Failure and Idea Evolution in an Elementary Engineering Workshop (Fundamental)

Chelsea Joy Andrews, Tufts Center for Engineering Education and Outreach

Chelsea Andrews is a Ph.D. candidate at Tufts University in the STEM education program. She received a B.S. from Texas A&M University in ocean engineering and an S.M. from MIT in civil and environmental engineering. Her current research includes investigating how children engage in engineering design through in-depth case study analysis.
Failure and Idea Evolution in an Elementary Engineering Workshop
(Fundamental)

Abstract

In the engineering education community, failure at the design stage has been promoted by researchers and policy documents, generally in an effort to engage learners in authentic engineering practices. Less often in engineering education, failure has been discussed as productive in its own right. One way failure may be beneficial is by encouraging students to revise and build on their ideas about why their designs are performing as they are. This connects directly to decades-old constructivist ideas that learning occurs through constructing, testing, and refining theories about how the world works.

If students are indeed learning from failure, then we can expect to see evidence that their ideas progress over the course of a design task. This raises the questions: How do students’ ideas evolve over the course of a failure-prone engineering design task? And, what differences are seen between tasks with repetitive failure and tasks with ready success? To investigate these questions, I draw from literature examining failure in science and mathematics education to better understand the role that failure can play in engineering. I examine video data from a single-day engineering design workshop for 13 upper elementary students. I focus on two consecutive tasks: the wind tunnel task, which featured rapid iteration cycles and repetitive failure, and the water transport task, which featured near-immediate success.

I closely examined student groups’ discourse and the changes they made to their design constructions following testing for evidence of the ideas informing their design decisions. For the wind task, the factors students attended to were coded, with codes such as, “weight of object,” and tracked both across groups and over time. The analysis revealed that some factors, such as weight, were common across all groups and persisted through the design task, while others, such as air flow, were taken up by a few groups, often after a long series of failures. Importantly, the initial factors, such as weight, were not abandoned in order to accommodate later factors, but rather factors appeared concurrently later in the task. In contrast, in the water transport task, there was little evidence that groups closely considered which factors led to the success of their designs and as a result there is little evidence that they built upon their initial ideas.

I found that all groups used ideas about how the world works to respond to testing failures, and indeed, students seemed to be designing in a more sophisticated manner, attending to multiple factors of the task, after experiencing repetitive failure. These findings provide empirical support for failure as a productive aspect of engineering design tasks, providing opportunities for students to articulate and build on their ideas. Additionally, facilitators were found to serve an essential role by questioning, modeling, and challenging students during the task. These results have direct implications for engineering curriculum design and teaching practices; in short, it is critical that students are given reasonably difficult tasks that incorporate physical testing, are given time for multiple iteration cycles, and are supported as they design.
Introduction

Engineering educators seek to engage learners in authentic engineering practices, including testing and iterating on designs (NAE and NRC, 2009). Failure is an integral part of the iteration cycle in the engineering design process; designs often fail, that is, do not meet all criteria and constraints, and a central disciplinary practice in engineering is interpreting this failure as feedback on those designs (Lottero-Perdue & Parry, 2014; Cunningham & Carlsen, 2015).

Recently, the Next Generation Science Standards (NGSS Lead States, 2013) as well as the Framework for K-12 Science Education (NRC, 2011) have required a focus on practices, including failure analysis for elementary students. These efforts are grounded in theories of learning that suggest students should engage in authentic disciplinary practices. However, while the engineering education community generally agrees that engaging in testing and iteration, including failure, holds a great deal of potential, there is great uncertainty about what failure can and should look like at the elementary level and what supports are necessary to make failure productive in elementary classrooms. In addition, teachers are hesitant to allow their students to fail or to even use the word failure in their classrooms (Lottero-Perdue & Parry, 2014).

Clarifying what engineering should look like for young students is especially important considering engineering has become increasingly present at the K-12 level over the last few decades. With engineering design explicitly included in the Next Generation Science Standards (NGSS Lead States, 2013), this trend is likely to accelerate. Despite the increased presence and awareness of engineering, there is a notable lack of basic research on how students engage with engineering design, especially at the elementary level (NAE and NRC, 2009; Brophy, Klein, Portsmore, & Rogers, 2008). A result of this lack of research is continued uncertainty about how best to create and facilitate engineering design tasks and curricula for the K-12 level.

