The Future of Engineering Education
2024 Annual Conference & Exposition
Oregon Convention Center
Portland, OR . June 23 - 26, 2024
Paper ID #43561

Elementary Student Teams' Design Failure Experiences and Factors that Affect their Opportunities to Learn from Failure (Fundamental)

Dr. Pamela S. Lottero-Perdue, Towson University

Pamela S. Lottero-Perdue, Ph.D., is Professor of Science and Engineering Education in the Department of Physics, Astronomy & Geosciences at Towson University. She has integrated engineering into courses for PreK-8 teacher candidates, developed and directed a graduate STEM program for PreK-6 teachers, and partnered with teachers to implement PreK-8 science-integrated engineering learning experiences. She has authored numerous engineering-focused teacher practitioner articles, chapters, and research articles, and presents her research regularly through the ASEE Pre-College Engineering Education Division, a division she has chaired. Her current research includes investigating how children plan, fail, and productively persist; how mixed-reality simulated classroom environments can be used to help pre-service and in-service teachers practice facilitating challenging discussions in science and engineering; and how undergraduate engineering design teaching assistants address (and may be able to practice addressing) team conflict within similar simulated environments.

Elementary Student Teams' Design Failure Experiences and Factors that Affect their Opportunities to Learn from Failure (Fundamental)

Abstract

The research literature has established that (a) learning from design failure and engaging in diagnostic troubleshooting are fundamental epistemic practices of engineering education, and (b) the ways in which teachers and students prepare for and respond to design failure is varied and complex. There is ample space for additional contributions to this literature, particularly with respect to how teams of students in K–12 classrooms negotiate failure experiences. This qualitative study examines 21 design teams across 8 classrooms in 8 elementary schools in the eastern United States as they engage in two Engineering is Elementary (EiE) units. There were 53 students and 2 to 5 students per team. Unit 1 for all teams was about bridge design. Unit 2 focused on the design of an electrical circuit, package to contain a plant, oil spill clean–up process, or site preparation to support piers for a bridge–like system. Research questions were: (1) To what extent do teams perceive that they have experienced design failure? (2) How do teams respond to and make sense of design failure? and (3) What factors within the classroom environment might challenge or support teams' opportunities to engage with design failure in meaningful ways? Data gathered included video footage of each team, student engineering journals, and post–unit video–recorded team interviews. One summary for each team and unit (42 summaries total) was generated using an analytic framework to distill data relevant to potential design failure experiences. Summaries included quotations, descriptions of team activity, and journal and interview excerpts, and were analyzed using collaborative, iterative analysis that involved defining and assigning a priori and emergent codes. Overall, 86% of Unit 1 teams and 90% of Unit 2 teams reported that at least one of their designs failed in full or part. Positive and productive responses to design failure included in that many teams engaged in diagnostic troubleshooting (62% Unit 1; 43% Unit 2) and some teams (fewer than 25% for each unit) persisted despite struggles. Negative or unproductive responses included that some teams made design decisions disconnected from testing evidence or design criteria (5% Unit 1; 14% Unit 2) or blamed other team members for design failures (19% Unit 1; 5% Unit 2). Students expressed emotions including satisfaction, joy, disappointment, and frustration as they responded to design failure experiences. Some teams determined success or failure based on competition or comparison with other teams. Factors that may have negatively affected some teams' opportunities to learn from design failure included alterations to constraints or criteria by teachers; inconsistencies in how students scored their designs in engineering journals; and unclear, inconsistent, or inaccurate testing processes. Three other factors—mid–create testing, interventions by teachers or parents, and intra– and inter–team dynamics—had the potential to either support or inhibit student learning from design failure.

Introduction

The engineering education literature asserts that learning from design failure and engaging in diagnostic troubleshooting are fundamental epistemic practices of kindergarten through grade 12 (K–12; ages 5 through 18) engineering education, in part because they are fundamental aspects of engineering practice [1-3]. Further, researchers have learned that the ways in which K–12 students experience and respond to design failure is varied and complex [4-8]. There is ample space for additional contributions to this literature, particularly with respect to how elementary students working in teams encounter and respond to design failures. (The elementary grades include kindergarten through grade 5 (K–5) or ages 5 through 11.) In what follows, I describe a qualitative study of 21 upper elementary student design teams' design failure experiences as they engaged in two Engineering is Elementary (EiE) units of instruction. I also explore how their opportunities to learn from design failure were supported or may have been inhibited.

Framework: Learning from Failure

This study is framed on the idea that design failure experiences can serve as opportunities to deepen learning. When positioned as a learning experience, design failure signals the need to change course—to learn how to alter the technology being designed to better solve the problem—as the designer proceeds in the design process. Andrews connected design failures to Piaget's "perturbations" in the environment that encourage students to get curious and learn [4, 9]. Another way to consider how design failure may inspire learning is to compare it to how anomalous data may do the same in science education. As Chinn and Malhotra found in four separate experiments, "upper elementary school children are fair minded in their observations of data about empirical regularities in science, and they are willing and able to change their beliefs in response to their observations" (p. 342) [10].

Squarely in the engineering education space, Crismond [11, 12] and Crismond and Adams [2] assert the importance of diagnostic troubleshooting in engineering whereby design performance is analyzed to allow for informed improvement. Specifically, diagnostic troubleshooting involves observing design behaviors, including failures, most often during testing; identifying design aspects or behaviors that failed; providing explanations for those failures; and then remedying the design. The explanations that follow observations and inform subsequent action may be scientific in nature or may have to with other aspects of the technologies that affect performance (e.g., material properties, dimensions); explanations depend on the nature of the design task. Thus, diagnostic troubleshooting may involve not only deeper learning about the design itself but also about underlying scientific or other concepts relevant to design performance.

In 2022, Jackson and colleagues conducted a systematized review of studies that conceptualized failure as an opportunity to learn [5]. Their inclusion criteria included empirical studies from across the STEM disciplines in K–16 (i.e., K–12 and university undergraduate) education between 2008 and 2019, resulting in a selection of 35 studies. They aimed to see what the research literature offered with regard to how students experience design failure and the "key elements in making failure a learning experience in design" (p. 1855) [5]. Jackson et al. identified five broad themes from the studies, three of which directly related to learning, i.e. that failure can help students learn key concepts; failure "induces thoughtfulness in problem solving," (p. 1863); and the classroom climate and messages conveyed about failure are important factors when the instructional intention is to have students learn from failures. Two other themes from the literature review were that failure has different meanings across studies and contexts, and that students have different reactions to failure experiences.

Finally, there is some evidence that failure, over success, is uniquely powerful for student learning. Andrews examined the nature of discourse around design challenges where failure experiences were more or less prevalent. Fewer failure experiences resulted in fewer opportunities to engage in diagnostic troubleshooting and learn (e.g., about critical factors for success, science ideas, etc.) from design failure. Andrews articulated: "When a design is successful, it seems like the success is attributed to the entire object, and why it is successful is not necessarily explored." (p.25). Experiencing design failure, then, represents the beginning of an opportunity to learn from that failure.

Review of the Literature: Elementary Engineering Design Failure Experiences

In this section, I examine the literature on elementary engineering design failure experiences by students as reported by teachers, reported by students, or observed during those experiences (Table 1). I include the three elementary engineering studies included in Jackson and colleagues' literature review [6, 13, 14]; the other 32 studies in the review were either about failure in engineering or technology experienced by university students (5 studies) or failure experienced by K–16 students in science or mathematics (27 studies). I also include studies that occurred after Jackson and colleagues' review period [7, 15-18], as well as studies that occurred before that review period but are worthy of note [4, 8, 19].

Perceptions about Whether Designs Failed and What Counts as Design Failure

All the studies included in this section and Table 1 included teacher reports, student reports, or observations of students experiencing design failure. The design challenges, be they about designing bridges, parachutes, packages for plants, or other technologies, resulted in at least some designs not meeting all criteria to the utmost degree or not following all constraints for one or more of their designs. This follows a definition of design failure that I have used in prior work that failure occurs "when a designed solution, or aspect of a designed solution, does not meet criteria under constraints as specified by the problem" (p. 2) [8].

Johnson, Kelly, and Cunningham organized design failure type according to stakes, extent, and referent [7]. Low stakes failures might occur out of public view (e.g., just within the team) whereas high stakes failures may be more public in nature, like a whole class testing experience. Some designs aim to be tested to different extents, including to failure (i.e., intentional failure as in loading a bridge with weights until it collapses) and other designs fail when they do not meet one or more criteria.

Johnson and colleagues also identified objective and subjective referents that help identify when failure occurs [7]. Objectively, and as in the definition of design failure provided above, failure occurs when a design does not "achieve the desired criteria within given constraints" (p. 76) . An example of subjective failure is when a design is considered to be a failure with respect to other teams' designs. This idea that teams may determine success or failure by comparing their performance with other teams was something I observed when I asked student teams about the extent to which they thought their designs failed [8]. Other subjective ways of defining failure or success in the literature included valuing one criterion, deemed the "most important criterion" (p. 21) over others when determining design failure or success [19].

