Elementary Student Reflections on Failure Within and Outside of the Engineering Design Process (Fundamental)

Dr. Pamela S. Lottero-Perdue, Towson University

Pamela S. Lottero-Perdue, Ph.D., is Associate Professor of Science Education in the Department of Physics, Astronomy & Geosciences at Towson University. She has a bachelor’s degree in mechanical engineering, worked briefly as a process engineer, and taught high school physics and pre-engineering. She has taught engineering and science to children in multiple informal settings. As a pre-service teacher educator, she includes engineering in her elementary and early childhood science methods courses, and has developed engineering education courses for middle school pre-service teachers and practicing elementary teachers. She has provided science and engineering professional development to multiple schools and school systems in Maryland, and has significantly contributed to the writing of many integrated STEM units of instruction used by teachers and school systems. Her research has examined factors that support and those that hinder elementary teachers as they learn to teach engineering, and currently focuses on how children learn to engineer and in the process, learn to fail and productively persist. She is the Director of the Integrated STEM Instructional Leadership (PreK-6) Post-Baccalaureate Certificate Program at TU. She currently serves as the Chair of the Pre-College Engineering Education Division of ASEE, and is a member of the ASEE Board of Directors Committee on P12 Engineering Education.
Elementary Student Reflections on Failure Within and Outside of the Engineering Design Process (Fundamental)

Introduction

In this qualitative study, I describe elementary students’ reflections on engineering design failure, their exposure to “fail words” (i.e., fail, failure, failed, etc.) in and out of school, and what they think it means to fail. This work draws from the idea that while design failure is a normal part of engineering practice, the concept of failure and the fail words themselves have very particular meanings in the context of engineering that are likely to be different than meanings these words and ideas take on in other contexts. The other contexts of interest in this study are those that are relevant to elementary students’ lives, including elementary schools, students’ homes, sports practice fields, and the television shows students watch. In what follows, I first address the research literature on this topic of design failure with respect to elementary engineering education. I then discuss the study context and participants, research questions, research approach and methods, findings, discussion and conclusions, and implications and future work.

Literature Review

The Normality of Failure within Engineering

Engineers design and analyze technologies via an engineering design process (EDP). EDPs typically include: defining the engineering problem via a problem statement or goal, constraints, and criteria; conducting background research about the problem and how others have tried to solve it; brainstorming multiple possible design ideas; selecting an idea to implement (a design) and creating a plan for that design; testing the design against criteria; analyzing test results to see where the design failed and succeeded against design criteria; planning improvements for the next design; and iterating, i.e., repeating parts of the design process to develop subsequent, and ideally improved, designs. “Design failure” occurs when designs fail to meet one or more criteria, and is an inherent part of the EDP. Design failure enables engineers to focus their improvement efforts on those aspects of the design that fail to meet criteria. In other words, engineers expect to learn from design failures.¹

A quote that captures the importance of failure within engineering design is from Henry Petroski:

Every successful design is the anticipation and obviation of failure, every new failure – no matter how seemingly benign – presents a further means towards a fuller understanding of how to achieve a fuller success.²

The idea here is that design failures are opportunities to learn how to improve designs in the iterative design process. In this way, design failures within EDPs are a means to an end, the end being a final product that meets or exceeds design criteria.
Engineering and Failure in Elementary Schools

According to the Committee on K12 Engineering Education, with support from the National Academy of Engineering and National Research Council, one principle of elementary through high school engineering education is that it should emphasize engineering design. Inherent in engineering design is that students have opportunities to “learn from failure and redesign,” one of six characteristics of high-quality STEM integration identified by Moore and colleagues. Further, engineering design knowledge is central to the Standards for the Preparation and Professional Development for Teachers of Engineering developed by Farmer, Klein-Gardner, and Nadelson. Two aspects of this engineering design knowledge for teachers to know and be able to teach are that engineering: “involves solving problems via an engineering design process (e.g., involving design under constraints, iterative design, optimization, improvement);” and “uses failure as a learning experience (e.g., when designed solutions fail, engineers learn from this failure and improve based on this new knowledge).” Additionally, elementary-level Engineering, Technology and Applications of Science (ETS) performance expectations within the Next Generation Science Standards (NGSS) involve students: including “specified criteria for success” as they go about defining problems, and planning and carrying out “fair tests in which variables are controlled and failure points are considered to identify aspects of a model or prototype that can be improved.”

Another principle for elementary through high school engineering education, according to the Committee on K12 Engineering Education, is that it promotes engineering habits of mind. Specifically, the committee referenced the following habits of mind: “systems thinking, collaboration, ethical considerations, creativity, communication and optimism.” Optimism “reflects a world view in which possibilities and opportunities can be found in every challenge and an understanding that every technology can be improved.” This is closely related to the idea of growth mindset from the research of Dweck and colleagues. Engineering education provides a fertile space for students to practice a growth mindset when they encounter design failures. Those who practice a growth mindset approach failures that arise in the learning process by analyzing the failure, trying again in a different way, and growing in knowledge and skill in the process. They learn to learn from failure; they practice optimism and perseverance. This is juxtaposed with a fixed mindset in which failure in the learning process signals thoughts such as “it’s not worth trying again since I’m just not good at this.”

While I and many others would argue that teaching engineering in elementary schools is worthwhile for students and society at large, its implementation is not a trivial matter. One of the challenges is that most elementary teachers have not had pre-service coursework or in-service professional learning experiences related to engineering education, and many elementary teachers lack self confidence or self efficacy with respect to teaching engineering. Another challenge has to do with the use of fail words and ideas about what failure means in the elementary context. What it means to fail in engineering is different than what it means to fail in education. In most elementary school environments the concept of failure and the fail words themselves have very negative connotations. A simple online search of “failing schools” will elucidate this point, bringing up, for example, a January 2017 article on “a new list of failing schools” in one U.S. State. This is but one example of many regarding how fail words can become negative labels in educational contexts. Because of the traditionally negative connotation
that fail words have, some in the maker movement have argued in blog posts that fail words should not be used in the classroom,\textsuperscript{15} or that fail word use must be considered more carefully before inserting slogans such as “fail early, fail often” or “fail forward” or “celebrate failure.”\textsuperscript{16}

\textit{Research on Students and Design Failure}

Research on design failure within elementary education limited, yet growing. Until recently, failure has been addressed as one topic of many others when researchers investigate students and teachers engaged in engineering design experiences. In these studies, there is some evidence how students generate failed designs, test designs to failure, conduct failure analysis and respond to design failure, or how teachers may respond to students whose designs fail.\textsuperscript{17-22} However, design failure has not been the focus of these studies.

In the last few years, researchers have been investigating failure in elementary engineering education more directly. The present study is an extension of my previous mixed methods and qualitative work that has focused on upper elementary teachers (teaching grades 3, 4, and 5), and how they and their students respond to design failure.\textsuperscript{9,23-25} The first of these studies examined elementary teachers’ perspectives on failure prior to teaching an engineering unit of instruction.\textsuperscript{9} This study, involving 254 survey respondents and 38 interviewees, found that most respondents (62\%) and interviewees (61\%) had a negative view of failure. For example, when asked what words or phrases come to mind when thinking of the word, failure, the highest response on surveys was “giving up” (40\%); the highest positive perspective on failure was “trying again” (21\%). Few teachers reported using fail words in their classrooms due to negative connotations of these words and past personal experiences with failure, and some were concerned that the students would identify themselves as failures when their designs failed or if they (teachers) used fail words.

Follow-up studies provided evidence that as teachers gained experience teaching engineering over the course of two years, most felt more comfortable teaching engineering, facilitating student teams whose designs failed, and using fail words.\textsuperscript{23,25} These studies also examined teachers’ reports of how students responded to design failures, as well as their reports of how they responded to students whose designs failed. Given the emphasis of the present study on students, I will reprint a table here that summarizes the range of student responses to design failures, as reported by teachers (Table 1). Note that these are divided into resilient, productive actions and positive emotions on the left side, and non-resilient, non-productive actions and negative emotions on the right.

Andrews has contributed to the elementary design failure conversation with her study of how students ideas evolve over time during design tasks.\textsuperscript{26} She found that there was little idea evolution with respect to a simple design tasks in which design success was easy to achieve; students were less able to identify why their designs succeeded, and were not pushed to alter their initial ideas. However, a design challenge that presented more opportunities for failure resulted in a more robust understanding of factors that influenced design success and failure. This points to the importance of design failure experiences in enhancing student learning and analytical thinking.
Table 1. Summary of students’ responses to design failure; from author and colleague’s previous studies, reprinted with permission.  

<table>
<thead>
<tr>
<th>Resilient, Productive Actions</th>
<th>Non-Resilient, Non-Productive Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledging design failure when it occurs</td>
<td>Denying that failure occurred by ignoring proper testing processes</td>
</tr>
<tr>
<td>Trying again</td>
<td>Giving up or losing interest</td>
</tr>
<tr>
<td>Engaging in failure analysis</td>
<td>Seeing the task as being too difficult</td>
</tr>
<tr>
<td>Focusing on improvement</td>
<td>Making changes to design without planning or thinking carefully</td>
</tr>
<tr>
<td>Working effectively as a team</td>
<td>Staying with the original failed design</td>
</tr>
<tr>
<td>Seeking help from peers and looking at other teams’ designs</td>
<td>Engaging in negative team dynamics</td>
</tr>
<tr>
<td>Using the EDP to guide next steps</td>
<td>Focusing on competition (worrying about performing less well than other teams)</td>
</tr>
<tr>
<td>Referencing background information to inform next steps</td>
<td>Ignoring background information that could inform next steps</td>
</tr>
<tr>
<td>Asking for help from the teacher</td>
<td>Seeking the “right answer” from the teacher</td>
</tr>
</tbody>
</table>

Both my own and Andrews’ work also touched upon the importance of having a reliable and informative testing process within engineering design challenges so that students can accurately interpret testing results, determine points of design failure, and subsequently improve in an informed way. One of my studies involved video analysis of a class learning an engineering design challenge. Students often performed design tests in an inaccurate, inconsistent way, leading to confusion and difficulty comparing first and second designs to one another. The associated conclusion was that students’ meaningful failure experiences are dependent upon them following proper testing procedures, criteria, and constraints. In Andrew’s study, one of the design tests involved trying to create an object out of simple materials that would hover for a period of time in a wind tunnel. The test often happened quickly (e.g., with the object flying out of the tunnel), making interpretation of what led to design failures difficult. A conclusion from her work was that “having [design challenge] tests that are straightforward to interpret should be a major consideration in creating design tasks.”