In this paper, I explore the ways learners engage in testing and iteration while working on both a failure-prone and a success-prone engineering design task using coding and conversation analysis. The data is drawn from a single-day engineering design workshop where 13 upper elementary students engaged in multiple hour-long design tasks, with clear objectives, constraints, and tests to determine success. In the wind tunnel task, students experienced repetitive failure, and only found success after many failed tests. In the water transport task, the students experienced ready success—many groups’ designs passed on the first attempt. In light of the vast differences seen in these tasks, my research questions are as follows: (1) In what ways did the ideas and factors used to inform designs evolve as groups engaged in failure-prone physical testing cycles? and (2) In what ways did groups’ responses to ready success differ from those seen after repetitive failure? I conclude with a discussion of implications for future research, classroom instruction, and the design of design tasks and curricula.

Background

This work is informed by constructivist and constructionist theories of learning, emphasizing building on learners’ prior understandings and experiences of the world (Piaget, 1952; Smith, diSessa, & Roschelle, 1994) through construction of public physical artifacts (Papert, 1980) as productive ways to engage learners in disciplinary practices and build conceptual knowledge. I
seek to do this by providing students with challenging and engaging engineering design tasks and supporting them during their designing.

Toward this goal, I am exploring what happens when students engage in engineering design tasks that feature repetitive failure, and how this differs from their engagement in tasks with ready success. I define failure as occurring when a design, design construction, prototype, or final product fails to meet some intended design criteria under intended constraints. This definition is aligned with descriptions in current policy documents and others in the engineering education community (e.g., Moore, Glancy, Tank, et al., 2014; Lottero-Perdue & Parry, 2014). By this definition, failure can be found at all stages of design—from drawings, to design constructions (if they fail a physical test), to finished products. It goes without saying that engineers work diligently to limit failures at the finished product stage; while these failures are inevitable and inform future design work, they are never intended (Petroski, 1992). On the other hand, failure during the design process, which can be described as failure-as-feedback (Lottero-Perdue & Parry, 2014), is necessary—this kind of failure informs engineers that their design needs improvement and hopefully gives insight into how to improve the design.

Although failure is often promoted by the engineering education community, there is great uncertainty about how elementary students respond to failure and how it should be approached at this level. This research provides empirical data that begins to unpack what we can expect from failure at the elementary level and what supports are necessary to make failure productive in elementary classrooms. To inform this analysis, I draw from literature on problem solving from outside of engineering that finds failure to be beneficial and perhaps even essential for learning, and then describe how failure has been treated in the engineering education community.

**Failure in other domains**

At least since Piaget introduced the concept of perturbations as necessary for learners to create more complex mental models, a body of literature has promoted allowing learners to struggle on their own with difficult problems without much structure. What I refer to as failure in this engineering context is similar to what other researchers have referred to as: perturbations (Piaget, 1977), impasses (Van Lehn, Siler, Murray, Yamauchi, & Baggett, 2003), failure (Kapur, 2008), script deviations (Schank, 1999), errors (Gartmeier et al., 2010), and micro-failure (Tawfik, Rong, & Choi, 2015).

The benefits of failure have been theorized for decades, with empirical support reported for a range of ages and content. Karmiloff-Smith and Inhelder (1974) found that as young children struggled with a balancing task they developed “theories-in-action” that they then tested and refined as they continued engaging with the task. Research in physics education has found that students were more likely to learn a physics principle when they reached an impasse than when instructors interceded before an impasse was reached (Van Lehn et al., 2003). Kapur and colleagues (Kapur, 2008; Kapur & Bielaczyc, 2012) found that grade 7 and 11 students who were allowed to struggle with complex, ill-structured problems in mathematics and science without support structures later outperformed students taught with a direct instruction approach.

Much of this research sees failure as an essential learning experience that affords students “an opportunity to refine their knowledge and understanding of the phenomenon” (Tawfik et al.,
The literature argues that failure should be welcomed as it: encourages students to generate ideas and consolidate solutions (Kapur, 2008), activates prior knowledge differentiation and allows students to critically assess their current knowledge (Bransford & Schwartz, 1999; Kapur, 2011), can set the stage for noticing critical features of a solution (Bransford & Schwartz, 1999), provides an impetus for prediction and causal reasoning (Schank, 1982; Gartmeier et al., 2010), helps students resolve misconceptions (Gartmeier et al., 2010), and makes it easier to retrieve experiences for future problem solving (Kapur, 2010; Schank, 1999; Gartmeier et al., 2008).

Tawfik and colleagues (2015) argue based on the prior research that failure “generates an additional inquiry process at the point of failure that may not exist during successful experience,” (p. 977) and they posit that this inquiry process leads to improved conceptual understanding of a phenomenon. Based on their review of the literature, they developed a model of failure-based learning that suggests learners progress through a cycle in failure-based problem solving that includes: experienced failure, challenge existing mental model, inquiry, identify potential reasons, find evidence to validate reasons, identify root cause, new solution generation, implement proposed solution, and assess viability of outcomes (p. 984).