Table 1

Elementary Engineering Design Failure Studies

Study	Year	Participants [b]	Setting	Designed Technology	Primary Data Sources [e]
Silvestri [16]	2023	12 Grade 3 students	Club	Pedestals, Bridges, Landing Systems, Candy Bags, Rubber Band Cars	Videos, audio, artifacts
Skinner and Harlow [17]	2022	24 Grade 4 students	Club	Parachutes	Videos, artifacts
Johnson, Kelly, & Cunningham [7]	2021	8 classrooms (Grade 3–5 students) including two table groups per classroom	School	Bridges [c]	Videos, artifacts
Lottero-Perdue & Tomayko [15]	2020	53 Kindergartners	School	Fences	Videos, interviews [f]
Simpson, Anderson & Maltese	2019	125 children ages 9–14 doing making activities	Museum (49), School (65), Club (11)	Various	Videos
Lottero-Perdue [19]	2017	29 Grade 3–5 students working in teams of 2–5	School	Bridges, Plant Packages, TarPul [d]	Interviews, artifacts
Lottero-Perdue & Parry [6] [a]	2017	74 teachers (survey); 10 teachers (interview)	School	Bridges, Alarm Circuits, Oil spill Clean-up Processes, Plant Packages TarPul [d]	Surveys, interviews
Lottero-Perdue & Parry [13] [a]	2017	254 teachers (survey); 38 teachers (interview)	School	Bridges, Alarm Circuits, Oil spill Clean-up Processes, Plant Packages TarPul [d]	Surveys, interviews
Andrews [4]	2016	13 Grade 4–6 students	Club	Hover Object and Floating Egg Transporter	Videos, artifacts
Lottero-Perdue & Parry [14] [a]	2015	108 teachers (survey); 14 teachers (interview)	School	Bridges, Alarm Circuits, Oil spill Clean-up Processes, Plant Packages TarPul [d]	Surveys, interviews
Lottero-Perdue [8]	2015	2 teachers; 7 students across 2 teams of focus	School	Bridges, Oil spill clean–up processes, Plant Packages [d]	Videos, interviews

Notes. [a] Included in the Jackson et al. literature review [5]. [b] Approximate ages for grades in the US school system: kindergarten, ages 5–6; Grade 1, ages 6–7; Grade 2, ages 7–8; Grade 3, ages 8–9; Grade 4, ages 9–10; Grade 5, ages 10–11. [c] Half of the classrooms used the Engineering is Elementary (EiE) curriculum bridge unit [20]; half used a comparison curriculum. [d] These units were from the EiE curriculum [20-24]. [e] Artifacts may include student journals/notebooks, images, or other student work during the design process. [f] Video–recorded cognitive clinical interviews of individual kindergartners moving through the design process with embedded interviews.

Whether or not students recognize design failure was a matter taken up by Skinner and Harlow [17]. They generated a conceptual framework that identified what students must know about or be able to do in order to recognize that design failure has occurred; this is the first step in diagnostic troubleshooting and thus, being able to respond to design failure [2, 11, 12]. Students must: (1) understand constraints and criteria of the challenge, (2) anticipate failure in their design process, (3) carefully observe during the design process including design creation and testing, (4) conduct fair tests and accept test results, and (5) realize/acknowledge that one or more constraints or criteria were not met. Skinner and Harlow found that the first was the largest barrier to students recognizing that their parachutes had failed. Additionally, they noted that most students had challenges with conducting fair tests and accepting test results, echoed in work by other researchers [4, 8, 19]. Having improper or inaccurate testing procedures or results does not provide students with an opportunity to learn from design failure in a meaningful way.

Simpson, Anderson, and Maltese asked about how the students noticed failure while engaged in maker activities, which included assembly tasks, construction tasks with opportunities for students to add creative ideas to what they constructed, and tinkering tasks [18]. Drawing from the mathematics education literature, specifically van Es [e.g., 25], Simpson and colleagues identified the first part of noticing failure as attending to it. One main finding from their study was "youth were more likely not to verbally attend to the failure …or provide a vague statement describing the failure … such as 'this is not staying together'" (p. 485).

Johnson and colleagues also identified reasons for failure, which they identified as a lack in student understanding about science, technology, or materials; poor craftsmanship in construction by students; or the "natural limits of [the] materials" students were able to use during design challenges (p. 79). Similarly, Skinner and Harlow identified failures due to problems with materials, not following constraints, and not meeting criteria [17].

Responses to Design Failure by Students

Researchers noted that students in their studies anticipated or predicted failure or success [8, 17], which is a sort of preemptive response to design failure. Students expressed hopes of succeeding or predictions that their design might fail, sometimes in response to teachers forewarning that design failure was a possibility [8]. Skinner and Harlow offered that "the practice of anticipating or predicting failure in conceptual designs and/or constructed prototypes was present in 90% of the examples of correct recognition of design failure" in their study (p. 203) [17].

Elementary students have varied reactions to design failure. Parry and I constructed a framework of productive/positive and unproductive/negative responses to design failures as identified by teachers [6, 14], connecting these to observed responses to design failure in video data [8]. Productive and positive responses included actions such as trying again, engaging in failure analysis, trying to improve, working effectively as a team, and using the EDP and background information to inform next steps. Unproductive or negative responses included actions such as making design changes without careful thinking, not making design changes, giving up, not working effectively as a team, focusing on competition with other teams, ignoring background information, and denying that failure has occurred (when it has). Other unproductive responses related to the first two on this list and articulated differently by others include: when students are persistent, yet with an ineffective strategy [7]; and when students do not interpret the reason for design failure [18]. Tomayko and I observed both of these, as well as a range of other mostly productive responses, in a study of individual kindergartners' design failure experiences as they created a fence to contain a small robot [15].

The framework in my and Parry's prior work also included emotions that were positioned as productive or positive (i.e., expressing a positive emotion such as excitement, not taking on a failure identity if a design failed) or unproductive or negative (i.e., expressing a negative emotion like frustration, taking on a failure identity, or appearing not to care) [6, 14]. Since this framework was generated, work by Jaber and colleagues has challenged the idea that emotions are apart from and necessarily productive/positive or unproductive/negative with respect to the learning process, asserting that they may be resources for and inherent to the learning process [26, 27]. In this reframing, for example, both joy and frustration may serve valuable roles to motivate sensemaking.

Silvestri and colleagues examined the possible actions that students take after design failure [16]. In some cases, students did not have the opportunity to discuss how they would improve their design or to create an improved design. In cases in which some improvement was possible, students could make alterations to the existing design or start over. They did so with no input, with input from peers, or with input from the teacher. A key aspect of their study was exploring the positions that students take up as they experience design failures. Students take on multiple positions including observer, tester, idea–sharer, tinkerer, and director. The context of these and other positions determine the extent to which they are productive. For example, Silvestri et al. observed power struggles and disagreements among some student teams as students within them took up different roles (e.g., one student shifting from observer to tester to director, ultimately not including other team members in the design process after failure). Parry and I also observed how team conflict can challenge productive responses to design failure [8, 14].

Factors that May Support or Inhibit Student Learning from Failure Experiences

Recall that one theme from Jackson and colleagues' review of the literature on learning from failure was that the messages conveyed about failure within the classroom—through the teacher and the overall climate of the classroom—may impact what students learn from design failure experiences [5]. Researchers have investigated factors that might support or inhibit this learning, beginning with teachers' comfort with (and normalization of) fail words, failure experiences, and supporting students as they engage in design failure and improvement, which may grow with more experience teaching engineering [6, 13, 28]. Supportive strategies reported by teachers or observed during instruction include (a) encouraging students to consider how to improve, engage in failure analysis, work effectively in teams, observe other teams' designs, and reference background information (e.g., relevant science, constraints, criteria) about the challenge; and (b) reminding students about proper testing procedures, criteria, or constraints [6, 8, 14].

Teachers may provide advice or guidance and offer evaluations of design success or failure, and the advice or evaluations they provide may be more or less directive [6]. Too much direction may hinder students' opportunities to learn from design failure [18]. Relatedly, teachers may purposefully choose to not intervene, allowing students to evaluate their own designs, and not insert their ideas into team discussions [8, 18, 28]. Silvestri and colleagues examined not only student positions, addressed earlier, but also teacher positions taken up during design failure experiences [16]. They found that teachers may position themselves as observers, questioners, or elicitors of ideas during engineering failure experiences, positions that served students well. However, teachers may overly emphasize roles as directors or evaluators when students should drive the learning process.

Other teacher–related factors to support students in their learning from design failure mentioned by Johnson and colleagues included allowing for time for improvement, having whole–class discussions about improvement, and allowing teams to engage in low–stakes testing in their groups prior to higher–stakes testing as a whole class [7]. Similarly, Simpson and colleagues said it was typical for the youth in their study to experience "micro–failures," which they described as "multiple failure patterns or iterations with in one making task" (p. 486) [18]. Johnson and colleagues also noted where the comparison curriculum in their study was potentially problematic with respect to learning from failure as it did not clearly delineate what constituted failure for a structure and did not allow for improvement [7].

Building the Literature

In their review, Jackson and colleagues called for more research involving the "close analysis of the interactions among students (or with the teacher)," suggesting that doing so would "perhaps uncover differentiating characteristics connected to variation in students' responses, even among teams in the same instructional model, and offer rich details about the students' experiences (p. 1867) [5]. As described in the previous section, since the review period, four studies have done exactly that, adding to what we as a community understand about how students experience and learn from engineering design failure [7, 15-17]. I have also included two video–based studies [4, 8], both of which are conference papers during the review period, for their contributions.

The present study also addresses Jackson and colleagues' call. It has commonalities with other studies in its focus on failure experiences, use of video data and other artifacts to examine student responses to and sensemaking about those experiences, and in its examination of factors that might support or hinder students' opportunities to learn from design failure. It is unique in that it explores teams' school–based design failure experiences across two sequential engineering design units.

Research Questions

Three research questions (RQs) guide this work and are as follows.

- RQ1. To what extent do teams perceive that they have engaged in design failure?
- RQ2. How do teams respond to and make sense of design failure?
- RQ3. What factors within the classroom environment might support or challenge teams' opportunities to engage with design failure in meaningful ways?

Context and Methods

Context

This qualitative study draws from data gathered as part of a larger efficacy study of the EiE curriculum [29]. In the present study, the primary unit of analysis is the design team. Participating design teams were in classrooms that participated in the larger study during its second year of data collection; were video–recorded as part of qualitative data collection for the larger project; and learned two science–integrated engineering units in sequence, Unit 1 and Unit 2. Unit 1 for all teams was an EiE bridges unit in which students used simple materials to construct a strong and stable bridge across two abutments [20]. Unit 2 involved the design of an electrical circuit; a package to contain a plant; an oil spill clean–up process; or site preparation to support piers for a "TarPul," a cable car that moves across wires supported by the piers [21-24]. Henceforth, I will refer to Unit 2 units as the alarm circuit, plant package, oil spill, or TarPul units, respectively. Each of the EiE units are structured in nature, emphasizing design problem solving across two iterations, with room for more iterations, time permitting. Children are provided with the problem, goal, constraints, and criteria for the challenge, as well as the allowable materials and testing procedures. They are provided with structured journals to record their engagement in the design process, including their brainstormed ideas, design specifications, testing results, and improvement ideas.