Study Context & Participants

Connections to the Parent Project: The E4 Project

This study is one part of a larger, multi-year project, the Exploring the Efficacy of Elementary Engineering (E4) Project, discussed in greater detail in my previous work. Pertinent to the present study, one aspect of the E4 Project was providing three days of professional development (PD) in the summer of 2013 to 135 teachers on the Engineering is
Elementary (EiE) curriculum. (Approximately 115 teachers received PD for a different curriculum, which is not a focus of the present study.) E4 Project teachers had not previously taught engineering, and were from a diverse cross-section of schools – with respect to ethnicity, socioeconomic status, and setting (i.e., rural, suburban, or urban) – across three states on in the eastern United States: Massachusetts, Maryland and North Carolina.

All 135 teachers received PD for the EiE unit, *To Get to the Other Side: Designing Bridges* (hereafter, the “bridge unit”).\(^{28}\) Depending upon the science units each teacher taught and included on their application forms for the E4 Project, they were also trained for one of four “assigned” EiE units that connected to one of those science units. Two of the assigned EiE units were: *A Stick in the Mud: Evaluating a Landscape* (hereafter, the “TarPul unit”); and *Thinking Inside the Box: Designing Plant Packages* (hereafter, the “plant package unit”).\(^{29,30}\)

In the 2013-2014 academic year, the EiE E4 Project teachers either taught: 1) the bridges unit, followed by their school system science unit and their assigned EiE unit; or 2) their school system science unit followed by their assigned EiE unit. Approximately half were in the first group, teaching the bridge unit, and half were not. In either case, the science unit was from each teacher’s science curriculum and provided science content relevant to the second EiE unit of study. Teachers were given the opportunity to attend a one-day refresher optional PD in the spring/summer of 2014, and then taught the units again in the 2014-2015 academic year. For both years of instruction, E4 Project leaders ensured that students in the teachers’ classrooms had not learned EiE or other engineering units via the school curriculum prior to learning engineering from E4 Project teachers.

Important to the present study is that neither the initial summer 2013 PD nor the optional follow-up 2014 PD explicitly focused on failure or design failure; rather, the PD focused on how to teach the units. Teachers referenced the teacher guides throughout the PD experience, and those teacher guides included fail words (Table 2). Note that of the three units, the bridge unit used fail words the most, and the plant package unit used them the least. Fail words were used as per the curriculum during PD instruction, and teachers were exposed to the broad idea that designs may fail, i.e., that they may not meet all or some of the criteria for the design challenge in a given unit. In the second, optional PD, teachers were asked to watch a video of students trying to figure out why their designs did not work well when tested, and then were encouraged to consider how the teacher in the video helped her students “persevere through failure.” This prompted a discussion about how to respond to students when their designs are not successful.

<table>
<thead>
<tr>
<th>Unit</th>
<th>Number of Fail Words in Teacher Guide</th>
<th>Number of Fail Words in Student Journal Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bridge</td>
<td>30</td>
<td>12</td>
</tr>
<tr>
<td>TarPul</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Plant Package</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

*Table 2.* Fail words used in EiE teacher guides for the bridge, TarPul and plant package units.
The E4 Project included the collection of data from both project teachers and students. Since students are the focus of this study, I will address student-related data collection here. Many forms of data were collected from the students in E4 Project teachers’ classrooms, including pre- and post-unit attitude surveys and engineering and science content assessments. Additionally, students’ EIE unit journals were collected and scanned. Altogether, approximately 8,500 students learned one or two EiE units for the E4 Project between 2013 and 2015.

A subset of E4 Project teachers were asked to have their classrooms be Classroom Intensive Observation (CIO) sites in which whole class video and focus group video were recorded each day of EiE unit instruction, and in which both teachers and student focus groups participated in interviews. See author’s previous work how teachers were selected to have their classrooms be a CIO site. There were one or two student focus groups in each site, with each focus group being a table or desk cluster of three to five students. Having either one or two focus groups in a classroom was dependent upon E4 Project staff availability. The particular group of students selected for focus group was dependent upon: 1) parental and student permission to be closely video recorded and interviewed; and 2) teacher discretion based upon what students would be willing to speak up and undistracted by the presence of cameras. Altogether, in the 2013-2014 year of the study, there were 16 EiE CIO sites; in 2014-2015, 12 of those 16 continued as CIO sites.

Participants

The present study set out to examine student reflections on failure, with the unit of analysis being student focus groups within CIO sites. I chose to include only second year CIO sites, given that by the second year of the E4 Project, teachers were more comfortable teaching the curriculum. I also wanted to select those teachers who had taught both the EiE bridge unit and one of the four aforementioned EiE units. Further, I aimed to select a subset of the four non-bridge units for analysis in this study; I plan to expand to all units in future publications. Ultimately, I chose to include in the study those focus groups of students who learned either of the following EiE unit combinations in the 2014-2015 academic year: 1) bridge and TarPul, and 2) bridge and plant package.

There were four classrooms who learned one of these two combinations: Ms. Lee’s class (Maryland), and Ms. Jefferson’s, Ms. Kraft’s, and Ms. Sheehan’s class (Massachusetts). These teacher names are pseudonyms. All four of the teachers participated in the 2013 PD, while just two – Ms. Lee and Ms. Jefferson – attended the optional 2013 PD. There were 29 total focus group students across the four classrooms. Maryland had one E4 Project staff member, and thus only had one focus group, while Massachusetts had multiple staff members to visit CIO sites and had two focus groups per class. In some cases, teachers split the focus groups into two teams who worked together on challenges. Table 3 shows the focus groups in each teacher’s classroom, as well as the relationship between focus groups and the teams of students who worked together during design challenges. Note that 11 teams in the study learned the bridge unit, 5 teams learned the TarPul unit after the bridge unit, and 6 teams learned the plant package unit after the bridge unit.
Table 3. Focus groups and teams in the present study. All names are pseudonyms.

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Unit(s)</th>
<th>Focus Groups</th>
<th>Teams</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lee</td>
<td>Bridge</td>
<td>Molly, David and Zora</td>
<td>Molly, David and Zora</td>
</tr>
<tr>
<td></td>
<td>TarPul</td>
<td>Molly, David and Erica*</td>
<td>Molly, David and Erica</td>
</tr>
<tr>
<td>Jefferson</td>
<td>Bridge and TarPul</td>
<td>Isabelle, Samar, Caroline and Gabriel</td>
<td>Isabelle and Samar</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Grant, Savannah, Arianna and Noah</td>
<td>Caroline and Gabriel</td>
</tr>
<tr>
<td>Sheehan</td>
<td>Bridge and Plant Package</td>
<td>Jocelyn, Brooke, Brennan, Kim</td>
<td>Jocelyn, Brooke, Brennan, Kim</td>
</tr>
<tr>
<td></td>
<td>Bridge</td>
<td>Spencer, Rachel, Olivia, Cole and Isaac</td>
<td>Spencer, Rachel, Olivia, Cole and Isaac</td>
</tr>
<tr>
<td></td>
<td>Plant Package</td>
<td>Spencer, Rachel, Olivia and Cole</td>
<td>Spencer, Rachel, Olivia and Cole</td>
</tr>
<tr>
<td>Kraft</td>
<td>Bridge</td>
<td>Kayla, Cassidy, Raja and Douglas</td>
<td>Cassidy and Douglas</td>
</tr>
<tr>
<td></td>
<td>Plant Package</td>
<td></td>
<td>Raja and Kayla</td>
</tr>
<tr>
<td></td>
<td>Bridge</td>
<td>Bethany, Brittany, Aiden, and Austin</td>
<td>Bethany and Aiden</td>
</tr>
<tr>
<td></td>
<td>Plant Package</td>
<td></td>
<td>Austin and Brittany</td>
</tr>
</tbody>
</table>

* Erica replaced Zora for the TarPul unit; Zora no longer wanted to be in the focus group.

Research Questions

In this study, I aim to answer the following research questions with respect to student teams and focus groups in Ms. Lee’s, Jefferson’s, Sheehan’s, and Kraft’s classrooms:

1. Under what circumstances do students identify design performance as a failure or as having failed, in whole or part?
2. In what contexts are students exposed to fail words in and outside of school?
3. What are students’ conceptions about what it means to fail?

Before sharing the answers to these questions, I turn to the methods used to address them.

Approach to Research and Methods

The research approach utilized for this study is primarily constructivist in nature. It is a worldview grounded in the belief that “individuals seek understanding of the world in which they live and work … [and] develop subjective meanings of their experiences.”31 With respect to this study as it contributes to my larger body of mixed methods, qualitative and quantitative work on failure, this constructivist approach is balanced with a bit of a pragmatism. The methods I have
chosen for a particular investigation have depended upon the needs, purposes, and constraints of the research context and units of analysis for that investigation. In this purely qualitative study, I explored in depth the ways in which students reflected upon their design challenge experiences, experiences with fail words, and conceptions of failure. These reflections were intended to be collected after students had some distance from the design challenge, and not in-the-moment as would be captured by video recordings of their participation whilst engineering. As described below, I looked to interview and journal data to examine students’ reflections of these design performance and failure topics.

Data Collection

The primary means of data collection for the present study was the use of video-recorded, semi-structured focus group interviews taken after each focus group had completed the bridge unit, and again, after each focus group had completed their second unit – either the TarPul or plant package unit. Three E4 Project staff members conducted interviews, which utilized an interview protocol containing questions about design performance, failure, and teamwork. I led protocol development, with input from the E4 Project leaders. Five interview topics were pertinent to the present study:

1. How each team’s first design (D1) and second design (D2) performed;
2. Improvements attempted or made between D1 and D2;
3. Whether or not the team thought that D1 or D2 failed in whole or part;
4. In what contexts (in and outside of school) have students heard fail words; and
5. Students’ ideas about what it means to fail.

Specific questions from the protocol for these sections of the interview can be found in the appendix of this paper. Altogether, the interview sections relevant to the present study constituted on average about 20 minutes of focus group interview time.

Students’ responses regarding their D1 and D2 design decisions and D1 and D2 performance were checked against scanned copies of their engineering journals. In most cases, the journals and interviews together provided clarity regarding design decisions and performance. In some cases, there was conflicting information between the two, and these instances were noted in the findings section.

Data Preparation

To begin the process of generating high-quality transcripts, I first sent audio from focus group interviews to a transcription service. I then reviewed each transcript from the service against each corresponding video recording. This enabled me to accurately match the speaker in the video to the speaker in the transcript, make corrections, clarify inaudible passages wherever possible, and add gestures to the transcripts. I gave each child from all 2014-2015 E4 Project focus groups unique pseudonyms, including those used in the present study.