Based on these conceptualizations and findings, researchers have recommended designing learning environments to intentionally induce failure (Schank, 1999; Gartmeier et al., 2010; Van Lehn et al., 2003; Kapur, 2011; Tawfik et al., 2015). In their canonical paper on preparation for future learning, Bransford and Schwartz (1999) contend that “[a]n important way that learners interact with their environments is by creating situations that allow them to ‘bump up against the world’ in order to test their thinking” (p. 82). Strategically employing failure within instructional design is one way to allow students to “bump up against the world” and test out their ideas. Most of these researchers suggest delaying support for learners until after failure has occurred and been recognized, and after learners have grappled with the failure (Kapur 2008; Van Lehn et al., 2003; Bransford & Schwartz, 1999). In contrast to typical lessons, when designing for failure, one would plan to engage in sustained inquiry after failure is encountered (Tawfik et al., 2015).

**Failure in engineering education**

When practicing engineers engage in designing small physical products (the kind of designing most similar to many tasks given to elementary students), they create and test models of their designs. Initial “models” may include mathematical models, then later digital models, and finally, sometimes, physical models (possibly prototypes, at full scale or model scale). Practicing engineers create and test these constructions, then use the previous test results as feedback to iterate and improve their design. In this way, interpreting failure (in the broad sense of not acceptably meeting all criteria) as feedback about designs is an essential aspect of engineering practice. In an effort to engage students with authentic engineering practices, many K-12 engineering tasks have been structured with the assumption that students will naturally follow these same steps: create design constructions, test them, and iterate on their designs using feedback from the test results. However, many engineering education researchers are now becoming concerned that this assumption is not warranted with novice designers, and that students need considerable support to iterate effectively.
Based on their meta-literature review, Crismond and Adams (2012) concluded that, overall, the test and redesign process is difficult and unnatural for young and inexperienced designers. They found that novice designers tended to evaluate a prototype’s performances “uncritically, in a coarse-grained, undifferentiated, and unfocused way,” leading to ineffective troubleshooting and possibly resulting in over-valuing flawed prototypes (p. 767). Similarly, Sadler, Coyle, and Schwartz (2000) found that elementary students have difficulty evaluating feedback from tests; in particular, they were likely to interpret experimental error as legitimate feedback on their designs.

Aligned with these conclusions that students have difficulty handling testing and failure independently, some researchers have suggested approaches to avoid or limit the failure students experience. Some advocate starting with a working prototype and improving it (Kolodner et al., 2003; Sadler, Coyle, & Schwartz, 2000; Schauble, Klopfer, & Raghavan, 1990), effectively avoiding failure, at least initially. In another approach, many published curricula feature a substantial amount of scaffolding, including task-specific fill-in worksheets and prescribed experiments (e.g., Engineering is Elementary, Cunningham, 2009; Learning by Design, Kolodner et al., 2003; Project Lead the Way, 2014). Indeed, the National Academy of Engineering report on K-12 Engineering Education, based on its investigation of literature and curricula, warns: “Although it may be tempting to allow students to direct their modeling themselves, the successful interventions reviewed here highlight the importance of the teacher providing explicit guidance and developing activities for investigating and negotiating contested claims” (NAE and NRC, 2009, p. 142).

The concern with allowing students to fail is also shared by many teachers, although they come to it from a different perspective. Lottero-Perdue and Parry (2014) found that a majority of teachers had an overall negative connotation of failure and that teachers were inclined to scaffold to avoid failure. However, many of these teachers associated failure with mistakes or saw failure as a personal trait, which is not consistent with the engineering view of failure as an “essential feedback mechanism” (p. 3).

Research questions and contributions of the current study

The literature reviewed above suggests that failure during physical tests in engineering tasks would give students feedback about their designs and the ideas that gave rise to those designs, leading to refinement of their understanding of the phenomenon. This supports my conjecture that engaging in failure-prone physical testing cycles provides opportunities for students to think deeply about the engineering design task and their design constructions, possibly prompting students to revise and build upon their initial ideas. In contrast, when success comes immediately, there is little need to closely consider what aspects of a design led to its success. It also suggests that students may require supports—such as facilitation practices—to productively engage in these practices.