Participants

There were eight classrooms and participating teachers in the study. Each classroom was either grade 3, 4, or 5 (ages 8 to 11). The teachers were Penny, Tammy, Diane, Teresa, Cathy, Jennifer, Janet, and Joy; these and all student names are pseudonyms. Seven of the teachers self–reported their race as White; one self–reported as Asian. The teachers reported having 3 to 28 years of teaching experience (M 11, SD 7). Prior to their involvement teaching the same units one year prior for the larger study, they had not taught engineering. They received professional development through the project to teach both Unit 1 and Unit 2.

There were 61 students who participated in the study (Table 2). Demographic information was gathered from a parent survey, a student survey, and/or a form completed by the teacher.

During engineering instruction, students typically sat at between six and eight table groups and each table group contained one or two teams. The participating teachers selected the table groups and teams with minor input from the research team. This input ensured that one or two table groups, the "table groups of focus" for our research, only included students who had parental consent and who had given their assent to be video recorded and participate in post–unit interviews. There was one table group of focus in each of three classrooms and two table groups of focus in each of five classrooms where there were more researchers available to operate cameras. Altogether the 61 participating students were in 21 teams for Unit 1 and 21 teams for Unit 2, each having between two and five students. Table 3 for student teams and units.

Table 2

Student Demographics

Demographic Variable	Source(s)	Subgroups	Percentage of Participants (n = 61)
Gender	Student Survey, Parent Survey, and Teacher Form	Female	54%
		Male	46%
Race/Ethnicity	Student Survey, Parent Survey, and Teacher Form	White	66%
		Black	11%
		Asian	15%
		Hispanic	8%
		Other	0%
Eligibility for Free and Reduced Lunch	Parent Survey	Eligible	13%
		Not eligible	64%
		Refuse to answer	23%
Individual Education Plan (IEP)	Parent Survey	Has an IEP	11%
		Does not have an IEP	85%
		Refuse to answer	3%
Language(s) Spoken at Home	Parent Survey	Always use English	52%
		Occasionally use another language	11%
		About half of the time use another language	8%
		All/most of the time use another language	25%
		Refuse to answer	3%

Data Collection

Each table group of focus, as well as the whole class, was video recorded throughout each unit. The present study focuses on the video record of Lesson 4 in each EiE unit, which entails students moving through the design process, beginning with the planning part of engineering design and ending with the second design test and how teams considered how the second design could be improved. Videos of this part of the design process typically occurred over several days and lasted in total between 2 and 3 hours. Additionally, each student completed a structured engineering design journal for each unit. The journal prompted students to record the results of relevant science investigations and students' brainstormed ideas, as well as their design plans, test results, and thoughts about improvement for at least two design attempts. After each unit, each table group of focus was interviewed about each team's design experiences by a researcher from the larger study using a semi structured interview protocol. See Appendix A for the interview questions from the protocol used in this study.

Table 3

Teams and Units

Teacher	Unit 2	Unit 1 (Bridge) Teams	Unit 2 Teams
Penny	Alarm Circuit	Larissa and Mark	Larissa and Mark
		Jadon and Tyler	Jadon and Tyler
		Elijah and Iris	Elijah and Iris
		Garrett and Payge	Garrett and Payge
Tammy	Alarm Circuit	April and Gopika	Eugene and Evan
		Clarence, Ron, and Sheldon	Liz and Makayla
		Chastity and Makayla	Jimmy and Peter
		George and Leo	Chastity, Karen, and Tiana
Diane	Oil Spill	Akeera, Billy, Ivey, and Timothy	Akeera, Billy, Ivey, and Timothy
Teresa	Oil Spill	Carla, Cody, Danny, and Dean	Carla, Cody, Danny, and Dean
Cathy	Plant Package	Cassidy and Douglas	Cassidy and Kayla
		Kayla and Raja	Douglas and Raja
		Austin and Brittany	Aiden and Austin
		Aiden and Bethany	Bethany and Brittany
Jennifer	Plant Package	Brennan, Brook, Jocelyn, and Kim	Brennan, Brook, Jocelyn, and Kim
		Cole, Isaac, Olivia, Rachel, and Spencer	Cole, Olivia, Rachel, and Spencer
Janet	TarPul	Caroline and Gabrielle	Caroline and Gabrielle
		Isabelle and Samar	Isabelle and Samar
		Arianna and Noah	Arianna and Noah
		Grant and Savannah	Grant and Savannah
Joy	TarPul	David, Molly, and Zora	David, Erica, and Molly

Data Analysis

For each unit and focusing on one team at a time, Parry and I analyzed the table group video, journals, and interview to construct what we called a "design summary." (Like me, Parry was directly involved in collecting data from some of the participating classrooms and familiar with the EiE units.) Generating design summaries was a highly iterative process typical of qualitative research [30], involving watching and rewatching video segments, checking video data against journal pages, and aligning interview statements with designs and experiences. The purpose of each design summary was to describe how the team moved through the design process focusing on the following elements: (1) design 1 (D1) planning and creation, (2) D1 testing and ideas about improvement, (3) design 2 (D2) planning and creation, and (4) D2 testing and ideas about improvement. These design process elements enabled us to understand the team designs and decisions about those designs, as well as design failure experiences that were most likely to occur during testing. As we watched and re–watched video, we searched for instances in which the students reacted and responded to design failure experiences, discussed design failures with one another or with other teams or people. We transcribed these instances. Although students did not need to use “fail words” (i.e., fail, failure, failing, failed) for us to capture their responses to design failure within design summaries, we noted when they did.

We used whole–class video as needed to better understand whole–class testing or try to fill in gaps left by the three main data sources in design summaries, yet we did not need to do so often. Together, the table group video, journals, and interviews both (a) overlapped, triangulating one another as data sources especially in response to RQ2 and RQ3; and (b) offered unique insights (e.g., interviews were more reflective while group videos were in the moment). There were cases in which data sources conflicted (e.g., one design plan written in a journal but another enacted); we noted those conflicts in the design summaries. Even when we primarily drew from one data source (e.g., interviews for RQ1) in answering a research question, we could interpret evidence from that data source in the design summary's context.

Our next level of analysis involved analyzing the 42 design summaries to answer our RQs. RQ1 was about team perceptions about whether their designs failed and was answered largely using interview data within the design summaries; again, these data were situated within design performance summaries that helped us to understand students’ perceptions about whether their design failed in the context of understanding what their designs entailed. The iterative coding process began with using simple “a priori” codes, i.e., *failed*, *did not fail*, *failed partially*. Emergent codes were added to fully describe the data, which included *team disagreement* (i.e., team members disagreed about whether or to what extent the design failed) and when design failure was either *not addressed* or *unclear* in the data. I employed qualitative content analysis to identify the frequency of these codes, as well as codes in response to RQ2 and RQ3 [31].

RQ2 was about how teams responded to and made sense of design failure. Evidence to answer this question within design summaries included what the students said as they designed, wrote in their journals, suggested through gestures, and shared during the interview. This analysis built upon framework from our prior work about *productive or positive* and *unproductive or negative responses* to design failure as reported by elementary teachers, as well as how students might anticipate design failure or success [6, 8]. However, given insights from work by Jaber and colleagues mentioned previously [26, 27], I removed emotions from productive/positive or unproductive/negative categories.

In our early discussions of the design summaries, we were curious about the factors within the classroom environment that had the potential to support or challenge teams’ opportunities to engage with design failure in meaningful ways (RQ3). Analysis of the summaries led to the development and refinement of factors subcodes. Ultimately, six factors subcodes were apparent across multiple students or teams.

In the findings section that follows, I will use “we” when referring to what we found since the analysis that led to the findings was collaborative in nature. (See acknowledgements section for explanation about authorship.)

Findings

RQ1 Perceptions about Engagement in Design Failure

During the interview, interviewers from the larger project team asked teams if they thought their first or their second design failed in part or whole. Table 4 provides a summary of those responses for both units. Overall, most teams for both units experienced design failure in whole or part. Overall, 18 of 21 Unit 1 teams (86%) and 19 of 21 Unit 2 teams (90%) shared that their first or second design either failed or partially failed. Only two Unit 1 teams (10%) and one Unit 2 team (5%) shared that neither their first nor their second designs failed.

Table 4

Team responses about whether design failure occurred

	Number (percentage) of Teams			
	Unit 1 (*n*=21)		Unit 2 (*n*=21)	
	D1	D2	D1	D2
Design failed	11 (52%)	6 (29%)	12 (57%)	3 (14%)
Design failed partially	4 (19%)	6 (29%)	3 (14%)	3 (14%)
Design did not fail	2 (10%)	3 (14%)	3 (14%)	5 (24%)
Team disagreement	2 (10%)	1 (14%)	1 (5%)	2 (10%)
Not addressed	2 (10%)	4 (19%)	0 (0%)	3 (14%)
Unclear	0 (0%)	3 (14%)	2 (10%)	5 (10%)

An example of a response coded as *design failed* was in Bethany and Brittany's interview about their Unit 2 plant package design. When asked if either design failed, both replied: "First." Brittany added that "the first one was horrible," and both team members identified the specific criteria that scored poorly for D1. However, Bethany and Brittany did not address whether D2 failed (and the interviewer did not follow up), thus, their response about the second design was coded as *not addressed*.

The *design failed partially* code was used when students suggested that parts of their design failed or the design failed somewhat—or as Elijah and Iris stated—"failed, kind of." Leo shared about his and Georgia's first bridge design in Unit 1: "Ours was okay. We were safe (strong), but our stability wasn't that good." Responses indicating that a design *did not fail* were often a simple "no" in response to whether the design failed. For example, the team of Carla, Dean, Danny, and Cody all agreed that their Unit 2 oil spill design did not fail, answering "no" immediately when the interview question about design failure was posed.

Team disagreement occurred when team members shared differing ideas about whether or to what extent design failure occurred. Isabelle and Samar, who worked together on Units 1 and 2, disagreed about their first and second Unit 1 bridge design performance and their Unit 2 TarPul design performance. What follows are excerpts from the design summaries with respect to the second designs for each unit:

> Unit 1: When asked if either design failed, Samar replied "both," and explained that the first design "definitely failed" given that it "couldn't hold 100 weights." He said that "certain things failed" on the second design "because it still fell over." Isabelle disagreed that both designs failed, and said that for the second design, they could have "gotten a higher score" on cost since "we did use a lot of money on tape."
>
> Unit 2: When asked if either design failed, Isabelle said that both designs failed, and Samar shared that their second design was better in that it could hold a person, and that only some parts failed on the second design where "a lot of parts failed" on the first.