I then set out to generate what I call a “design performance summary” for each team of students, distilling both interview data (addressing topics 1-3, above) and journal data to create a story line for each team regarding: the design decisions for D1, design performance of D1,
improvements attempted for D2, design performance of D2, and team member responses regarding whether and to what extent D1 and/or D2 failed. Altogether, I generated 22 DPSs.

Analysis

Design performance summary analysis aimed to address the first research question: Under what circumstances do students identify design performance as a failure or as having failed, in whole or part? Analysis of topics 4 and 5 of the interview addressed the second and third research questions: In what contexts are students exposed to fail words in and outside of school?; and What are students’ conceptions about what it means to fail?

I employed a general qualitative approach of iteratively coding for themes to answer research questions. The design performance summaries were hand-coded apart from other interview topics, going through three rounds of coding until I reached a refined and organized ultimate code set of 12 codes and 3 subcodes to describe students’ design performance and assessment of failure with regard to D1 and D2. While seven of these codes were gleaned more directly from interview responses and journal entries, five codes and the three subcodes emerged from the data without being specifically associated with an interview question. Additionally, one code prompted an additional review of interview transcripts.

The interview sections in which students responded to the direct questions about fail words they have heard in and out of school and about what it means to fail also went through three rounds of coding. The first was by hand, generating an initial set of codes. Those codes were then collapsed and reorganized for the second round of coding. In the third round, I generated a final set of codes as I coded the interviews in the HyperResearch coding program. Altogether, 39 codes were used to describe these data; some of those codes were combined in the writing sections of the finding sections that follow.

A caveat: Throughout the paper, I not only offer quotes and paraphrased passages to describe the range of student responses, I also provide percentages representing the frequencies of coded responses (e.g., 6 of 29 participants, or 21%). These percentages are only used to give readers a sense of response frequency among interviewees. They are not meant to be interpreted with any sort of statistical importance.

My Role

In addition to leading the development of interview protocols for this study, I was also the E4 Project member who collected CIO site video data, took fields notes, and conducted student and teacher interviews for Maryland. Over the two years of the project, I have observed three bridge units, three TarPul units, and one plant package unit in addition to other units not included in the present study. I observed all of Ms. Lee’s teaching of the bridges and TarPul units in the second year of the project, and, with respect to the present study, interviewed Molly, David, and Zora after the bridges unit and Molly, David, and Erica after the TarPul unit. Additionally, I assisted with the delivery of the summer 2013 PD, and led the optional, follow-up 2014 PD for Maryland. With respect to the Maryland study participants, I may be construed as a participant-observer – a familiar face in the classroom when I sat down to interview the students.
— with emphasis on observation role. With respect to the majority of study participants from Massachusetts, my role was purely as a researcher who had knowledge related to the units that the participants learned.

Although I have tried to be as objective as possible in interpreting students’ meaning making, I have surely brought bias to my role as the qualitative researcher for this study. I am an advocate for engineering education and see its promise, so I may be less likely to identify ways in which it — or its enactments — might be detrimental to students. I have placed importance on design failure within the EDP; importance that others may not similarly place. Also, while I have tried to represent students’ meaning-making accurately, I may have misinterpreted those meanings in my analysis.

Study Validity

I have implemented multiple validity strategies to strengthen this work. Protocol development was informed by my growing understanding of the topic of failure and how to ask students about their design and failure experiences. (I learned how not to ask students about failure in the first year of the study.) I reviewed transcripts thoroughly after initial transcription, checking them carefully against video recordings. I coded and re-coded, keeping a record of codes and their descriptions throughout the process, altering those as needed to better capture the data. I also present the full range of responses in the following section, including major and more minor themes. I compared student engineering journals to interview responses; I also checked team members’ journals against one another. In one case, this comparison process led to the exclusion of one team from analysis given the inconsistencies across these data sources. All cases in which there were inconsistencies have been noted in Tables 1-3 in the findings section. In other cases, I used a more tentative interpretation of students’ views on failure with respect to their designs.

All these strategies notwithstanding, there are ways in which the study validity could be improved. For purposes of triangulation, I could compare journals and interview responses against focus group video of the students as they moved through the EDP. (This is planned for future work, discussed in the final section of this paper.) I could employ peer debriefing as a means to cross check my analysis.

Findings

This section is organized into three subsections. In the first subsection, I summarize the design challenge for each unit, the performance of each team in the study who learned the unit, and the team’s assessment of whether or not and to what extent their first or second designs for that unit failed. The second examines students’ exposure to fail words in and out of school, and the third discusses students’ sense making around what it means to fail.

Unit Design Challenge Descriptions and Design Performance

In what follows, I share design challenge and team design performance summaries for the EiE bridges, TarPul, and plant package units, respectively. This is followed by cross-unit
findings regarding students’ assessment of the extent to which their designs failed, organized into four parts: overall failure, failing somewhat, no failure, and broad themes.

The Bridge Unit Design Challenge and Design Performance

Students were presented with the EiE unit bridge design challenge after testing the relative strength of beam, deep beam, and arch bridges made of index cards (Figure 1). The ends of the bridges rested on abutments that were about six inches apart. Students learned that deep beam bridges, thickened by a layer of accordion-folded index cards, and arch bridges were the strongest.

Figure 1. Three bridge types in the EiE bridge unit. Reprinted with permission.

The bridge unit design challenge involved using simple materials to design and construct a bridge that would span 15 inches from one 6-inch high abutment to another. The abutments were made of blocks or textbooks stacked on top of one another. Bridges were not allowed to be taped to the abutments, but could rest on them. Allowable materials were: craft sticks, copy paper, index cards, drinking straws (hereafter, straws), tape, paper clips and string. Each material had an assigned cost. Bridge criteria included barge passage, cost, stability and strength:

- **Barge Passage**: The 5” width of a 2” high model barge had to fit under the bridge without touching any parts of the bridge. (No score associated with this.)
- **Cost**: The higher the total cost of materials to build the bridge, the lower the cost score. For example, if a team spent $10.00 or more on the design, they earned a cost score of 1, and if they spent less than $4.00, they earned a cost score of 5. There were also cost scores of 2, 3, and 4 for values between $4.00 and $9.99.
• Stability: This was measured by how many times out of four trials a wind-up plastic toy car could roll across the bridge without falling off. The number of successful rolls was equivalent to the stability score; e.g., three successful rolls would be a stability score of 3.

• Strength: The strength score was determined by how many standardized weights (washers of a certain size) placed into a cup could be supported by the bridge span before the deck of the bridge dipped below the “failure line” – a line 2” below the top of the abutments (Figure 2). This should be tested in the middle of the widest unsupported part of the bridge deck, and not on top of a pier. If more than 100 weights could be held by the bridge, the strength score was 5; if fewer than 25 weights are held, the strength score was 1. There were also strength scores of 2, 3, and 4 for values between 25 and 99 weights.

The total score for the bridge is calculated by adding the cost, stability, and strength scores, with the maximum score being 14.

Figure 2. Failure line for the bridge in the EiE bridge unit. Reprinted with permission.

Table 4 summarizes the design performance of the 11 teams who learned the bridge unit, gleaned from interview and journal data. It also includes the teams’ interview responses to the questions: Do you think that D1 or D2 failed? and Did parts of D1 or D2 fail? Responses to these questions were coded as answers to the question: Do you think that your design failed? Answers included: yes, overall; somewhat; no; and no response. These answers are shown in Table 1 in bold italics, and are followed by any elaboration provided by team members from interview data.
Table 4. Bridge unit design performance and team responses regarding D1 and D2 failure; higher score means better design performance.

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Team</th>
<th>D1 Score</th>
<th>Do you think D1 failed?</th>
<th>D2 Score</th>
<th>Do you think D2 failed?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jefferson</td>
<td>Savannah and Grant</td>
<td>6</td>
<td><em>Yes, overall.</em> Savannah said: “Pretty much everything” didn’t work well.</td>
<td>10</td>
<td><em>Somewhat.</em> Savannah said: “Kind of.”</td>
</tr>
<tr>
<td></td>
<td>Gabriel and Caroline</td>
<td>10</td>
<td><em>Somewhat.</em> Agreed that certain parts failed.</td>
<td>13</td>
<td><em>Somewhat.</em> Agreed that certain parts failed.</td>
</tr>
<tr>
<td></td>
<td>Noah and Arianna</td>
<td>Unsure*</td>
<td><em>Yes, overall.</em> (Noah). “Everything failed” (Noah).</td>
<td>7</td>
<td><em>Somewhat.</em> Agreed that it “failed to do certain things – for weights” (Arianna).</td>
</tr>
<tr>
<td>Sheehan</td>
<td>Jocelyn, Brooke, Brennan and Kim</td>
<td>9</td>
<td><em>Somewhat.</em> First design failed with respect to strength.</td>
<td>11</td>
<td><em>No response.</em></td>
</tr>
<tr>
<td></td>
<td>Spencer, Rachel, Olivia, Isaac and Cole</td>
<td>7</td>
<td><em>Somewhat.</em> First design failed due to lack of strength and stability.</td>
<td>9</td>
<td><em>No response.</em></td>
</tr>
<tr>
<td>Kraft</td>
<td>Kayla and Raja</td>
<td>5</td>
<td><em>Yes, overall.</em></td>
<td>9</td>
<td><em>No response.</em></td>
</tr>
<tr>
<td></td>
<td>Austin and Brittany</td>
<td>9*</td>
<td><em>Yes, overall.</em></td>
<td>10</td>
<td>*Yes, overall (Brittany); <em>Somewhat</em> (Austin). Austin blamed weight failure on Brittany.</td>
</tr>
<tr>
<td></td>
<td>Cassidy and Douglas</td>
<td>9</td>
<td><em>Yes, overall.</em> Considered it a failure because it received a low weight score.</td>
<td>9</td>
<td>*Yes, overall (Douglas) or <em>somewhat</em> (Cassidy). Same issue with low weight score as for first design.</td>
</tr>
<tr>
<td>Jefferson</td>
<td>Isabelle and Samar</td>
<td>10 or 11*</td>
<td><em>Yes, overall</em> (Samar) “because it couldn’t hold 100 weights”; <em>Somewhat</em> (Isabelle).</td>
<td>8 or 9*</td>
<td><em>Somewhat</em> since “certain things failed” (Samar) and since could have spent less (Isabelle).</td>
</tr>
<tr>
<td>Lee</td>
<td>Molly, David and Zora</td>
<td>6</td>
<td><em>Yes, overall.</em> Bridge could not support itself. Molly said: “it could not support a straw.”</td>
<td>5</td>
<td><em>Yes, overall.</em> As with first, bridge could not support itself – even for “a millisecond” (Zora).</td>
</tr>
<tr>
<td>Kraft</td>
<td>Bethany and Aiden</td>
<td>8</td>
<td><em>Yes, overall.</em> Because the arches did not work.</td>
<td>10 or 7**</td>
<td><em>Yes, overall.</em> Because the arches did not work.</td>
</tr>
</tbody>
</table>

* Numbers are suspect here because of inconsistencies across interviews and journals or incomplete journals. ** The team miscalculated the cost score; they determined cost score based only on added materials for D2, not based on all materials used in D2. The overall D2 score according to the team was 10, an improvement from D1; in actuality, it should have been 7, a reduction in overall score.
The EIE TarPul unit involved the students making design decisions to identify and prepare a building site for a TarPul. A TarPul is like a bridge in that it allows people to cross over areas, such as rivers, which cannot be easily traversed (Figure 3). However, rather than walking across a bridge, people board a simple carriage that is elevated above the ground and connected to a pulley system. The cables of the pulley system stretch horizontally from a support pole on one side of the river to a support pole on another side of the river. Passengers board the TarPul on one side, and then while seated in the TarPul carriage, reach up to the pulley system and pull themselves from one side of the river to the other. Tarpuls are used in areas where bridges may not be affordable or otherwise feasible.