However, still little is known about how learners might respond to repetitive failure during design tasks, or how to support them in doing so. Very little research has looked directly at how students engage in testing and iteration, especially at the elementary level. Allowing students to freely test designs and encounter failure is not without risks. Students may interpret failure during testing as a personal failure, rather than a failure of a physical object for a physical reason
(Lottero-Perdue & Parry, 2014; Sadler, Coyle, & Schwartz, 2000). Even if they do attend to more factors, they may not recognize how these factors interact, or that multiple factors may be valid at the same time. Or, students may identify factors that are unlikely to have any effect on their design or a given design task, or incorrectly attribute variations due to experimental error to changes they made to their designs (Sadler, Coyle, & Schwartz, 2000; Crismond & Adams, 2012). However, if failure can be productive, by prompting students to revise and build on their ideas about why their designs are working or not (Kolodner et al., 2003), then it is important to investigate how failure can be included in elementary engineering tasks.

My goal is to explore the potential of engaging learners in failure-prone engineering design tasks and to begin to identify what supports can encourage them to persist and make progress. Specifically, I ask:

1. In what ways did the factors used to inform designs evolve as groups engaged in failure-prone physical testing cycles?
2. In what ways did groups’ responses to ready success differ from those seen after repetitive failure?

Methods

The data for this study comes from a single-day engineering workshop for upper elementary students held at a university engineering education center in the greater Boston area.

Participants

A total of 13 students, entering fourth to sixth graders, attended the single-day workshop; 5 of these students were girls. The workshop facilitators, researchers affiliated with the center, included an undergraduate engineering student, graduate education students, and an elementary teacher on sabbatical with the center.

Design tasks

When engineering is presented in elementary classrooms, it is often in the form of design tasks (Cunningham, Knight, Carlsen, & Kelly, 2007; Douglas, Iverson, & Kalyandurg, 2004), in which students design, construct, and test an object to fulfill a specific function. Like all design problems, design tasks “set a goal, some constraints within which the goal must be achieved, and some criteria by which a successful solution might be recognized” (Cross, 2000, p. 13). A common feature of many elementary classroom engineering design tasks is the use of a physical test to evaluate groups’ design constructions (e.g., NAE and NRC, 2009). The workshop consisted of a series of design tasks to mirror what often occurs in classrooms.

During the workshop, students worked in groups of 2 or 3 on a series of four well-defined engineering design tasks, each lasting about an hour. For each task, groups created design constructions that met a set of design criteria, as evaluated by a physical test. Data for this paper is drawn from the second and third tasks, the wind tunnel task and the water transport task. For

---

1 All participants assented to the use of their real first names in published materials.
each of these tasks, students were given craft materials to build with, including: cardboard, straws, pipe cleaners, balloons, popsicle sticks, paper lollipop sticks, aluminum foil, coffee filters, rubber bands, egg cartons, and masking tape.

The goal of the wind tunnel task is to create an object out of craft materials that will hover in the top half of a plastic tube set above a vertically-oriented fan (Figure 1) for about ten seconds without flying out of the top or falling to the bottom. The task was introduced verbally and reinforced with a poster that stated: “Your Goal is: To build an object that can hover in a wind tunnel for at least 10 seconds. Your criteria: (1) it must stay between the top 2 bands, (2) Hover for 10 seconds before losing height or flying out. (Tip: Test! Test! Test many times)”.

![Figure 1](image1.png)

Figure 1: The test setup for the wind tunnel task, consisting of a clear plastic tube set atop an upwards-facing fan. The tube is 14 inches across and 40 inches tall.

The goal of the water transport task is to create an object that can transport an egg “passenger” across a tub of water (Figure 2), with a fan blowing down the length of the tub, without submerging the egg. The poster for the water transport task stated: “Your Goal is: Build an object that will float Mr. Egg across a container of water using wind power. Your Criteria: • Holds egg but cannot be taped down. Cannot fall out. • Must go the entire length. • Can sink a little—but not bottom.”

![Figure 2](image2.png)

Figure 2: The test setup for the water transport task, consisting of a long tub of water with a fan pointing down the length of it (the camera taking this picture was situated above the fan).
During each task, students were free to test their designs as often as they desired and were encouraged at the beginning to also test individual materials. There was a final share out at the end of each task where each group publically tested their designs; the share out was framed as a low-stakes venue to share their most successful designs and design history with the other groups.

**Data collection**

All work was recorded using multiple video cameras set to capture small group work at tables and testing at two testing stations. Pictures were taken of all final artifacts and of any drawings that were created. Both talk and participant artifacts were analyzed.

**Analysis**

I will report on three complementary analyses of the workshop data, conducted to address the research questions posed above. First, I paint a broad sketch of how ideas during the failure-prone wind task evolved over time. Second, I present a deeper analysis of the coding, providing transcript excerpts to both better describe and justify the codes used in the first analyses. Third, I briefly present an example episode from the water task as a contrasting case, as this task featured ready success.