The *unclear* code was used when students seemed to address the performance of the design somehow, but students' perceptions whether or to what degree failure occurred were not clear. Two of the unclear statements for Unit 1 and three for Unit 2 were suggestions that the second design was better than the first; we did not infer the meaning of "better" with respect to failure. All these examples occurred with respect to a second design where one or more team members thought the first design failed in whole or part.

One team in the *unclear* category responded to the question about whether their first or second designs failed by sharing how their team's Unit 2 plant package design scored as compared to others in the class, i.e.: "first we got an 11, which is the second highest, and then we got 13, which is the highest." This, however, did not address design failure. Two teams were uncertain about how to classify how their Unit 2 alarm circuit design performed because the circuit worked sometimes and did not work other times. For example, Garrett shared about his and Payge's design: "Every time we tried it, um, it worked, but then when (the teacher) called everyone over—of course it didn't work."

RQ2: Responses to and Alternative Sensemaking about Design Failure

This section is divided into five subsections: (1) *preemptive responses* to design failure (and success) during the design process, (2) *productive or positive responses* to design failure, (3) *negative responses* to design failure, (4) *emotions* expressed during design failure experiences, and (5) *alternative sensemaking* about failure. See Table 5 for a summary of codes and subcodes.

Preemptive Responses: Anticipating Failure (or Success)

Three Unit 1 teams (14% of 21) shared aloud that they thought their bridges would fail. For example, as Mark and Larissa worked, Mark said, "We are totally going to fail this," to which Larissa responded, "Just do what we can," and later offered, "We're going to need a miracle!" Two Unit 2 teams (10% of 21) suggested that their design might fail. Erica offered, "We're probably going to fail on this one." Tyler shared that his and Jadon's alarm circuit could work if the switch connected in a specific way but said it would be "really unlikely it's gonna do that."

Three Unit 1 (14%) teams and four Unit 2 teams (19%) explicitly anticipated success from their designs or were optimistic about their designs. One Unit 1 team said that they were "so going to win this." Another team expressed optimism through Molly's assertion early in the design process that she wanted it to be "perfect." A Unit 2 team reflected on their second design success with: "we knew it would work!" For some teams, optimism was sparked or rekindled when teams recalled that they would have at least one additional design attempt. When Molly was constructing her team's first bridge—which she initially hoped to be perfect—she struggled to get it to hold itself up, asserting: "It's not going to work!" Immediately following this, Molly shared, "At least it's only design 1." Another Unit 1 team suggested that "tomorrow we can rebuild it" and "this is only our first design." One Unit 2 team was similarly optimistic when the teacher responded affirmatively to Austin's question, "So we're going to have another time to redesign it?" Austin offered to his partner, Aiden, "We're going to nail it the second time!"

Table 5

Coding Frequencies for Teams' Responses to and Alternative Sensemaking about Design Failure

Major Code	Subcode	Unit 1 Teams (*n*=21)	Unit 2 Teams (*n*=21)
Preemptive Responses: Anticipating Failure or Success	Anticipating failure	14%	10%
	Anticipating success	14%	19%
Productive or Positive Responses to Design Failure Experiences	Failure analysis and/or improvement	62%	43%
	Persistence	14%	19%
Unproductive or Negative Responses to Design Failure Experiences	Misdirected failure analysis	5%	14%
	Blame	19%	5%
Emotions	Satisfaction	10%	5%
	Joy	10%	5%
	Disappointment	10%	10%
	Frustration	19%	10%
Alternative Sensemaking about Design Failure	Comparison and competition	24%	19%
	Most important criterion	19%	10%
	Grappling with fail words	10%	0%

We also found evidence within Unit 2 of three teams (14%) expressing their anticipation of or hope for a particular numerical score for their design. Two of the teams, team Raja and Douglas and team Cassidy and Kayla, sat at the same table. Douglas and Raja said to Cassidy and Kayla that they were going to get "a 13" – "the highest score we can get" to which Cassidy and Kayla replied, "Yeah, we are too—13; we were going for 12."

Productive or Positive Responses

We identified two productive or positive responses in the data. These included: *failure analysis and/or improvement* (62% of Unit 1 teams; 43% of Unit 2 teams) and *persistence* (14%; 19%).

We gathered direct evidence of 13 Unit 1 (62%) and 9 Unit 2 teams (43%) *analyzing design failure* or *planning for improvement*. For Unit 1, most of the examples we identified were about failure analysis (10 of 13) and fewer (5 of 13) addressed improvement. For example, Douglas identified the piers he and Cassidy had placed on the outside of the bridge as having failed: "they almost broke straight off" and "didn't help keep the weights up." Jadon and Tyler commented on an improvement they made, putting straws on the sides of their bridge "like railings," which improved their bridge stability. Georgia and Leo were one of two teams that discussed both failure analysis and improvement. In the interview, Georgia offered the following about their second bridge, "I put the cones [piers] all different because of … how we did it before, our bridge test failed. It just failed." She demonstrated with her hands to show how she arranged them so it "can support it." Leo added that "they [the piers] weren't separated enough (the first time)." For Unit 2, 7 teams addressed failure analysis, 7 mentioned improvement, and 5 teams addressed both failure analysis and improvement. In their design to clean up oil spills, the team of Ivey, Akeera, Billy, and Timothy identified the rubber band they hoped to contain the oil spill in Design 1 as problematic. They decided to use yarn instead in their second design to try to contain the oil more effectively. Ivey explained: "We all had the same idea. We were all thinking yarn and then Billy just said, 'Why don't we do yarn?' and then we all agreed."

Seven teams, three in Unit 1 (14%) and four in Unit 2 (19%), demonstrated evidence of *persistence* by acknowledging their own effort or persisting despite multiple difficulties. Raja offered at one point, "I can't believe no one gave up yet," referencing his own team's persistence and perhaps that of other teams. Our design summary described Gabrielle and Caroline's work together: "Gabriel and Caroline … persisted through multiple failures and collaborated to try new ideas, with little success." Persistent troubleshooting and improvement by Liz and Makalya were evident in the following Unit 2 design summary, referencing notes from the team video:

> After reviewing the contractor form, Makayla and Liz began to troubleshoot and improve their circuit, by adding a buzzer and another battery. Makayla realized the battery connections were switched and turned it around to get the polarity correct. They created a circuit without the switch and successfully made the buzzer work but not the bulb. They isolated the bulb and were able to get it to light but still struggled with both elements working. They then began to include the switch, and at the teacher's suggestion, removed the buzzer to first test if the switch connection works. They pushed down on the connection end to make the bulb light … They continued to troubleshoot … to make a better connection and were finally successful.

Unproductive or Negative Responses

We identified two unproductive or negative responses. These included *misdirected failure analysis* (5% of Unit 1 teams; 14% of Unit 2 teams) and *blame* (19%, 5%).

While failure analysis and subsequent improvement are often productive, that is only the case if the failure analysis upon which improvements are based is careful. We noted occasions of *misdirected failure analysis* in which teams responded to design failure with improvement ideas that were disconnected from a robust failure analysis. We noted a case of this for one Unit 1 team (5%) where reasons for design failure written in journals and described by students (i.e., that columns need to be more spread out or that paper was weak) were not in alignment with some key problems that we observed in the video (i.e., that the columns were different heights, fell over, and were not affixed to the deck). We suspect that there were more cases of misdirected failure analysis for the bridges unit that we did not capture; however, this may be due to other issues with the testing process for this unit described in the RQ3 findings section. Three Unit 2 teams (14%) with misdirected failure analysis were designing alarm circuits; in this unit, there were multiple possible causes of circuit success or failure. In one case, Elijah and Iris tried persistently to fix the switch in the circuit, assuming that the switch was not closing with enough force to complete the circuit, only to figure out after they agreed their switch was not successful that they noticed a wire missing from the circuit. Iris shared, "we didn't see the wire wasn't attached to the battery." Elijah continued, "So it was an open circuit. So, our thing might have actually worked."

There were four Unit 1 teams (19%) who *blamed* a team member for the failure of their respective bridges. For example, Austin blamed Brittany for not placing the weights on the bridge correctly during testing and Makalya and Chastity blamed one another for their bridge failure. During the interview, Makayla offered that "our second bridge failed … really fast because of *someone* [emphasis in original] who did the index cards." Chastity retorted, "Well I'm sorry. I wanted to make an edge that the car wouldn't go over … Maybe it was your fault for putting paper clips on it." We observed just one Unit 2 team (5%) responding to design failure by blaming. In this case, they blamed another team who served as "contractors" who tested their alarm circuit design. Payge shared and Garrett agreed that "the contractor built it wrong." Payge and Garrett did not consider that their schematic for the contractors could have been improved.

Emotions

There were four emotions that we observed in the data. These were *satisfaction* (10%; 5%), *joy* (10%; 5%), *disappointment* (10%, 10%); and *frustration* (19%, 10%).

We were witness to two Unit 1 teams (10%) and one Unit 2 team (5%) expressing their *satisfaction* after improving between their first and second designs. By satisfaction here, we mean a mix of pride and happiness. Austin was pleased that his team made it through one phase of the bridge testing process and Tyler, on another team, said "we got one more weight. It was so good!" and later, "Our bridge got better." Mid–way through the Unit 2 test of her team's TarPul design, Molly happily offered, "It's going to hold a person!!"; the first design had not been able to do so. Her teammates shared their excitement with me: "We held 11 [weights]!"

Joy–filled responses to design failure included laughter or humor, which occurred within two Unit 1 teams (10%) and one Unit 2 (5%) team. By this, we mean that the students found a way to see humor in the failure, not to dismiss it but rather to make light of it. Larissa laughed when she said that she and Mark would "need a miracle" for their bridge to work, and Mark laughed when responding "um, no" to a question in the interview about whether their first bridge design was good. All the interviewees in the interview that included Savannah and Grant laughed when Savannah responded to a question about what did not work well in their first design with, "pretty much everything." David, a member of Unit 2 TarPul design team, did so when his team realized the TarPul could not hold enough weights to represent an adult, and thus, was ineligible for scoring. He joked that the TarPul could hold a baby (but not an adult).