Figure 3. A sketch of a TarPul from the EiE TarPul unit.²⁹ Reprinted with permission.

In the TarPul unit, students read a story about a village that is considering installing a TarPul so that villagers can get to a medical facility across the river. The villagers would like for the TarPul to be placed right next to the village (i.e., the village site). However, there are other possible places to build along the river that are one, two, three or four sites away from the village. Each site along the river may have rocky or organic soil on either side, and sites are on
straight, somewhat curved or very curved portions of the river (see Figure 4 for field map). The sites may be “treated” by compacting them, providing a stronger base for the TarPul poles. Before the design challenge, students learned that: rocky soil is better at holding up TarPul poles than organic soil, and that compacted soil is better at supporting the TarPul than is uncompacted soil. They also learn that erosion of riverbanks, which is bad for maintaining the support of TarPul poles over periods of time, occurs at a faster rate for more curved sections of a river than for straighter sections.

Figure 4. Field map from the EiE TarPul unit. Reprinted with permission from the EiE team.

For the design challenge, students used a model TarPul system, i.e.: two 4-cup cylindrical containers of soil separated by about 12” of space, two ~10” dowels placed into the center of each of the containers, a string that stretches from the upper portion of one dowel to the upper portion of the other, and a paper cup that hangs from the stretched string to represent the carriage. For each design, teams considered all of the possible sites along the river, picked one, and determined if they will compact the soil and to what extent: not at all, slightly (pushed down by ¼”) or significantly (½”). Before testing how much weight the TarPul can hold, teams located two cups representing the type(s) of soil on each side of the river for that site, compacted the soil (if they planned to do so), placed the poles, and tested to see how many metal washers the paper cup carriage can hold before it dipped below the soil/water level. Figure 5 shows a model TarPul system with the soil/water level indicated.
The criteria scoring system for the TarPul challenge was as follows:

- **Location:** If the site chosen was 4 sites away from the village site, the location score was a 4. If it was 3 sites away, it was a 3, and so on.
- **Erosion:** If the site was on a very curved portion of the river, the erosion score was a 2; slightly curved was 0 points, and straight was -2 points.
- **Compaction:** There were two compaction scores; one per river side. If a side was compacted significantly, the compaction score was 2 points; slightly was 1 point; and not at all was 0 points.
- **Strength:** The strength score was determined by how many metal washers can be placed in the cup before the bottom of the cup dipped below the soil line, which represented the top of the river water. The minimum weight to carry one person in the TarPul was 11. If the TarPul holds between 0 and 10 weights, the strength and overall score for the TarPul was a “no score;” in effect, the TarPul was disqualified. If it held between 11 and 20 weights, the TarPul had a strength score of 4; if it held 51 or more weights, it had a score of 0. Scores of 3, 2, and 1 were also provided for holding 21 to 50 weights.

The goal for the TarPul design challenge was to have the lowest overall score possible, and to avoid having a “no score.”

Table 5 summarizes the design performance of the four of the five teams who learned the TarPul unit. The fifth team was not included in the table since the team’s design performance was unclear after analysis of interview data and team journals. As with Table 1, this table includes the teams’ interview responses to the questions: Do you think that D1 or D2 failed? and Did parts of D1 or D2 fail?
Table 5. TarPul unit design performance and team responses regarding D1 and D2 failure; lower score means better design performance, and “no score” means that it was ineligible for scoring.*

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Team</th>
<th>D1 Score</th>
<th>Do you think D1 failed?</th>
<th>D2 Score</th>
<th>Do you think D2 failed?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lee</td>
<td>Molly, David and Erica</td>
<td>no score</td>
<td>Yes, overall. Molly said: “Failed overall.”</td>
<td>12</td>
<td>No.</td>
</tr>
<tr>
<td>Jefferson</td>
<td>Samar and Isabelle</td>
<td>no score</td>
<td>Yes, overall. (Isabelle). Somewhat (Samar) since “a lot of parts failed.”</td>
<td>7</td>
<td>Yes, overall. (Isabelle). Somewhat (Samar) since “some parts failed.”</td>
</tr>
<tr>
<td></td>
<td>Arianna and Noah</td>
<td>10</td>
<td>Yes, overall (Noah) since “very curved, so … a big failure”; Somewhat (Arianna) due to curve.</td>
<td>5</td>
<td>No response.</td>
</tr>
<tr>
<td></td>
<td>Caroline and Gabriel</td>
<td>2**</td>
<td>Yes, overall. Because it only held one person.</td>
<td>3</td>
<td>No.</td>
</tr>
</tbody>
</table>

* Data for Savannah and Grant in Ms. Jefferson’s class are not included in this table. Their design performance was unclear from the interview and journal data. Journals were partially or fully incomplete, and conflicted with interview data. ** This score is somewhat suspect because of the soil type (organic) coupled with the lack of compaction; however, I include it here since this is what the team documented.

The Plant Package Unit Design Challenge and Design Performance

Prior to the plant package design challenge, students learned about plant needs (water, sunlight, air) and plant parts (roots, stem, leaves, flower). They also investigated a range of everyday packages, and learned that most packages address some or all of the following seven functions: containing the product, having a special means of carrying the product, communicating information about the product, displaying all or part of the product, preserving the product, protecting the product, and dispensing the product.

For the design challenge, students were challenged to design and create a package for plant (e.g., a daisy) that would sit on a store shelf and not need to be watered for a three- or four-day period of time. Teams could only use certain materials available to construct their packages. Each team needed to select one “base package”: either a 2-L clear soda bottle, or a 2-quart (half gallon) juice carton. They could also use manila folder cardstock, copy paper, transparency film, aluminum foil, plastic wrap, wax paper, paper towel, tape, straws, and cotton balls. Each item was assigned a cost per amount (e.g., $0.25 per cotton ball or 0.50 for a 12”x12” sheet of wax paper). Figure 6 shows two D1 plan drawings: Raja’s, made of a bottle base package; and Olivia’s, made of a juice carton base package.
The design performance of plant packages was assessed against seven criteria, including six product functions and cost. The six product functions were contain, carry, communicate, display, preserve and protect. (The function, dispense, was not relevant for the plant package.) The seventh criterion was cost. The criteria were evaluated in the following way:

- **Contain**: If the plant was held in the package, the contain score was 1; if not, it was 0.
- **Communicate**: If the package clearly communicated how to care for the plant (air, sunlight, water), the communicate score was 2 points; partial communication was a score of 1; and no information was a score of 0.
- **Carry**: If the package clearly shows or tells the consumer how to carry the plant, the carry score was 2 points; if there was some communication of this information, but it was confusing, the carry score was 1; if no information was provided, the carry score was 0.
- **Display**: If the plant could be seen, the display score was a 1; if not, it was a 0.
- **Protect**: Students conducted a shake test on their plant package, holding it upright and shaking it back and forth horizontally and somewhat vigorously 10 times. After the shake test, they evaluated whether the plant was: entirely protected, not tipped over (2 points); somewhat protected, partially tipped over (1 point); or damaged, tipped over (0 points).
- **Preserve**: Before the design challenge, students compared a dead plant and a healthy plant, and developed a 3 point rubric for plant health. For example, unhealthy plants may have brown/yellow, wilted or dry leaves (score of 1), while healthy plants are likely to have green, firm leaves (score of 3). Students used this rubric to evaluate plant health.
- **Cost**: The higher the cost of the materials, the lower the cost score. Classrooms had an option to use two different scales for cost. In either case, the lowest cost packages earned a cost score of 3 and utilized only the base package (no other materials). Somewhat high cost packages earned a cost score of 2, and high cost packages earned a score of 1.

The higher the overall plant package score, the better the design. The maximum number of points for a plant package was 14. See Table 6 for summaries of plant package unit design performance and team responses to whether or not their designs fully or partially failed.
Table 6. Plant package unit design performance and team responses regarding D1 and D2 failure; higher score means better design performance.

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Team</th>
<th>D1 Score</th>
<th>Do you think D1 failed?</th>
<th>D2 Score</th>
<th>Do you think D2 failed?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kraft</td>
<td>Raja and Douglas</td>
<td>11</td>
<td>No.</td>
<td>13</td>
<td>No.</td>
</tr>
<tr>
<td></td>
<td>Cassidy and Kayla</td>
<td>11</td>
<td>No.</td>
<td>13</td>
<td>No.</td>
</tr>
<tr>
<td></td>
<td>Bethany and Brittany</td>
<td>8</td>
<td>Yes, overall. Brittany said, “the first one was horrible.” Both mentioned specific criteria for which the design performed poorly.</td>
<td>11</td>
<td>No response.</td>
</tr>
<tr>
<td></td>
<td>Aiden and Austin</td>
<td>8</td>
<td>Somewhat.</td>
<td>12</td>
<td>No response. However, team mentioned that D2 was better than D1.</td>
</tr>
<tr>
<td>Sheehan</td>
<td>Jocelyn, Brooke, Brennan, and Kim</td>
<td>12</td>
<td>No.</td>
<td>11*</td>
<td>Yes, overall. Because the plant was in poor health. (But later team shared that the padding for D2 was effective, suggesting that it only somewhat failed.)</td>
</tr>
<tr>
<td></td>
<td>Spencer, Rachel, Olivia and Cole</td>
<td>13</td>
<td>No. (Cole) No responses from other team members.</td>
<td>12*</td>
<td>Somewhat. (Spencer). Spencer said that the preserve score failed and the cotton balls, which were there to protect the plant, started to slip. No responses from other team members.</td>
</tr>
</tbody>
</table>

* Ms. Sheehan likely did not share one of the constraints for the plant package design: that the plant, once placed on the shelf, could not be watered for the three-to-four day shelf period. The teams in her class did not design an internal watering mechanism (e.g., a means to wick water into the plant’s soil) within their packages. D1 plants had been watered while on the shelf, however, teams were surprised that for D2 plants were not watered over the weekend, resulting in poor plant health.

