notable exception to Assignments 2-5, in which ‘design’ is once again the word most frequently used. This is consistent with our goals for students completing Assignment 1, which focus on two areas: forming and organizing the team and understanding the design problem. Interestingly, we observe from Figures 6(a-d) that ‘concept’ and ‘concepts’ in
Assignment 2, ‘device’ and ‘process’ in Assignment 3, ‘concept/concepts’ and ‘materials’ in Assignment 4, and ‘materials,’ ‘concepts,’ and ‘components’ in Assignment 5 are all the most frequently used words after ‘design.’ Given that, in Assignment 2, students generate concepts, in Assignment 3 they refine them into a primary design concept, in Assignment 4 they develop CAD models and plan the prototype construction, and in Assignment 5 they perform a post-mortem on the device as-built, we observe in the results evidence that students are internalizing the POED as intended. We note also the fact that students around Assignment 3 begin to more readily discuss their learning in the context of a design process, indicating that by the stage of concept refinement the value of the structured approach is becoming evident to the students. Overall, while we are confident that, within the scope of this paper, the text mining approach has enabled us to satisfactorily resolve Question 2, in the future, we will likely focus on further analyzing the text mining results and developing a more rigorous framework for tying word frequency to a POED.

In Question 3, we ask, do students whose reflections on doing are assessed to be more insightful focus on different “doings” than their peers? What can we infer by text mining the data about the ways that ‘insightful’ students write their LS? We perform the original analysis (analysis of LS over the entire design project) but this time controlling for the ratings of the LS. As a reminder, we evaluate each LS according to a rubric discussed in lecture and provided to the students for reference before the students are ever required to write. In our rubric, we rate each statement from zero to three points, with zero representing an improperly formatted statement, one representing ‘low-hanging’ fruit, two representing ‘medium’ insightfulness, and three representing ‘insightful.’ Controlling for the ratings, we observe in Figure 11 that, for team LS rated one, topics covered largely focus on ‘concept/concepts’ and ‘device.’ In contrast, we observe that students with statements in the three-point category largely focused on ‘team,’ ‘process,’ and ‘plan.’ From this we gather that students most fully internalize, by our standards, the non-technical aspects of the POED. Meanwhile, we observe that student LS in the two-point category also heavily focused on ‘concept/concepts’ as well as ‘materials’ and ‘process.’ We also note that there were too few ‘zero-point’ statements to perform meaningful analysis, which itself seems to indicate that the students have adequately internalized the ‘reflection on doing’ approach embodied in the LS format. We are excited to observe that the text mining results broken down by rating have allowed us to reframe our understanding of the LS data in a way not possible in the past. However, to more rigorously address Question 3, in the future we may need to develop a ratings scale with higher fidelity (rather than a simple four-point scale), so that we can organize the data in greater detail.

5.2 Contributions and way forward

In this paper, we provide evidence for our interpretation of the process of student learning in our course, AME4163. In our prior work [4], we note that LS dealing with POED 1 (1a-forming a team, 1b-implementing a team contract) are heavily featured in Assignment 1 and, to a lesser degree, Assignment 2. Further, they highlight how concept evaluation and refinement, POED 3, become more prominent in Assignments 2-5. In addition, they note that, over the course of the entire design project, team formation, planning, and management dominate the subjects explored in the team LS. Our results from the text mining analysis provide additional support for these observations.
In addition, we observe in this paper notable underlying patterns that had not been noted in the past. We highlight the phenomenon by which students begin focusing on the process of design by Assignment 3. Whether this constitutes a similarity or change from previous years is a topic we plan to address next year. To that end we need to perform the same data analysis techniques employed here on the LS from the previous iteration of the course. Additionally, we are pleased to find that, in answering Question 2, we find that student internalization of the POED are occurring at the planned moments in the course, following closely with the subject matter outlined in Assignments 1-5.

One area in which we posit that this analysis might be useful as a complement to traditional forms of assessment used in engineering design courses such as grading. Frequently, students in such courses are evaluated based on how well they resolve the design challenge. We find this approach problematic for two reasons. First, we do not believe that the performance of a mechanism in a design, build and test project is an appropriate or adequate metric to assess student learning; the design artifact is an output, not an outcome. In fact, much of the insight expressed in our data set here is from students who produced mechanisms that performed only fairly or poorly. We posit that student learning should be the outcome which instructors should pay attention to. Second, we suggest that grading the design reports, the design mechanism and its performance does not capture what students have learned as a result of doing. Unlike the method explored in this paper, evaluation of the design mechanism does not facilitate an understanding of how students learn over the course of a design project, how insightful that learning is, nor how it changes in response to particular tasks (such as assignments).

In closing, we are confident that the use of text mining to interpret student LS in this work provides useful information regarding patterns in the student learning and internalization of the course material. We plan to improve this analytical tool, perhaps by automating the process of moving the student submitted LS into the developed database. Consequently, this may allow us to develop these software tools for use by other educators and researchers in their own work. In addition, we identify that the analysis of the LS by rating using text mining left us with additional questions. We would like to know if the ratings of statements were influenced by the POED the student chose to write about. Resolving this would allow us to see what POED students most struggle with internalizing. Additionally, we have not in this paper addressed the degree to which students are transitioning to Junior Engineers, one of our stated course aims. To resolve this, we will, in future work, use the text mining approach to analyze the LS present in the Semester Learning Essay and in the Capstone the following semester. Finally, we seek collaborators who are interested in publishing scholarly papers and pursuing funded research.

6. Acknowledgements

Jackson Autrey acknowledges the Graduate Teaching Assistantship from the School of Aerospace and Mechanical Engineering at the University of Oklahoma, Norman and the financial support granted through the Dolese Teaching Fellowship program. Jennifer Sieber acknowledges the Graduate Teaching and Research Assistantships from the School of Aerospace and Mechanical Engineering at the University of Oklahoma. Farrokh Mistree acknowledges the financial support that he received from the LA Comp Chair. Zahed Siddique acknowledges the
financial support that he received from the Dick and Shirley O’Shields Professorship in Mechanical Engineering.

7. References