Evidence of *disappointment* was difficult to capture with certainty. There were two Unit 1 teams (10%) and two Unit 2 teams (10%) where we noted strong evidence of this emotion. Although Tiana refused help from her teammates during her second unit alarm circuit design, she expressed that she was "starting to feel sad." Also, Aiden was disappointed when his and Austin's plant tipped over during the shake test for their plant package design. After the test, Aiden came back to his desk, put his head in his hands a few times and then said to Austin, "Guess what? This is going to be our score [makes a zero with his hands]: a big fat zero." Aiden also shared this with his teacher, Cathy, who replied:

> Right, so instead of wasting your time sitting and getting discouraged, they [engineers] don't get discouraged, what they do is they think … and they ask questions, they go look and figure out, well, how can we make it so that it doesn't tip when it shakes.

Like a teacher in Jaber and colleagues' study, Cathy was "tuned into" (p. 149) Aiden's frustration, and used it as motivation to continue in the design process, analyzing the design failure and improving accordingly [26].

Frustration was like disappointment in that it was difficult to capture and may have occurred more often than we were able to capture. Four Unit 1 teams (19%) had team members who clearly experienced frustration during their bridge design. Both Spencer and Olivia, who worked together, voiced their frustration (also evident in their actions) in different ways, with Spencer saying "this is really hard" and Olivia saying "I'm sleepy. I don't want to write anymore." Austin, working with Aiden on his bridge design, at one point threw the bridge deck down and said: "Oh my God! I don't even care anymore. So frustrating!" Two Unit 2 teams (10%), both designing alarm circuits, had notably frustrating experiences; one student was particularly frustrated when other teams' circuits were successful but his team's circuit was not. We did not observe teacher intervention during these occasions. Certainly, these frustrations could have compelled some students to continue working and trying; yet are data were not able to clarify whether this was the case, if other factors contributed to their ongoing learning (e.g., the need to complete journal pages), or if the frustration students seemed to experience reached a level that may have inhibited students' sensemaking during design.

Alternative Sensemaking about Failure

Students made sense of design failure experiences outside of the context of comparing design performance to criteria and constraints followed by improvement. We identified three ways in which students engaged in alternative sensemaking about failure: *comparison or competition* (24% of Unit 1 teams; 19% of Unit 2 teams); *most important criterion* (19%, 10%); and *grappling with the word, "fail"* (10%, 0%).

Some teams' determination of success or failure seemed more so about how their design performed in *comparison* to or in *competition* with other teams rather than in comparison to criteria (for D1 or D2) or their own past design performance (i.e., comparing D2 to their own D1). Five Unit 1 teams (24%) and four Unit 2 teams (19%) fell into this category. For example, Bethany and Aiden waited to test their D1 bridge after all the other teams, saying that she didn't mind waiting so that "ours is the one that can be good" and so they could "see what others have done to [make] ours and make sure it doesn't collapse like theirs." In another class, Jocelyn, excited by her and her teammates' first design bridge planning, offered "So cool! We're so going to win this!" This win/lose language was apparent in their second design, as well.

Two of the Unit 2 teams who were designing TarPuls sat across from one another at a table and regularly compared one another's scores throughout the design challenge. In another classroom where students were designing plant packages, score comparisons were also prominent, with Raja sharing that his team's first design scored an 11 "which is the second highest" and their second design scored a 13 "which is the highest." A student from another team offered that the class all received the same score for the second score and Raja's teammate, Douglas replied, "we were all tied." In another case, a student on a team was so invested in having his team's TarPul hold the most weight that what he tested (even with the objections of his teammates) was not representative of an actual site along the river with certain site characteristics.

In Unit 2 classrooms, an authority figure encouraged this sense of comparison or competition. In one, we observed a parent helper interject herself into her son's team a few times, sharing that their team is doing better than other teams in the classroom. For example, at one point, she said, "maybe you're at $7.00 but some people are at $9.00 or $11.00, so if you need more, don't worry." In another classroom, the teacher informed Liz and Makayla's team that they were only the second group in the class to make the circuit work.

For all design challenges in the curriculum, there are multiple criteria that collectively determined an overall design score. Teams attended to these as they scored their designs in their journals and many teams discussed how their designs attended to multiple criteria to various levels of success. Some teams, however, seemed to favor one criterion more than others. We saw evidence of this with respect to both the bridges unit (Unit 1) and the Unit 2 TarPul unit, both of which involved testing strength to a point of failure (among other criteria). For four Unit 1 teams (19%) in three different classes, strength—the number of weights the bridge could hold—seemed to be the *most important criterion* as compared to other bridge criteria including stability and barge accessibility. For one team, this focus was in part related to comparison/competition because they were trying to have their bridge hold as many weights as another team in the class.

For the TarPul unit, two teams (10%) in two different classes focused on having their second design model TarPul hold more weight, which represented the TarPul being able to carry more people across a river during the rainy season. They were not as concerned with other criteria. This was in part due to the curriculum, which specified that a certain number of weights needed to be held to represent at least one adult person, else the design received a "no score" overall. Caroline and Gabriel received a score for both their first and second designs; their first design had an overall higher score but held fewer people. Reflecting on their design performance in the interview, Caroline offered: "well, actually, I think the first one failed. The second one actually held more people." 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.

For two Unit 1 teams (10%), when asked in the interview if their first or second bridge design failed, we noticed *students grappling with a fail word*. During the interview with one of those teams, Dean shared, "Well, it didn't exactly fail … I wouldn't say they failed because (it) had good stability, and the barge was able to pass through, but it, its strength wasn't enough." Later he continued: "So it didn't exactly fail … it just, it didn't have enough strength to hold up that many [weights]." His teammate, Carla, added, "It didn't meet our goal." Later, she explained:

> I wouldn't say it failed, but … it wasn't … as successful—So in a way it did fail but it didn't fail but—we needed to put more piers, so it was our fault it sorta failed, and then the weights came down too fast on the second one, so like, it was its fault and our fault.

In general, the team agreed that the testing did not go as they'd planned but were uncertain about whether it failed or not. Fault of the design or the team may play a role here regarding whether failure occurred, but it's unclear exactly how. In another team, Bethany offered that her and Aiden's bridge "fell" rather than "failed." We did not observe a similar apprehension about using fail words by teams with respect to Unit 2.

Despite what seems like discomfort by some students, we tracked students' use of fail words (i.e., fail, failed, failing, failure) in the data we collected, and observed the use of fail words by 11 Unit 1 (52%) and 9 Unit 2 (43%) teams. For example, during Unit 1 testing, Douglas was disappointed that he and his teammate, Cassidy, couldn't try to see how many weights the bridge would hold. Cassidy said they couldn't since they failed the last stability test, saying, "No. It's a fail on all our stuff." In Noah's Unit 2 journal, in a letter to village elders about where to place the TarPul, Noah wrote: "The site we have chosen is E. We have tried site D but failed." Most uses of fail words were in the interview responses in response to questions about whether or to what extent designs failed.

RQ3. Factors that May have Affected Design Failure Experiences

Several factors emerged from our analysis of the data that may have impacted the way in which students working in teams respond to design failure experiences. These factors and their potential to support or impede productive learning from design failure are listed in Table 6. It is important to note that we cannot know with certainty that these factors did support or hinder students' design failure experiences; yet our data—coupled with understandings from the literature about things such as the importance of testing and design failure experiences—suggest that they were likely to have affected these experiences. Additionally, we recognize that teachers have a great deal to juggle as they teach engineering and may intentionally or unintentionally trade off maximizing children's design failure experiences for logistical reasons or to respond to student needs.

Altered Criteria or Testing

We observed four cases across three classrooms where teachers changed the criteria and thus how teams experienced design failure. The first two cases altered the way that cost operated as a criterion—in one case making it easier for overall design success and in the other making it more difficult. The other cases involved changes to testing duration and the timing of materials provided to students, both of which made it difficult to compare D1 and D2 and interpret design success or failure.

Table 6

Factors that May Support and/or Impede Productive Learning from Design Failure

Factors	**May support** student learning from design failure experiences	**May impede** student learning from design failure experiences
Alterations to design criteria or testing processes		✓
Other testing challenges		✓
Journal inconsistencies with respect to scoring		✓
Mid–create testing and failures	✓	✓
Interventions by teachers or other adults	✓	✓
Intra– and inter–team dynamics	✓	✓

One Unit 1 teacher, Penny, shared with the class that cost would not be considered in the bridge design. This criterion removal may have contributed to students making not as careful decisions about which materials to use in their bridge design. For example, after hearing that cost was not included as a criterion, Jadon's response was, "Let's just get one of everything!" His teammate responded: "Let's get two!" All four teams that we observed in this class used materials liberally and without documenting their material use in their journals.

During Unit 2, another teacher, Cathy, told students creating second design plant packages that any materials they added to their second design would be added to the total cost of their first design. Thus, students were unable to start over with an entirely new design with new materials starting at $0.00. For those who followed this suggestion, their D2 cost scores were artificially elevated.

In a Unit 2 plant package challenge, the curriculum specified that plants in plant packages were to be tested over about three days without watering. However, in teacher Jennifer's class, D1 was watered during the testing period but D2 was not watered during the testing period. This caused confusion about failure by students who blamed their "disgusting" and "droopy" (according to Kim, Brennan, and Brooke) plants in D2 packages on a lack of watering. The teams in this classroom were unable to compare the plant health of D1 and D2 due to this change in the testing process.

The curriculum specifies that for the alarm circuit unit, the students construct a switch and circuit that, for the first design, turns on a light when the switch is operated. If that is successful, then a buzzer is added for the second design so that when the switch is operated, both the light goes on and the buzzer sounds; adding the buzzer requires that a parallel circuit be created. Tammy, perhaps inadvertently, handed out both the buzzer and the light for the first design, which led to a high rate of failure and frustration across the teams who struggled to manage both. Realizing the mistake, the teacher said that they could remove the buzzer from the second design, but some teams still attempted both.

Other Testing Challenges

For students to be able to make informed decisions about how to improve their designs after design testing, it is imperative to have a clear, consistent, and accurate testing process. While design testing sometimes was clear, consistent, and accurate, there were instances in which it was not, making design failure analyses and knowing how to improve difficult. Our observations helped us to identify challenges within testing processes for the Unit 1 bridge unit and the Unit 2 plant package and TarPul units.