Overall Failure

Many teams suggested that one or both designs were failures, overall, as shown in Tables 1-3 and summarized in Table 7. There were two more obvious cases of this: 1) bridges that could not support themselves, and 2) TarPuls that received a “no score” evaluation. Molly, David and Zora in Ms. Lee’s class could not engineer a bridge that could support itself – much less support weights or allow a car to pass over or a barge to pass under it. The team was clear in expressing that both D1 and D2 were failures, without having to get into detailed statements about criteria. Two teams – Molly, David and Erica in Ms. Lee’s class and Samar and Isabelle in Ms. Jefferson’s class – earned a “no score” for their D1 TarPuls. Recall that a “no score” means that the TarPul is ineligible to receive an overall score since it was not able to carry an adult person.
Molly shared that D1 “failed overall,” and her teammates agreed. Isabelle indicated that her “no score” D1 failed, and Samar shared that “a lot of parts failed” on this design.

Table 7. Summary of team responses to the question: Do you think that your design failed?

<table>
<thead>
<tr>
<th>Unit</th>
<th>Design</th>
<th># Teams</th>
<th>Team Responses to the Question: Do you think that your design failed?*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Yes, Overall</td>
</tr>
<tr>
<td>Bridge</td>
<td>D1</td>
<td>11</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>D2</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>TarPul</td>
<td>D1</td>
<td>4**</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>D2</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Plant Package</td>
<td>D1</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>D2</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

* In some teams, there was disagreement. In these cases two responses per team were included in this table. ** One team was not included in this analysis since the team’s design performance was unclear.

Similar to Samar’s idea that “a lot of parts failed,” other teams labeled entire designs as failures because these designs did not score well across multiple criteria. This evident in the design experiences of four teams: Savannah and Grant and Noah and Arianna in Ms. Jefferson’s class and Kayla and Raja in Ms. Kraft’s class for the bridges unit, and Bethany and Brittany in Ms. Kraft’s class for the plant packages unit. Their bridges and packages scored low on more than one criterion. For example, the following is an excerpt from Savannah and Grant’s design performance summary:

Savannah and Grant’s first bridge design utilized paper and index cards to make a deep beam bridge with some piers. When tested, it did not allow the barge to pass through, allowed the cars to fall off the edge for the stability test, and only held three weights. When asked what didn’t work well in this first design, the two shared these details about the barge, stability and strength. Savannah summarized that “pretty much everything” didn’t work well in D1, after which she and the other interviewees laughed. Grant and Savannah agreed that the first design failed.

These low scoring areas became the focus for D2, in which the team improved their cost, stability and strength scores.

In other cases, team members seemed to identify designs as failures if the team scored poorly on what I have deemed “the most important criterion.” This criterion was not identified in the curriculum as such, but seems to have been selected by some teams to be the most important criterion above the others – superseding the importance of the overall design score. For some teams or team members, this criterion simply seemed to be whichever criterion they struggled with the most. For Arianna and Noah’s D1 TarPul, it was the erosion score that was problematic, leading Noah to assert: “very curved, so … a big failure;” however, Arianna suggested that D1 failed in this way, but not overall. The poor health of the plant in Jocelyn, Brooke, Brennan and Kim’s D2 package led them to initially define D2 as a failure; yet later in the interview, they
admitted that the protective elements of the design worked well. Members of two bridge teams – Cassidy and Douglas in Ms. Kraft’s class and Samar of the Samar/Isabelle team in Ms. Jefferson’s class – identified the weight score as the most important. While Cassidy and Douglas were unable to support much weight at all in either design, Samar and Isabelle received a weight score of 4 for D1, a high score. Despite this, Samar suggested that D1 failed “because it couldn’t hold 100 weights.” He ignored the idea that the overall score was impacted most significantly by his team’s high cost score, not the weight score. Although D2 had a lower overall score, since it held more weights, he did not consider it to be an overall failure. Rather, he said that “certain things” failed on the design (perhaps referring to cost). Isabelle disagreed with his assessment that D1 failed overall, suggesting that D1 (and D2) failed only somewhat could have been improved by lowering the cost.

One of the more interesting examples of a team using the most important criterion was with Caroline and Gabriel’s team in Ms. Jefferson’s class. Their overall D1 score was a 2; about this, Gabriel shared, “[Ms. Jefferson] said two was a pretty great score so there wasn’t any problems.” Caroline offered that it “turned out pretty good.” For their second design, they stayed at the same site, but chose to do some compaction of the soil to attempt to hold more weights. They indeed held more weights (33; 3 people in the TarPul), but their overall score was a 3 – an increase from D1 – because of the cost of compaction. When asked if they thought their first or second design failed, Caroline offered: “Well, actually, I think the first one. The second one actually held more people … even though we got a higher score, it was good.” A short while later, the following exchange occurred:

Gabriel:  I think [design] two [was better] because … if you were carrying sick people across the TarPul you could get – like –

Caroline:  Three at a time.

Gabriel:  Yeah, three people across quick.

This is an interesting perspective in that both Gabriel and Caroline suggested that failure should largely be determined by the weight score – not the overall score – because the most important thing is for more people to be able to use the TarPul. Clearly, the team did not seem to think that D1 was an abysmal failure; however, their thoughtful response put their initial thoughts about D1 success in context.

Failing Somewhat

Many teams and team members suggested that their designs failed in certain ways, but not in others. I coded this as the response, “somewhat,” to the question: Do you think that D1 or D2 failed? These responses sometimes included a general response such as “kind of” or “parts of it”, or a recounting of criteria that the design did and did not meet well. Some examples of this have already been mentioned, e.g., both Arianna and Isabelle disagreed with their respective teammates’ suggestions about overall design failure, instead suggesting that the designs failed only somewhat.

The following design performance summary for Aiden and Austin’s team in Ms. Kraft’s class reflects the idea of failing somewhat – or “a little bit” – quite well:
Aiden and Austin’s total D1 score for their plant package was an 8. They lost points for the cost (score of 2), communicate (score of 1), carry (score of 0), and protect (score of 0) criteria. Despite this, they commented that their plant was healthy and “still growing” (Aiden). For their second design, they aimed to focus on improving protection, as well as the communicate and carry scores. They added more materials, increasing cost and lowering their cost score to 1, yet all other scores improved. Their D2 overall score was a 12, with the two points deducted because of cost. When asked if either design failed, Austin said that D1 failed. He reflected: “I think it failed a little bit. And our second design was better, and then our third one was even better. So I think … we learned from our mistakes to fix it.

This is an ideal design experience in which team members engaged in failure analysis – determining how D1 performed with regard to the criteria and what to attend to in subsequent designs– made improvements, and were able to see an increase in performance as a result.

No Failure

As is evident in Table 4, it was relatively rare for teams to suggest that their designs were completely successful – i.e., that they did not fail in whole or part. No bridge teams asserted this for either D1 or D2, and many teams shared in interviews that the bridge unit was the more challenging of the two that they learned. Two teams said that their second TarPul designs did not fail. One of these teams was Molly, David and Erica from Ms. Lee’s class. They had a “no score” for their first design, and earned 12 points for their D2 overall score – a high score for a design challenge in which the lower the score, the better. In their case, I suspect that the jump from a D2 “no score” to being eligible to score, and thus being able to carry a person in the TarPul, made this team feel successful. It’s also important to note that this was the first experience of design challenge success for Molly and David, whose D1 and D2 bridges were not able to support themselves. The other case of D2 success for the TarPul unit was for Caroline and Gabriel, whose D2 TarPul was able to carry three people; this was the most important criterion for this team, discussed earlier.

Four teams who learned the plant package unit suggested that their first and/or second designs did not fail. Two teams in Ms. Kraft’s class who were in a focus group and interviewed together – Raja and Douglas and Cassidy and Kayla – said that both D1 and D2 did not fail. When Raja and Douglas were asked if either design failed, they didn’t answer the question directly, but Raja offered: “first we got an 11, which is the second highest, and then we got 13, which is the highest.” After sharing this, Cassidy shared that her team received the same score, to which Douglas replied “we were all tied.” Raja and Douglas’s package designs scored relatively high; additionally, since they were high or the same compared to other groups in the class, this may have made them feel that their designs did not fail.

Two teams in Ms. Sheehan’s class had first designs that scored high, and indicated that those designs did not fail; in comparison, the teams suggested that the second designs failed somewhat – especially when compared to the first designs. There is an underlying issue here that is worthy of note related to a lack of consistency in testing procedures by Ms. Sheehan. According to the EiE curriculum, part of the plant package design challenge is for teams to consider how the plant can stay hydrated without it being watered from the outside while it sits
on the shelf for three to four days. (Teams may design systems like reservoirs of water and wicks to deliver water to plants for this purpose.) However, Ms. Sheehan did not follow this part of the challenge, and allowed for D1 packages to be watered throughout the testing period. The D2 packages also sat on a shelf for testing, but since this testing occurred over the weekend, they did not get watered externally; as a result, the D2 plants were wilted and discolored. Thus, the high scores for D1 were actually artificially high for the plant package design challenge.

**Broad Themes**

Three broad themes emerged from analyzing the design performance of the study participants, and have been thread throughout the aforementioned examples and discussion: disagreement, relative failure, and disappointment and humor. First, and as has been shared thus far in multiple examples, members on a team who share the same design challenge experience do not necessarily agree about whether or not or how much a design failed. I see evidence of this in team interviews for at least four of the teams in five separate interviews, with one team – Isabelle and Samar – having disagreements across both the bridges and TarPul units.

Second, a team’s or student’s assessment of whether or not a design has failed in whole or part is somewhat relative, despite having a common overall scoring system for each unit via the curriculum. Students and teams may consider how they performed across all criteria, and take a sort of average to determine failure. Douglas took this approach when another team in his focus group was discussing their D1 strength, stability and cost results. That team said that their design failed, but when they mentioned that cost and stability were fine, Douglas interjected: “It would be less of a fail because it’s two on one instead of one on two.” Other teams or team members may use the most important criterion to determine failure. Still, another approach is to compare scores – feeling that if your team’s overall or criteria score is lower than other teams in the class, this is more of a failure than it is if you score similar to or above those teams. Design failure may also be relative within a team when comparing D1 and D2. While Molly, David and Erica’s D2 TarPul score of 12 was not great, it was a significant improvement when compared to their D1 “no score.” To other teams, a score of 12 may have been considered a failure.

A third point is about the emotions that are evident in interviews when teams are reflecting on their design failure experiences, which include disappointment and humor mixed in with relatively unemotional retellings of those experiences. Disappointment was largely evident in the bridges unit for those teams who did not see improvement from D1 to D2, or whose scoring was lower for D2 as it was for D1. Samar seemed disappointed that he wasn’t able to hold as many weights as he wanted – or perhaps, as many weights as other teams could hold. Molly, David and Zora were clearly disappointed by their non-self-supporting bridge, and Douglas looked down as he reported in the interview that “they both failed.” The plant package teams who discovered that their plants had not been watered over the weekend seemed disappointed, as well, holding droopy looking flowers during the interview. These demonstrations of disappointment were balanced, though, by examples of humor and laughter around failure experiences as retold in interviews. David joked that his team’s “no score” TarPul might not be able to hold an adult, but it could hold a baby, and his team laughed and smiled after sharing in unison, “the first,” when asked if their first or second TarPul design failed.
Brooke, Jocelyn, Brennan and Kim demonstrated some levity when retelling how their first bridge design failed after holding only four weights:

Brooke: First design failed at 4 weights.
Jocelyn: It was much better …[the second time]
Kim: I know (laughing). Sounds so pathetic.
Interviewer: How would you say that the first one failed?
Kim: It failed dramatically. It was like, "Oooahhh" (arms sweeping downward)
Jocelyn: It fell, like going down, going down - going down, then collapse. But it wasn't going down too much - because –
Brooke: [Few words, inaudible; trying to jump in]
Kim: And the weight cup just flew like ahhhhhh (lifting hands up) –

As was the case for David and his team, this exchange was more spirited than somber, despite the team’s D1 strength failure. Similarly, Grant and Savannah smiled and laughed a bit after sharing that their first bridge only held three weights. It is important to note that these emotions were expressed at the end of the design challenge, and may not be the same as the emotions that students express either in the moment – when they are determining their test results for a design – or in an individual setting (i.e., not in a focus group interview).

**Fail Words In and Out of School**

Of the 29 students who took part in focus group interviews, 25 (86%) shared specific examples of where they have heard fail words in or out of school. This count of “specific examples” did not include cases in which students agreed with another student’s contribution. Four students did not provide specific examples: Caroline did not respond to the question at all; Zora responded that she had heard fail words in and out of school, but did not elaborate; Isaac said that he had heard fail words out of school, but did not elaborate; and Savannah was unsure, responding with statements such as “maybe in a book?” and “I don’t know.”

**Fail Words within the Context of Engineering**

Eleven students (38%) included examples of fail word use from the engineering units that they studied: bridges and TarPul, or bridges and plant packages. For example, in his post-bridge unit interview, Douglas said that “We heard failure in school ‘cause when Ms. Kraft was talking about it when we were building our bridges.” Four other students, Molly, Samar, Jocelyn, and Cassidy, reflected on fail word use related to bridges – and in one case, the tower activity that teachers taught at the beginning of the bridge unit - during their post-bridge unit interviews. In her post-TarPul interview, Erica shared that she had heard the term, “failure line,” which was from the bridges unit. Three other students, Kayla, Brooke, and Isabelle, also shared that they heard fail words in the bridge unit during their post-TarPul or post-plant-package interviews. In this way, their examples extended back one unit. For four of the students (14%), responses to the question - “Do you ever hear ‘fail’ or ‘failure’ out side of school or inside of school?” – only referenced the EiE engineering units that they learned; in other words, they shared no other examples of fail word use from in or outside of school. Olivia’s only response was: “Only when I’m doing engineering.”
Five students (17%) provided examples of fail word use that were related to bridges or towers, but were not specifically from the unit. Kayla and Raja recalled fail word use related to tower building at home, Grant remembered reading a book about bridges and buildings that referenced failure, and Jocelyn recalled a conversation she had with her mother about failure and her bridge design. Noah shared a story of hearing bridge workers on a bridge talking about failure; this story is either an example of incredible luck and coincidence – to have overheard bridge workers talking about failure – or somewhat questionable.

**School Related Fail Word Use**

Beyond the engineering units students studied in school, six students (21%) provided other examples of school-related fail word use. Samar, Kayla, Brittany and Cole all mentioned failing a test or a project. Some of these were specific examples of their own or sibiling’s test failures. For example, Kayla recalled: “My mom said one time that you almost failed your math test, because I got a D.” Cole offered: “Like we do projects and sometimes they fail.” Recess was also identified as a place where fail words might be heard. Gabriel mentioned, “if you fail at climbing monkey bars,” and Spencer offered a tentative “maybe recess” as a place where fail words were heard, after which fellow interviewees Isaac, Rachel and Cole readily agreed.

**Books, Television and Movies**

Books, television and movies are places where seven students (24%) mentioned hearing fail words. Grant mentioned reading a book about bridges and building – *Feats and Failures* – in which there are “bridges like, falling.” In their post-bridge interview, Kim started a conversation about failure in books like those in the *Harry Potter* series and *Lord of the Rings* trilogy:

Kim: I hear it in books all the time because I've been reading *Harry Potter*. I've read all those and they say failure all the time.

Jocelyn: I'm reading *Harry Potter*, they say failure like every single page.

Brennan: And a lot of magic books, they - they always have failure because like the spells don't work. And sometimes it’s a failure.

Kim: I know, like *Harry Potter, Lord of the Rings* … [another student interjected an off-topic comment] … Gandalf says failure a lot.

Brennan recalled reading fail words in books – “especially ones where they give morals” – in the post-plant-package interview.

Students mentioned hearing fail words in the television shows *Wipeout* and *Odd Squad*, as well as in the movie, *Despicable Me.* David inspired an interesting exchange with his team members about fail words within *Wipeout.*

David: I heard it … on TV, there's this show when the people do stunts …

Molly: Like *Wipeout*.

Zora: Fail.

Molly: You hear a lot of the word "fail" and "failure."

Zora: *(Swaying back and forth).* Fail fail fail fail fail.
David: Yeah, like in that – fail –
Molly: And failed.
Interviewer: When do they say – when do they say it?
Zora: Fail.
Interviewer: When somebody messes up?
David: Yeah. Whenever they get – yeah – when they fall into the water and all that. They have a whole bunch of failures.
Molly: Yeah, cuz, there's a lot of messing up.

David also readily and excitedly responded with “TV!” in the post-TarPul unit interview.

Fail Words in Informal Play and Organized Sports

Fail words are also a part of nine students’ experiences (31%) with informal play and organized sports. Gabriel mentioned failing to “[jump] off the diving board,” and Cole mentioned trying to jump a racecar on his racetrack at home and “every time I do it, and I miss, I’m like: fail!” Cole brought up the game, Jenga, in his post-bridges focus group interview, and the team talked about the Jenga blocks collapsing. Olivia added: “it’s like (shows collapse with hands) oh, I failed!” Grant, Noah, Arianna, Rachel and Austin all mentioned hearing fail words during organized sports. Grant heard it in his soccer practice: “We had to dribble through the cones, but a person didn’t do it so then the coach said, ‘you failed.’” Noah’s baseball coach called his team “a big failure” after losing a game, and Austin’s friend’s skateboard coach told his friend that he was “being a failure.” Aside from coach commentary, Rachel mentioned that she sometimes says to herself that she fails when doing gymnastics, and Arianna mentioned that her dad uses fail words when her sister is not running or taking the ball in her soccer games.

Parents Using Fail Words

Arianna’s father is one of many examples that students shared about parents using fail words. Altogether, seven students (24%) shared such examples. Parents sometimes used fail words related to their children; this was the case for Arianna’s sister. Bethany recalled her mother telling her brother that he was failing a grade, and Douglas recalled his mother saying that he almost failed when he nearly knocked over something at a grocery store. Raja and Kayla who were interviewed together, both recalled very similar stories of their mothers saying that the towers they built at home out of blocks failed. In addition, Aiden and Erica mentioned fail words related to their father’s employment. Erica shared: “My dad keeps saying that he fails on his job a lot because he has a depression kind of thing.” In the post-bridges unit interview, Aiden recalled hearing his father on a work-related phone call saying that something was not a failure. In the post-plant-package unit interview, he discussed how his father described a work project as a failure.

Siblings Using Fail Words

Siblings were the often the subjects of fail word discussions. In some cases, they were mentioned because it was they who were doing the failing (e.g., Bethany’s brother failing a grade). More directly, Erica, Spencer and Rachel mentioned hearing fail words from their
siblings – and Raja recalled a dream in which his brother used the word; thus, four students (14%) heard fail words from siblings in real or dream worlds. Erica shared: “My sister taunts me with my homework and says I’m gonna fail.” In the post-bridges unit interview, Spencer said that his brother “always says, awe, you failed (laughing).” He mentioned that his brother says failure again in the post-plant-package unit, after which, Rachel said: “so does my sister.” Raja recounted a dream that he had in which “I had a dream where my brother, he only knew how to say failure … like: ‘failure,’ ‘failure,’ ‘failure,’ (singing/chanting it, laughing).”

What does it mean to Fail?

After students were asked where they have heard fail words in and outside of school, they were asked: What does it mean to fail? This was meant to be an open-ended inquiry to explore students’ ideas about failure. Later in the interview, I asked students if their ideas about failure changed after learning more about engineering. Students responded in a wide range of ways, telling us that this question was a bit too abstract for most students. Many, for example, responded by recounting what they learned from the design challenge, rather than what the other interviewers and I were after, which was how their ideas about failure changed. Many more said that their ideas remained the same, or were unsure how to respond. Just four interview excerpts across four codes from this “how did your ideas change” section of the interview answered the question with some degree of elaboration. Since these excerpts all have to do with students’ conceptions of failure, I have captured those responses in this section along with students’ answers to the question: What does it mean to fail?