Challenges related to Unit 1 bridge unit testing were related to (1) weight testing, (2) the bridge–abutment connection, and (3) car quality. The bridge unit teacher materials state that the strength of the bridge is to be tested "by placing a plastic cup on the center of the span and adding weights to the cup, one at a time …" (p. 122); the curriculum defines "span" as "the distance between supports" (p. 82) [20]. That said, we observed bridge testing in three classrooms where teams either placed weight cups directly on top of piers or distributed the weights all over the deck of the bridge rather than in a single cup in the center of the span. Additionally, the curriculum specifies that bridges may rest upon but not be affixed to the abutments, yet there were teams in three classrooms that affixed their bridges to the abutments with tape. Finally, wind–up toy cars were used to test the stability of the bridge. However, their quality was poor, and they often veered off in different directions, confusing students regarding whether veering off was a fault of the cars or the bridge. We have direct evidence of this for one team but observed the veering of cars outside of this team. A member of this team, Ivey, suggested that if a car drives off a bridge like for their D2 bridge, "that's not our fault." However, her teammate, Billy, suggested that if the car falls off because they don't have railings on the bridge, "I mean, that would be somewhat our fault because we didn't build a high enough railing and a long enough railing."

Challenges related to the Unit 2 TarPul unit were related to: (1) the failure line, (2) pole holding, and (3) compaction. To prepare the TarPul site, each side of the model river (two cups of soil) could be compacted to a particular measured amount (0, ¼, or ½ inch) that teams needed to account for as they scored their designs. More compaction is more expensive and "counts against" the overall score; however, as students learn earlier in the unit, more compaction makes for a stronger TarPul. For two teams in particular, compaction was not measured or accounted for accurately. After compaction, the TarPul's support poles are placed into each side of the model river and the model TarPul carriage is connected to those two poles by a string. Weights (washers) are dropped into the carriage during testing, when the cups can be held but not the poles; we observed three teams holding the poles. The number of weights the TarPul can hold is determined by how many weights cause the model carriage to dip below a "failure line," a piece of masking tape marking the top of soil on one side of the river to the other. Two teams tested their TarPuls with no failure line in place. Once these teams added the failure line, they both continued to add weight after the carriage dropped well below the line, as did another team.

Inconsistencies in Scoring within Journals

One of the engineering journals' purposes was for each student to document their scoring and testing performance. Instruction was typically broken up over multiple days, so the scoring of D1 may have been a day or more prior to students being asked to consider improvement ideas and develop D2. However, there were multiple cases in which journals within a team presented differing accounts of scoring or when what was accounted for in a journal did not quite match with testing.

In Unit 1, we observed cost– and materials–related discrepancies in seven of the teams (33%) who were expected to keep track of materials and cost, suggesting some inaccuracies in accounting related to the cost criterion; for Unit 2, we observed this in 2 of the teams (10%). For example, in Unit 1, Leo's materials list did not account for the 11 extra inches of tape he and his teammate, Georgia, decided to use; Georgia did account for this in her materials list and total cost calculation; video review confirmed Georgia's accounting. On another team in another class, Bethany and Aiden only calculated materials they added onto D1 to create D2, not including the existing D1 materials they were also using. This was less a discrepancy across team members and more so an accounting mistake. In one Unit 2 team, there was disagreement in the cost of their D2 plant package within the team journals. Kim wrote $6.50, Brooke wrote $11.00, Brennan did not have a complete cost page, and another team member, Jocelyn, had been absent for D2 and did not have a complete cost page either.

There were discrepancies regarding other scores, specifically weight and stability, in Unit 1 for six teams (29%) and in Unit 2 for one team. For example, on the team of Isabelle and Samar, Samar's strength scores were a point higher than Isabelle's based on his assessment of what the bridge "could have held" if the cups holding the weights "wouldn't have fallen." Isabelle tried to explain that the cups falling was a failure, but Samar was adamant that it was not the fault of the bridge. Thus, their scores differed in their minds and journals.

Mid–Create Testing and Failures

Another factor that may have impacted teams design failure experiences during formal testing was that we observed teams doing forms of testing as they created their designs (e.g., putting objects on their bridge, doing a quick shake test of their plant package). We observed this in six Unit 1 teams (29%) and three Unit 2 teams (14%). These "mid–creation tests" often led to changes to the design prior to the formal testing process. For example, after conducting a mid–create shake test, Aiden and Austin added more cotton balls to their plant package design to better protect the plant during the shake test. We see these tests as smart ways that students are learning and creating their designs or what Johnson and colleagues referred to as low–stakes opportunities for students to improve their designs [7]. That said, mid–create testing is not always possible (e.g., in the Unit 2 oil spill unit) and students may not remember to document mid–create alterations in their plans to accurately track costs and decisions in their journals.

Failure Responses and Interventions by Teachers and Other Adults

Across both units, we witnessed teachers in three classrooms explicitly encouraging design teams after they had experienced design failure. One of these cases was when teacher Jennifer said to a Unit 1 bridge design team: "What's awesome is that you can come back tomorrow with some new ideas." Also, teacher Joy, after witnessing Molly, Zora, and David's bridge that, as Zora noted, "can't even hold a straw," responded: "Oh dear. That's okay. As engineers, we keep working and reflecting." Teacher Cathy also mentioned what engineers do in response to Aiden and Austin's disappointment after their plant package shake test, sharing that instead of getting discouraged, engineers "ask questions … [and] go look and figure out … how we can make it so that it doesn't tip when it shakes." Austin replied, "So we're going to have another time to redesign it?" and Cathy responded: "Absolutely."

There were times when teachers positioned themselves more so than the students as the design evaluators, determiners of scoring, and analyzers of failure experiences. We observed this in four classrooms. In Unit 1, we observed this primarily in one teacher, Cathy, who frequently made evaluative comments about students designs, for example, calling one team's bridge a failure and "done" when the teacher observed that the bridge was elevated above the abutments. In another case, the teacher said there was too much sagging in a bridge and stopped the test but did not refer to a specific measure of sagging that was or was not allowable. Cathy also told the class in Unit 2 that the "perfect score" for the plant package was a 13 since it was impossible to create the package with the bottle or carton alone; this assumption is an arguable constraint and may have inhibited students' improvement. Another teacher, Joy, stated that if Molly, Zora, and David's Unit 1 bridge cannot hold a straw (as Zora shared), it will not be able to hold a cup of weights. In Unit 2, teacher Janet did some design evaluation, at one point intervening on a disagreement about how much weight the TarPul the team was testing held; Samar said 10 and Isabelle said 6 because she said that Samar had been holding the poles (Samar denied this). To resolve the dispute, teacher Janet asked the team to record 10 weights.

Relatedly, we observed cases in which authority figures in the classroom—teachers, a researcher, or a parent—were directive with respect to improvement ideas after design testing and failure experiences. In teacher Cathy's classroom during Unit 1, the teacher and researcher suggested that teams might consider using columns in their D2 bridge designs since they had learned about their strength and stability in prior investigations; this was leading, done since no teams had used columns to this point, and effective in encouraging the use of columns. Teacher Jennifer provided very specific suggestions to teams about how to improve their plant packages (i.e., to use paper towels since nobody else was using those).

A parent helper in teacher Jennifer's classroom was highly directive to her son's design team for both Unit 1 and 2 when students were to be reflecting on D1 (to do failure analysis and consider improvements to include in D2). At one point, the teacher inserted their own failure analysis, saying: "your columns are great, [but] it seems like they weren't very stable." The mom replied to this, "You have to think of another way to attach them [the columns] maybe." Then the teacher said, "Or maybe try an arch." Later, the mom pushed the arch idea: "Have you guys thought about an arch? … because I think the columns are not going to work." [The team next to them used an arch.] Shortly thereafter, the mom came back from looking at other teams and told them what to do for D2, i.e.: use paper railings by folding up deck sides, use a cluster of columns, put weight cup on top of that cluster, add more paper to the deck. In Unit 2, the mom interjected again when she reassured the team that they were doing better than other teams and could afford to spend more since other teams had spent more.

Intra–and Inter–Team Dynamics

In many cases, we observed teams working well enough together: dividing up the tasks, co–constructing, sharing ideas, etc. For two Unit 1 teams (10%) and two Unit 2 (10%) we observed especially positive team dynamics and in one case, a team that commented on their growth as teammates. The two Unit 2 examples include not only our observations of positive team dynamics, but students' commentary on that topic. Towards the end of their plant package creation Raja and Douglas had the following exchange:

> Raja: We're doing good at this. This is the most I've actually compromised with someone.
> Douglas: Same with me.

During their interview, Jadon and Tyler both agreed that their teamwork improved for their second alarm circuit design. For the first design, "[Tyler] was doing all the building" and "[Jadon] was like, writing about it." However, for the second design, they both contributed to these tasks; we confirmed this through our video review of their design process.

There were more teams who were notable for experiencing team conflict. This included three Unit 1 teams (14%) and seven Unit 2 teams (33%). Across these 10 teams, we identified four themes related to team disfunction: (1) displays of anger towards another team member (6 teams); (2) a team member or members monopolizing hands–on work and decision making, while others were left out (3 teams); (3) frustration over one or more team member's lack of contribution (2 teams); and (4) team members working in parallel rather than together (2 teams). Two teams demonstrated three of these four themes. The most bickering that we witnessed was within the team of Bethany and Brittany, paired together for their Unit 2 plant package design challenge. The following from the design summary summarizes and exemplifies the nature of their interaction throughout the challenge:

> We [researchers] observed the following patterns [for this team]: (1) a struggle for who was to create the written parts of the package (even though Brittany won rock–paper–scissors twice to determine who would lead this, Bethany still felt that she was the rightful writer); (2) anger, largely directed from Bethany to Brittany; and (3) Bethany not really accepting Brittany's ideas or listening to her. Even Austin and Aiden, sitting near these two, tried to intervene. At one point, Austin said, "So work as a team while you can and you're not mad …" The following exchange was at the peak of the anger between Bethany and Brittany, shortly after Austin's suggestion:
>
> > Brittany: Wait. Are we still doing windows [for the plant package]?
> > Bethany: Whatever you feel like, Brittany [angrily]. I don't care anymore.
> > Brittany: I asked you a question. Are we or are we not?
> > Bethany: We have to or else we don't get a good score.

Additionally, we noticed many occasions in which Brittany and Bethany were not communicating and, rather, contributing their own ideas and features to their plant package design, evidence of parallel work.

We did not observe evidence of Brittany or Bethany being frustrated by one another's lack of action (our third theme related to team disfunction), but this theme, as well as anger and parallel work were all evident for Evan and Eugene. Some notes from the design summaries included that Evan and Eugene "sat at opposite ends of the table despite the teacher stopping several times to tell them to work together" and that "the teacher eventually came over to remind the boys to work together and to sit next to each other." We described Eugene as "clearly frustrated" and Evan as "quiet." The final note about this team in the design summary was as follows.

> During the troubleshooting part of the lesson, the teacher overheard the other team at their table chastising Eugene for how he was speaking to Evan. She moved Eugene to another table, and asked Evan if he wanted to work with a group or alone. He stayed at the table and worked alone from that point forward.

We see challenging team dynamics like that of Brittany and Bethany and Evan and Eugene and others as inhibiting what students might be able to learn from engaging in engineering design, including in collaboratively and productively responding to design failures.

In addition to issues of comparison and competition across teams, there were cases within Unit 2 where teams interacted with one another in other ways. We have already shared one example of an intersection between intra–team and inter–team dynamics, i.e., in the Unit 2 plant package design challenge when Austin urged Brittany and Bethany to work more productively together. The example we shared of this was not the only case of this happening; Aiden also encouraged Brittany and Bethany that the score they were going to receive was "not the end of the world as long as you work together."

Another pair of student teams had a constant back–and–forth engagement where the teams not only compared scores but also argued with one another about proper testing procedures (many of which were not followed) and whether a negative score is possible. There were also some negative dynamics between Grant, on one team, and Noah, on another. At one point, Noah accidentally knocked over part of Noah's team's apparatus and Grant laughed. Later, when Grant and Noah's teammate, Arianna, convince Noah that a negative score is possible, the following exchange transpires:

> Noah: Ohhh.
> Grant: I thought you screwed up again. I'm joking. I'm joking.
> Noah: Even smart people make mistakes. And I'm not the only one.
> Grant: I'm joking, I'm joking. Everyone is smart here.

Our interpretation of the banter between Noah and Grant was that it was largely competitive and negative.

Conclusion and Discussion

I began this paper by asserting that design failure is an important part of engineering education not only because it is part of the normal process of engineering design, but also because it represents an opportunity for learning [1-3, 5]. RQ1 sought to explore the extent to which teams of students experienced design failure from their point of view in part to establish the frequency with which this aspect of engineering design occurred for students. RQ2 explored the ways in which teams responded to design failure experiences and made sense of those experiences. Finally, RQ3 examined the factors within classrooms that may support or impede student learning about design failure. This work overlaps with research in elementary engineering education on design failure and aims to expand it.

In my past work and in this study, I have found it to be interesting to ask student teams to reflect on whether their designs failed in whole or in part [19]. One major conclusion from this inquiry is that design failure occurred within most design teams for both the Unit 1 bridges unit (which included collapse as a measure of design failure) as well as for the Unit 2 units where failure to meet design criteria was prevalent but perhaps less momentous. Another conclusion that is less discussed in the literature is that within a design team, not all team members may agree on the extent to which their designs failed. This relates to my and Johnson and colleagues' findings about students' sensemaking about design failure that may be objective or subjective, with subjective assessments of design failure including how well a design performs against other teams' designs or with respect to a criterion that a student or team values more than others [7, 8, 19]. It is also relevant to the hesitation of some students about applying fail words to the performance of a design, likely due to the loaded nature of the word in schools and the extent to which fail words are normalized (or not) within classrooms during engineering instruction [6].

Team responses to design failure in this study were similar to those observed in other studies [6, 8, 14]. Some students anticipated design failure or success prior to design testing, as did Skinner and Harlow and as in my prior work [8, 17]. Most of these cases in the present study were broad statements like "we're totally gonna fail this" or "we are so going to win this," less about predicting where or how a design might go wrong and more about a binary fail–or–not–fail prediction. However, one example from this study was different and worthy of note, i.e., Tyler's observation that his and Jadon's alarm circuit was unlikely to work in the test since there was a very specific way in which the circuit needed to connect. This thoughtful prediction is one that begins to consider diagnostic troubleshooting, albeit before testing. Considering this, I wonder if more time should be spent in engineering design experiences having students make predictions about how their designs might perform.

Responses to design failure after it occurred were consistent with those in the literature [6-8, 14-16, 18]. Productive responses included engaging in failure analysis and/or improvement and persisting. Other codes in my and Parry's previous work were also evident and were captured elsewhere in our findings. For example, student use of background knowledge to inform next steps was included in our codes for productive failure analysis and improvement, and positive team dynamics was included as a factor that supported learning from failure (in response to RQ3). Unproductive responses included misdirected failure analysis, which may overlap somewhat with Johnson and colleagues' identification of students persisting albeit with an ineffective strategy [7]; Simpson and colleagues' observation that they youth in their study infrequently interpreted the reasons for design failure [18]; and my and Parry's prior work that observed students making hasty changes to their design without careful thinking or ignoring background information that could productively inform design changes [6, 14]. Further, team focus on competition with other teams more so than on the improvement of each team's design was an unproductive response evident in the present study and in past studies [6, 14].

With a new lens from work by Jaber and colleagues [26, 27], I recast student emotions outside of necessarily productive/positive or unproductive/negative categories. Student improvements in the design process often resulted in joy and satisfaction. Their disappointments and frustrations during design failure experiences represented the potential to motivate continued effort. We did observe some emotions that were less likely to be productive, especially the act of blame which might arise from resentment or anger, and anger itself, which was apparent in some team dynamics.

Although I did not frame this study as Silvestri and colleagues did with respect to exploring the positions students take up as they respond to design failure [16], certainly, the teams in this study took on various positions. They observed, tested, shared ideas, and tinkered, and thus were observers, testers, idea sharers, tinkerers, etc. Some also took on positions as directors, which at times was associated with team conflict.

The factors identified in the present work that may support and/or hinder students' opportunities to learn from failure have implications for how engineering education can better support students' opportunities to learn from failure. Regarding supportive factors, teachers can help students learn to normalize design failure, interpret design failures during failure analysis, and consider how they might improve [5-7]. Teachers can also pose questions and elicit ideas from students to encourage them and focus them on diagnostic troubleshooting and improvement in response to design failure [2, 8, 11, 16]. Teachers may choose to allow for low–stakes testing [7], micro–failures [18], or what we called mid–create testing, allowing for students to learn from failure prior to higher stakes testing opportunities.

Teachers or other adults in the classroom may act in ways that decrease student opportunities to learn from failure. They may direct responses to design failure, as Silvestri and colleagues observed [16], or even emphasize competition among teams. Other teacher decisions to alter criteria or testing processes during the design process may have contributed to lost opportunities to respond to design failure in meaningful ways, at times confusing whether design failure occurred or making it difficult to compare between D1 and D2. I mean not to disparage teachers in making these observations! I recognize that teaching engineering design challenges in an elementary setting is complex, and the intentions of our teacher participants were to give their students the best experience possible given their unique classroom contexts.

Whether in professional learning experiences or in pre–service teacher education, teacher preparation to teach engineering design challenges should address the importance of supporting but not directing students' design and design failure experiences, facilitating consistent and accurate testing, and emphasizing a focus on learning from design failure experiences over competing with other teams. Also, curricular materials should be as clear as possible with respect to criteria, constraints, testing procedures, etc., and provide guidance about ways to support student learning from design failure experiences.

Another finding from our study was that journals may or may not be accurate representations of scoring or design decisions across students within a team. Teacher oversight of journals—perhaps including a review of team plans (within each student's journal) before material collection and a review of team scoring prior to subsequent design—would be helpful. A clear downside is that this creates more work for the teacher.

Finally, some team dynamics likely contributed to productive learning from design failure experiences and other team dynamics clearly did not. Some studies, have examined elementary student teams while doing engineering [32-35] or experiencing design failure [7, 8, 16]. However, the literature focused specifically on team conflict in elementary design teams is relatively sparse and few have focused on team dynamics during design failure. More work is needed in this area.

Like any study, this one has limitations. Given the way in which we included focus group tables and teams within those tables, we cannot claim that the 21 Unit 1 and 21 Unit 2 teams are representative of all upper–elementary engineering design teams. Also, the two researchers who co–generated and analyzed design summaries, Parry and I, were not present in all the classrooms as data were collected. Rather, each of us were present to collect data in one classroom, respectively. There are limitations in what can be interpreted from video data, journal data, and interviews alone. That said, these sources triangulated one another in helping to compose an accurate and complete as possible story of teams' engagement in the design process and their design failure experiences.

I aim to continue to study student responses to engineering design failure, moving into the secondary space, which according to Jackson and colleagues [5], needs attention by the research community. In doing so, we also aim to focus in on team dynamics as an influential factor in how teams respond to design failure in more or less productive ways.

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. I would like to thank the participating teachers and students, as well as the E4 Project team for their support during data collection, especially Liz Parry, Christine Cunningham, Cathy Lachapelle, Jonathan Hertel, and Chris San Antonio-Tunis.

A special thanks to Liz Parry for her assistance with analysis for the present study and for being a thoughtful partner over the years in this work about design failure. She asked not to be included as an author in the present study—not due to a lack of support for the work—but because I took on the synthesizing part of the analysis and writing myself. That said, I am grateful for her early contributions and ongoing support!

References

[1] C. M. Cunningham and G. J. Kelly, "Epistemic practices of engineering for education," *Science Education,* vol. 101, no. 3, pp. 486-505, 2017, doi: 10.1002/sce.21271.

[2] D. P. Crismond and R. S. Adams, "The Informed Design Teaching and Learning Matrix," *Journal of Engineering Education,* vol. 101, no. 4, pp. 738-797, 2012, doi: 10.1002/j.2168-9830.2012.tb01127.x.