Failure Means That it Doesn’t Work

By far, the most common meaning that failure had for students was something did not work – it performed poorly in some way or another. This was mentioned by over half (16, 55%) of the students in various ways. Here are some specific examples of how students communicated this idea: “you didn’t get a good score” (Isabelle); “[to] not work … to not function” (Kim); “something doesn’t work” (Spencer) “it’s not strong enough” (Rachel and Cole); “you can’t really fix it” (Cole); “if you plan for it to do this and it ends up doing the complete opposite” (Brennan); “to not succeed” (Douglas); and “it’s always sagging [and] cars can’t roll over the bridge” (Kayla).

How Things Fail Changes According to Context

In some cases, students suggested that what it means to fail – and specifically, the nature of what did not work or performed poorly – changes according to the context. For example, Cassidy shared the following in her post-plant-package interview: “Failure is like if the plant is tipped over, or the bridge is sinking or falling.” Here she references examples from both the plant package and bridge units. This is also evident in the following exchange in response to the question about how ideas about failure have changed. Initially, Kim, Jocelyn and Brenna responded with “not really” and “a little bit,” but then Brooke shared another idea, which is the beginning of this excerpt:
Brooke: Yes. (Gesturing to plant package.) Because that doesn't collapse … [some other team responses to initial question] … The bridge design - failure, to us, meant collapse, or –
Charlotte: It means a lot of things to fail. [a few interjections by other students] …
Interviewer: To some of you, failure meant collapse like a bridge but, to others of you, failure means more than just that.
Jocelyn: Now it means to me - If it was a plant, it will shrivel up.
Brooke: Like it would shrivel up.
Jocelyn: If it's something like a car, it will probably explode if you're doing something wrong. And if it's a bridge, it will collapse or something.
Brooke: Or maybe if it was a car it could just – it could just stop working or the engine could blow.
Brennan: Failure / to different things means something else. Cars can mean explode, crash. And then bridge, collapse, fall down.

Here, students drew from both the bridge and plant package units, and interjected their own ideas about vehicle failure. Note that some of the ways in which failure is discussed in this excerpt are what I would call “big failures;” i.e. collapse or explode. While collapsing was discussed by multiple students due its prominence in the bridges unit, explosions were only identified by this group (Brooke, Charlotte, Brennan, Kim and Jocelyn.). In a different excerpt, Kim also mentioned that to fail mean to “lose, die … [and] spoil.”

You Fail when it’s Too Difficult

Other conceptions of failure included that something was too difficult, as suggested by four students (14%). Molly, David and Zora, in their post-bridges unit interview and shortly after their D1 and D2 bridges could not support themselves, discussed what failure meant to them:
David: Like, it means that you can't … well, you try to do it, but you can't do it.
Zora: You can’t do it right.
David: You can’t.
Molly: You try to accomplish, but it was [too –
David: It was too hard.
Molly: Hard and you couldn't really do it.

Coupled with this idea of a task being too difficult is that it is so difficult that you cannot complete it. I see this in David and Molly’s contributions, above, as well as in Gabriel’s statement that to fail means to “not complete your objective.”

Failure Means Doing it Incorrectly

In the above exchange, Zora offered: “You can’t do it right.” This is related to something being too difficult, but is somewhat different. The idea here is that a task could not be done correctly. Six students (21%) shared this idea. Samar, Rachel, and Cole offered statements
similar to what Zora shared. Arianna and Grant’s sports examples both contained suggestions that someone was told that they failed because they did not do something correctly in soccer.

**Failure Signals the Need to Improve**

Six students (21%) suggested that failure indicates the need to redo, try again, and improve. For example, Aiden shared: “But if your thing is a failure, that just means that it’s wrong and you have to go back and use the EDP again to improve it.” Douglas offered: “So failure is that you don’t succeed, but it doesn’t mean to stop trying.” Kayla agreed with this, saying: “That’s what I was trying to say!”

Three other students – Kim, Caroline, and Spencer – alluded to the ideas of improvement and iteration. Kim shared: “Just look at Alfred Nobel. He invented gun powder. This guy must have failed a million times.” Caroline described the kind of failure analysis that would occur prior to improvement, offering: “Now you … learned your lesson, and now, in the future, you know you don’t wanna do that again or else you’re going to fail again.” Finally, after learning the bridges unit, Cole, Rachel and Spencer had the following exchange:

- Cole: I thought failing was like during football and you could be running for a touchdown and then trip. That's like a fail.
- Rachel: Yeah. And like I said, fail at gymastics like you fail at football …
- Cole: Yeah, Yeah I fail too.
- Spencer: Like in this, you can fail when you need to.

My interpretation of Spencer’s contribution, “you fail when you need to,” is that failure is not a dramatic mistake like tripping and preventing the team from getting a touchdown. Rather, it is okay to fail in engineering. Extending this interpretation perhaps too far, Spencer’s comment might also suggest that designs fail when they need to be improved in some way.

**Failing = Quitting**

One focus group of students within Ms. Kraft’s class spent many minutes of their interview discussing what they thought of as failure: giving up or quitting. The following was the first of these exchanges in the post-bridge unit interview:

- Brittany: Mrs. Kraft told us.
- Bethany: It means to – I know what it is, I just - okay, you can go to Brittany.
- Interviewer: Okay.
- Brittany: It means that - you fail when you like don't do something right, but you don't fail until…you have to fix it then it won’t, it wouldn’t be failed.
- Interviewer: So, if you fix it then it's not, than it doesn't fail? Do you fail when you don't do something right but you can fix it? …
- Brittany: Yeah
- Austin: I know exactly the definition … Something fails when – something fails when it doesn't work and you give up.
- Interviewer: All right, Aiden, how about you?
Aiden: It’s when - a failure is when you - when something fails and then you give up because it fails.

Interviewer: And how about you Austin?

Austin: It means that like if you're building bridges and it falls down, you fix it and you don't quit. … A failure means that you quit from doing it ever again.

Note that the students seem to be referencing a definition of failure-as-quitting from their teacher, Ms. Kraft. This actually caused some confusion on what it meant to fail, as was evident in this exchange later in the interview:

Bethany: It only fails until you fix it.

Austin: No, it fails when it fails.

Bethany: And we fixed it a few times

Aiden: We didn't fix it because – because it was still - [inaudible 00:20:22; shifts into an off-task argument]

(Recall from the first part of the findings section that Aiden and Bethany’s second bridge design continued to have problems holding weights.) Another point of confusion in the interview was when Bethany shared that the main character, Gru, in the movie, Despicable Me, said “I failed.” This eventually prompted an argument about whether he really failed:

Bethany: Yeah, I heard on - on in a movie, "Despicable Me." (laughter) That's my dad's favorite. So we watched it and it said, "I failed" -

Aiden: “I failed!” (Mimicking the main character, Gru in a funny voice.)

Bethany: And it was weird because Gru failed something -

Aiden: “I failed!” (Again, mimicking in a funny voice.)

Bethany: I forget what, but I think it was with his mother.

Brittany: Yeah, but he didn’t really fail because he tried again to fix it.

Bethany: No he didn’t.

Brittany: Yes he did.

Bethany does not seem to think that Gru actually failed since he tried to fix whatever it was that he had failed. This same definition was evident in the post-plant-package unit interview. Austin seemed to try to resolve the issue by differentiating between “failure” and “fail”:

I heard it inside of school. And failure, I think, is when somebody fails and they just give up and do not try how to fix it. And a fail is when you do it wrong but then you go back and fix it so it's right. (Austin)

Later in that interview, Brittany offered, “It sounds like fail and failure are almost the same thing.” Bethany ultimately replied, “not exactly, they kind of are but not exactly.”

Although Ms. Kraft may have been very well intended in her effort to share that failing is only associated with giving up, it was clearly a matter of confusion for the students. The other focus group in Ms. Kraft’s class did not seem to spend as much time on this issue. However, aspects of this failure = quitting messaging were evident in this response by Douglas with regard to how his ideas about failure have changed since learning the plant package unit:
I don't know. It changed by knowing that if you work on it more … that you can work on it a long time, and it won't be a failure, so if you work on it for a long time that means that it will get stronger and stronger, if you keep working on it, improving it. (Douglas)

This suggests that failure is only failure when one stops working on a design.

**Failure as an Identity**

Six students (21%) suggested that the word failure – rather than fail – can be associated with people. Molly said that if “somebody calls me a failure ... I just don’t listen;” she also said that nobody had called her a failure, but she was prepared to handle that situation and clearly saw a failure label as being something to avoid. Brennan recalled that some books suggest “you’re a failure, then it ends up not being what they actually say.” Some students in Ms. Kraft’s class debated what it meant for a person (the engineer) or a thing to be a failure. Raja, Kayla and Douglas, from one focus group, debated this:

Raja: Another failure is, failure is something that didn't work out, like if you plant something and you give it too much water, and then it dies, it was a failure, or you are the failure because -

Kayla: No, your plant would be the failure, wouldn't it?

Raja: I don't know, actually.

Kayla: Would your plant be the failure, or would you be the failure?

Douglas: Technically … it would be both.

Raja: Yeah, both.

Kayla: Yes, actually it would.

Douglas: The plant failed by not staying alive, but you failed at doing the right thing.

Aiden, in another focus group in Ms. Kraft’s class shared:

If you’re a failure, that’s what you do. But if your thing is a failure, that just means that it’s wrong and you have to go back and use the engineering design process again to improve it. That’s all you have to do, if only the machine is a failure, not you. (Aiden)

Given the theme of failure = quitting in Ms. Kraft’s classroom, the comment, “if you’re a failure, that’s what you do” likely means “if you’re a quitter, that’s what you do.”

**Conclusion and Discussion**

*Design Performance Summaries and Students’ Assessments Regarding Design Failure*

The following conclusions were based upon the design performance summaries, constructed via analysis of interview data and students’ engineering journals.

- Students may identify designs as overall failures, and do so if designs:
  - Fail in a fundamental way (e.g., they design a bridge that cannot support itself);
  - Performed poorly across many criteria;
  - Failed to perform well on a “most important criterion” – a criterion that the students deem to be important; or
Perform less well than other teams’ designs in the classroom.

- Students may see a given design as having failed somewhat, given that the design performed well with respect to some criteria and poorly with respect to others.
- Students may identify designs as being overall successful if:
  - They perform well with respect to the most important criterion;
  - Meet all or most of the criteria;
  - Are a significant improvement with respect to a past design; or
  - Perform as well as or better than other teams in the class.