[3] H. Petroski, *To forgive design: Understanding failure*. Cambridge, MA: The Belknap Press of Harvard University Press, 2012.
[4] C. J. Andrews, "Failure and idea evolution in an elementary engineering workshop," in *ASEE Annual Conference & Exposition*, New Orleans, LA, 2016, doi: 10.18260/p.26895.
[5] A. Jackson, A. Godwin, S. Bartholomew, and N. Mentzer, "Learning from failure: A systematized review," *International Journal of Technology & Design Education,* vol. 32, no. 3, pp. 1853-1873, 2022, doi: 10.1007/s10798-021-09661-x.
[6] P. S. Lottero-Perdue and E. A. Parry, "Elementary teachers' reflections on design failures and use of fail words after teaching engineering for two years," *Journal of Pre-College Engineering Education Research,* vol. 7, 1, Article 1, pp. 1-24, 2017, doi: https://doi.org/10.7771/2157-9288.1160.
[7] M. Johnson, G. Kelly, and C. Cunningham, "Failure and Improvement in Elementary Engineering.," *Journal of Research in STEM Education,* vol. 7, no. 2, pp. 69-92, 2021, doi: https://doi.org/10.51355/jstem.2021.101.
[8] P. S. Lottero-Perdue, "The engineering design process as a safe place to try again: Responses to failure by elementary teachers and students," presented at the NARST Annual International Conference, Chicago, IL, 2015.
[9] J. Piaget, *The development of thought: Equilibration of cognitive structures*. Viking Press, 1977.
[10] C. A. Chinn and B. A. Malhotra, "Children's responses to anomalous scientific data: How is conceptual change impeded?," *Journal of Educational Psychology,* vol. 94, no. 2, pp. 327-343, 2002, doi: 10.1037/0022-0663.94.2.327.
[11] D. P. Crismond, "The role of diagnostic troubleshooting in engineering design: Case studies," in *ASEE Annual Conference and Exposition*, Pittsburgh, PA, June 22-25 2008. [Online]. Available: https://peer.asee.org/4481. [Online]. Available: https://peer.asee.org/4481
[12] D. P. Crismond, "Scaffolding strategies that integrate engineering design and scientific inquiry in project-based learning environments," in *Fostering human development through engineering and technology education*, M. Barak and M. Hacker Eds. Rotterdam, Netherlands: Sense Publishers, 2011.
[13] P. S. Lottero-Perdue and E. A. Parry, "Perspectives on failure in the classroom by elementary teachers new to teaching engineering," *Journal of Pre-College Engineering Education Research,* vol. 7, 1, Article 4, pp. 1-21, 2017, doi: https://doi.org/10.7771/2157-9288.1158.
[14] P. S. Lottero-Perdue and E. A. Parry, "Elementary teachers' reported responses to student design failures," in *American Society for Enginering Education Annual Conference & Exposition*, Seattle, Washington, 2015, doi: 10.18260/p.23930.
[15] P. S. Lottero-Perdue and M. Tomayko, "Kindergartners' engagement in an epistemic practice of engineering: Persisting and learning from failure," in *2020 American Society for Engineering Education Virtual Annual Conference*, Online, 2020, doi: 10.18260/1-2--34892.
[16] K. N. Silvestri, McVee, M. B., Shanahan, L. E., & English, K., "A positioning theory analysis of interaction surrounding design failures in an elementary engineering club," *Journal of Pre-College Engineering Education Research,* vol. 13, no. 1, p. Article 7, 2023.

[17] R. K. Skinner and D. B. Harlow, "Recognition of design failure by fourth-grade students during an engineering design challenge," *Journal of Pre-College Engineering Education Research,* vol. 12, no. 2, Article 10, 2023, doi: https://doi.org/10.7771/2157-9288.1377.
[18] A. Simpson, A. Anderson, and A. V. Maltese, "Caught on camera: Youth and educators' noticing of and responding to failure within making contexts," *Journal of Science Education and Technology,* vol. 28, no. 5, pp. 480-492, 2019, doi: https://doi.org/10.1007/s10956-019-09780-0.
[19] P. S. Lottero-Perdue, "Elementary student reflections on failure within and outside of the engineering design process," in *American Society for Engineering Education Annual Conference & Exposition*, Columbus, Ohio, 2017, doi: 10.18260/1-2--28213.
[20] Engineering is Elementary Team, *To get to the other side: Designing bridges*. Museum of Science, Boston: National Center for Technological Literacy, 2011.
[21] Engineering is Elementary Team, *A stick in the mud: Evaluating a landscape*. Museum of Science, Boston: National Center for Technological Literacy, 2011.
[22] Engineering is Elementary Team, *An alarming idea: Designing alarm circuits*. Museum of Science, Boston: National Center for Technological Literacy, 2011.
[23] Engineering is Elementary Team, *A slick solution: Cleaning an oil spill*. Museum of Science, Boston: National Center for Technological Literacy, 2011.
[24] Engineering is Elementary Team, *Thinking inside the box: Designing plant packages*. Museum of Science, Boston: National Center for Technological Literacy, 2011.
[25] E. A. van Es and M. G. Sherin, "Learning to notice: Scaffolding new teachers' interpretations of classroom interactions," *Journal of Technology and Teacher Education,* vol. 10, no. 4, pp. 571-596, 01/01/ 2002.
[26] L. Z. Jaber, S. G. Davidson, and A. Metcalf, ""I loved seeing how their brains worked!"—Examining the role of epistemic empathy in responsive teaching," *Journal of Teacher Education,* vol. 75, no. 2, pp. 141-154, 2024, doi: 10.1177/00224871231187313.
[27] L. Z. Jaber and D. Hammer, "Learning to feel like a scientist," (in English), *Science Education,* vol. 100, no. 2, pp. 189-220, 2016, doi: 10.1002/sce.21202.
[28] A. V. Maltese, A. Simpson, and A. Anderson, "Failing to learn: The impact of failures during making activities," *Thinking Skills and Creativity,* vol. 30, pp. 116-124, 2018, doi: https://doi.org/10.1016/j.tsc.2018.01.003.
[29] C. M. Cunningham, C. P. Lachapelle, R. T. Brennan, G. J. Kelly, C. S. A. Tunis, and C. A. Gentry, "The impact of engineering curriculum design principles on elementary students' engineering and science learning," *Journal of Research in Science Teaching,* vol. 57, no. 3, pp. 423-453, 2019, doi: 10.1002/tea.21601.
[30] J. Saldaña and M. Omasta, *Qualitative research: Analyzing life*. Thousand Oaks, CA: Sage, 2018.
[31] M. Schreier, *Qualitative content analysis in practice*. Sage Publications Ltd., 2012.
[32] M. E. Jordan and R. R. McDaniel, "Managing uncertainty during collaborative problem solving in elementary school teams: The role of peer influence in robotics engineering activity," *The Journal of the Learning Sciences,* vol. 23, no. 4, pp. 490-536, 2014. [Online]. Available: http://www.jstor.org.proxy-tu.researchport.umd.edu/stable/43828355
[33] J. Schnittka and C. Schnittka, ""Can I drop It this time?" Gender and collaborative group dynamics in an engineering design-based afterschool program," *Journal of Pre-College Engineering Education Research,* vol. 6, no. 2, 2016, doi: 10.7771/2157-9288.1120.

[34] E. Silva Mangiante, E. Carpenter, and H. Robichaud, "Elementary student engineers: Their tamwork, negotiations, and problem-Solving," (in English), *Kappa Delta Pi Record,* Article vol. 56, no. 2, pp. 70-76, 2020, doi: 10.1080/00228958.2020.1729637.
[35] J. R. Wieselmann, E. A. Dare, E. A. Ring-Whalen, and G. H. Roehrig, ""I just do what the boys tell me": Exploring small group student interactions in an Integrated STEM unit," *Journal of Research in Science Teaching,* vol. 57, no. 1, pp. 112-144, 2020, doi: 10.1002/tea.21587.

Appendix A Post Unit Interview Protocol Questions

Post Unit 1 Interview Protocol Questions

1. What did you learn from the [Unit 1 Bridges] unit you just finished?
    a. What did you learn about engineering?
2. Tell me about your first design decision about how to design your bridge.
    a. What kind of bridge did you decide to build and why?
    b. How did you decide on this kind of bridge (or these kinds of bridges together)?
    c. What materials did you decide to use and why?
3. What parts of your first bridge design worked well? How do you know?
4. What parts of your first design did not work well? How do you know?
5. What changes did you make to your bridge design to try to improve it? (Ask for detail, like: Did you change the bridge type? Did you change the number or type of materials?)
    a. What parts of your second bridge design worked well? How do you know?
    b. What parts of your second bridge design did not work well? How do you know?
6. Would you say that your first (or second) bridge design failed? How so?
    a. Did parts of it fail?
    b. Did it fail to do certain things?
7. 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”?
8. Did your ideas about what it means to fail change after you did this engineering design challenge?
9. Let’s talk about working together as a team. Do you think your team worked well together?
    a. What was the hardest thing about being together on a team? What was the easiest?
    b. Did you always agree about your test results? Did you always agree on what plan to make?
    c. What is one thing that you learned about working on a team that you’re going to remember next time you’re on a team?
10. Do you have anything else you want to share with me today?

Post Unit 2 Interview Protocol Questions

1. How well did you first [Unit 2] design work?
    a. What parts of your first [Unit 2] design worked well? How do you know?
    b. What parts of your first [Unit 2] design did not work well? How do you know?
2. What changes did you make to your first [Unit 2] design to try to improve it? (ask for detail/elaboration)
3. Would you say that your first (or second) [Unit 2] design failed? How so?
    a. Did parts of it fail?
    b. Did it fail to do certain things?
4. Did your ideas about what it means to fail change after you did the [Unit 2] engineering design challenge?
5. Let's talk about working together as a team. Do you think your team worked well together?
    a. What was the hardest thing about being together on a team? What was the easiest?
    b. How did you work through your disagreements?
    c. What is one thing that you learned about working on a team that you're going to remember next time you're on a team?
6. Let's compare the [Unit 1] Bridges unit and the [Unit 2] unit:
    a. Which unit did you like better?
    b. Which was harder/easier? Why?
    c. Was it easier for your design to fail in the [Unit 1] Bridges unit or the [Unit 2] unit?
7. Do you have anything else you want to share with me today?