Also, not all students on a team will necessarily agree about the extent to which a design has failed (e.g., overall versus somewhat). Finally, while reflecting on past design failures during interviews, some students expressed disappointment, consistent with student responses to failure identified in previous research. Some expressed humor and made light of those failure experiences, and most were able to relay design failure experiences in a relatively straightforward, unemotional way.

Perhaps the biggest takeaway from design performance summary analysis is there are multiple ways that students assess the failure or success of their designs. Some of those align with how engineers engage in failure analysis, i.e.: by testing the design and identifying how well the design performed with respect to design criteria. I shared an example of this from Aiden and Austin’s plant package design performance summary in the Failing Somewhat section. However, not all students took this criterion-by-criterion approach to assessing design failure. There were two ways of assessing failure that diverged from this approach. The first of these is when students focused on one most important criterion as a measure of success. In doing so, tradeoffs among criteria are ignored in favor of a single-criterion measure of performance. Students’ subsequent analysis of design performance may not be as robust. Another way of assessing failure is to compare overall design performance (i.e., the overall design score) against that of other teams in the class. This is okay so long as when asked why designs failed or performed well, students can answer not only with respect to team comparisons in final scoring, but also with respect to how their design performed against the design criteria.

Finally, and as has been discussed in previous studies, students cannot learn as much from failure experiences if design testing procedures are not followed properly. The teams in Ms. Sheehan’s class, for example, could not easily compare their D1 and D2 designs since D1 was watered while on the shelf, and D2 was not. Further, these groups were not sufficiently challenged by the package design challenge because they were not prompted to consider how to create a mechanism inside the package to keep the plant moist while it sat on the shelf, as the curriculum intended.

*Fail Word Use*

Nearly three quarters of the students in the study (72%) offered specific examples of fail words to which they have been exposed outside of the EiE units that they learned as a part of the E4 Project. They heard these words in academic settings, e.g., failing a test or a grade, and read or heard fail words in books, television shows, and movies. Games, play and organized sports were spaces where they heard fail words. Coaches, parents and siblings used fail words. In all of
these examples, the connotation was negative, and the implication was that someone or something failed or, less often, was a failure; with just a few exceptions (e.g., collapsing towers) examples typically did not specify failure against a particular criterion, as they would for engineering. Students also mentioned their EiE units of study as places where they encountered fail words. Additionally, in interview excerpts shared throughout this paper, some students—specifically, Zora, Austin and Raja—used fail words in a repeated, silly way within interview discussions. Although it is difficult to surmise the meaning of these interjections, it is clear that fail words seem to have attention-grabbing power here and in the popular culture in which students are immersed.

Many elementary teachers hesitate to use fail words in their classrooms because of the negative connotation of those words within education, and for fear that students may, through this exposure, think of themselves as failures.9,23,25 Further, some in the maker movement worry about using fail words because of this negative connotation. It is indeed advisable for teachers to use fail words with care in their classrooms, and to always associate failure with a design and not a class, team or a student. However, students have already heard and used these words in their lives both in and outside of school. Students have been exposed to more and less ominous uses of fail words. Engineering education provides an opportunity for teachers to recast fail words in the specific contexts of engineering design challenges, suggesting that design failure is determined with respect to criteria and providing valuable insight towards growth and improvement.8

What it Means to Fail

Students assign meaning to the fail words that they encounter in and out of the context of engineering design. Most associate failure with poor performance or something not working; this is a helpful way of thinking about failure in engineering, in that design failures identify what does not work to meet criteria within a design. Especially promising are those cases in which students begin to see that what it means to fail is different depending on the design context: a bridge fails in different ways than does a plant package or TarPul. Further, it is exciting to see students articulating ways in which design failure signals the need to conduct failure analysis and improve—insights gleaned most likely from their engineering learning experiences and that are consistent with resilient, productive student responses to failure (see Table 1).23,25

As was evident from this study, one of the more problematic ways that teachers can define failure is by equating it with quitting. Recall from a previous study involving 254 teachers that the word or phrase most frequently associated failure was “giving up,” which included words like “quitting.” There are myriad clichés to be found online and elsewhere that say something to the effect of “you only fail when you give up.” The trouble with this motivational expression for engineering education is that it confuses students’ interpretation of when failure has occurred in the design process. Indeed, failure does not simply happen when one gives up in engineering; rather, failure occurs when designs do not perform well against one or more criteria—and engineers expect this to happen in design process. Thus, teams in Ms. Kraft’s classroom who have been given the failure = quitting formula are collectively not really sure if their designs have failed since they kept working on the designs through the end of the unit.
Implications and Future Work

This work has implications for inservice and preservice elementary teachers learning to teach engineering, as well as the preservice teacher educators and professional development providers who guide their learning. Six suggestions we share here are that teachers of engineering should:

- Have explicit discussions with students about fail words and their meanings both inside and outside of engineering instruction;
- Present a definition of engineering design failure as when a design does not perform well against one or more criteria;
- Model appropriately directed fail word use, associating failure with designs and criteria, not with the class, teams or students;
- Have explicit discussions with students about the importance of multiple criteria – and tradeoffs among them – to determine the overall success or failure of a design;
- Avoid clichés about failure, since many of those confuse students’ understanding of design failure within the EDP.
- Ensure that students follow proper testing procedures throughout the design process so that they can accurately analyze design performance and learn from design failures.

Each of these suggestions will be discussed in more detail in the sections that follow.

Fail Word Use

Given that most students have already encountered fail words in their lives, engineering instruction should include explicit discussions with students about fail words and their meanings. Teachers might start by asking students in a whole class discussion what fail words they have been exposed to and what fail means in those contexts, and then present the specific way in which fail words will be used in the context of engineering design (e.g., via terms such as “failure line” or more generally, design failure). Students are capable of learning that words have different meanings in different contexts. Following from other studies, teachers should be sure to associate fail words with the design, not the students. As was the case in the present study, students will likely share examples of fail words in which individuals have been identified as failures. Students can learn that for engineering, the idea is not to identify people as failures, but rather to identify ways in which designs failed to meet particular criteria so that in the next part of the design process, they can improve their designs.

Multiple Criteria and Moving Beyond Comparison Across Teams

Teachers of engineering should also help students understand that there are multiple criteria to assess designs for good reason, and that it is important to attend to all criteria when assessing the overall performance of designs. A related discussion is about tradeoffs among criteria (e.g., increasing bridge strength with more materials may increase a strength score, but decreases the cost score.) If a whole class agrees that one criterion is more important than others, consider weighing this criterion in the overall scoring of the design; this could be a healthy conversation with students. Teachers should also be aware that students may use other means to judge the success or failure of their designs, most especially team-to-team comparison. This is
somewhat unavoidable given that whole-class discussions of design performance after D1 or D2 are an important part of engineering instruction. Comparison in and of itself is not necessarily bad; however, the discussion should not stop there. Failure analysis should not consist of: we failed because everyone else did better than us, or we did well because we got the same score as everyone else. Rather, students should be pushed throughout the design process and in those whole class discussions to articulate ways in which their designs performed against criteria.

Clichés

Also, when teaching engineering, avoid using clichés that confuse what it means to fail in the context of engineering. This includes “failing is only when you give up.” The problems with this association between failure and quitting were discussed earlier in the paper. It also includes “failure is not an option.” Design failure is always an option in engineering design, and can present greater learning opportunities for students than successes. Further, even the expression that FAIL simply means “First Attempt in Learning,” a phrase seemingly used to temper negative reactions to the word, is problematic. Quite often, designs fail the second time, too, as some teams in the study discovered, and 408 times, as those who developed the Formula 409® cleaning spray experienced. The Testing Process

Finally, it is impossible to truly learn about the performance of a design if testing procedures are not followed accurately and consistently. This is a foundation on which analysis of design performance and learning from design failure is built, and has been identified in studies mentioned earlier in this paper. In the present study, the students in Ms. Sheehan’s class did not get the full benefit of learning from design failure given that the testing process was different for D1 than for D2.

Future Work

In the near future, I plan to expand this study to include the two other assigned units that E4 Project CIO site focus groups learned in 2014-2015. In addition, I hope to analyze the focus group video for these focus groups, strengthening the evidence within design performance summaries, and examining students’ real-time reactions to design failures. I also aim to conduct case study analyses within select CIO classrooms in the E4 Project, incorporating whole class video, focus group video, teacher interview, focus group interview, and class journals. I remain interested in how students and teachers make sense of failure and fail words in and outside of the design process, and how students and teachers respond to design failures in the moment and in retrospect.
References

5. Reimers et al. (2015) [Quote from p. 47]


27. Andrews (2016) [Quote from p. 28]


38. Wipeout television show on ABC (ended 2014) by creators Matt Kunitz and Scott Larsen.

39. Odd Squad television show on PBS Kids by creators Timothy McKeon and Adam Peltzman.

40. Despicable Me movie by Illumination Entertainment and Universal Studios.

41. Jenga game by Hasbro.


**ACKNOWLEDGEMENTS**

This material is based upon work supported by the National Science Foundation under Grant No. 1220305. 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. The author would like to thank the student participants for their insights, and the E4 Project team for their support, especially Christine Gentry, Chris San Antonio-Tunis, Jonathan Hertel, Liz Parry, Cathy Lachapelle, and Christine Cunningham.
APPENDIX

Interview Protocol for the TarPul Unit

Bridge and plant package unit protocols are similar, yet address specific design contexts for those units. This is a partial protocol, displaying those questions that were relevant to this study.

Questions relevant to this study:

1. Tell me about your first design decision about where to put the TarPul and how to prepare the site for the TarPul.
   a. Which site did you choose? Why?
   b. What was the soil like at that site?
   c. Did you decide to compact either side? By how much? Why?

2. How did your TarPul design score the first time?
   a. What parts of your first TarPul design scored well? How do you know?
   b. What parts of your first design did not score well? How do you know?

3. What changes did you make to your TarPul design to try to improve it? (different site? Different soil? Different compaction?)
   a. What parts of your second TarPul design scored well? How do you know?
   b. What parts of your second TarPul design did not score well? How do you know?

4. Would you say that your first (or second) TarPul design failed? How so?
   a. Did parts of it fail?
   b. Did it fail to do certain things?

5. Where else have you heard the words, “fail” or “failure” in school or out of school?
   a. (In contexts mentioned by kids): What does it mean to fail?
   b. How do you think engineers might use the words “fail” or “failure”? (Not analyzed for the present study; will be analyzed in future studies).

6. Did your ideas about what it means to fail change after you did this engineering design challenge?