Because I am interested in testing and failure, the first task was to identify tests and failures for each group. For the wind task, each time an object was released into a wind tube by a student was counted as a single test. Most of the time, students made changes to their designs between tests, often at their group working area. However, sometimes, there were multiple tests in a single visit to the testing station without any changes to the design between tests. For most of these repeated tests, the facilitators encouraged the students to test again to get a better understanding of the test and the result. The test result (success or the mode of failure) was also recorded. Finally, the videos were analyzed to determine, as much as possible, the changes made to designs between tests. Every test, test result, and change to the design was tabulated, but for simplicity, who desired the test (either a student or a facilitator) was not.

**Group-level coding of ideas.** To explore how learners engage in failure-prone physical testing cycles, I first sought to document how their ideas evolved over the course of the wind design task. Because I was interested in looking at the ideas mentioned in this specific activity, codes were derived from the data itself, rather than from published literature. I used a bottom-up verbal analysis (Chi, 1997) to identify and mark the presence of a number of ideas concerning the designs and the interactions between the designs and the physical test, such as weight and air flow. I chose to operationalize student ideas as the factors they attended to as they engaged in the task, as evidenced by their discourse, supplemented with knowledge of their building actions and test results. These factors, ideas about what would affect a design’s performance or test result, were generally expressed as reasons for design features or changes to designs, or explanations for why a design failed or passed a test. Each transcript for each group, including both building and testing cameras, was coded by turn. Only student turns that referred directly to an idea about the design or the test were coded, about 16% of the total number of student turns.

An initial list of a few codes was generated from knowledge of the data. For example, throughout the design task, groups referred to the weight of their designs or of materials, using phrases such
as “it was too heavy” or “it’s still too light.” After noticing this pattern, turns that referred to the weight of materials or the design as a factor that affects design performance were coded as “weight.” Upon examining all six group transcripts, the coding scheme was refined to include four codes that described the ideas and factors considered by multiple groups: (1) weight, (2) size, (3) air pushing/catching air, and (4) air flow. A fifth code, “other factors,” was also included because aside from the first four codes, the factors mentioned by groups were so diverse that it was not possible to create reasonable codes that crossed groups, but it was considered important to capture these ideas. Turns were double-coded for multiple codes if multiple factors were mentioned. This approach allowed me to construct a larger-scale representation of general patterns across groups.

The appendix features a table with explanations and examples of each code and extended examples of three of the codes (weight, air flow, and other factors) is given in the results. At least two other educational researchers independently coded between 22%-33% of the transcript for each group. Raw agreement across all codes varied between 94% and 98% for different groups. All differences were resolved through discussion and viewing the video data.

**In-depth analysis of transcript excerpts.** The patterns of different ideas that emerged from the coding in the first analysis suggests that over the course of the design task, after encountering repetitive failure, groups began to consider additional factors in their designing and in interpreting the test. To better understand what this pattern represents, I present closer analysis of excerpts from the wind task that serve as examples of a few of the codes.

Finally, I present a contrasting case from the second considered task, the water task, as an example of what test responses can look like when success is readily achieved.

**Results**

To address research question 1, I present results from the group-level coding analysis, which revealed that the students began attending to additional factors over the course of the wind task. In part 2, I present examples of the codes and deeper analyses of the factors groups attended to over time. To address research question 2, in part 3 I present a short analysis of the water task, a design task with ready success, to serve as a contrasting case.

**Part 1: Group-level coding of wind task ideas**

Figure 3 presents the results of the overall coding analysis for all six groups. Each graph shows both tests and factors, marked at the time elapsed during the design task. This section will describe the patterns in the coding results and the next section will provide more details and examples about the codes themselves.

In each of the graphs in Figure 3, the bottom-most series of markers shows when tests were conducted and the results of those tests; successes are marked by red squares and all other tests (black dots) were failures. The other five horizontal rows show when a member of the group mentioned one of the coded factors.
Figure 3: Results from the coding analysis for all six groups. The bottom line marks tests (red squares indicate success, black dots indicate failure); the other lines mark coded speech turns. Note that the cameras for different groups are not synced, so times do not line up with each other perfectly (up to 5-minute offsets are likely). The bottom group, Julian and Alex, only have six coded turns; this group was difficult to hear on the camera because they spoke very softly, answered many facilitator questions with “I don’t know,” and spent much of the task time off task building “launchers.” Even so, they had the most successful designs of any group (three unique designs, one was tested twice).

Note that ideas, factors, and designs are related but distinct. While ideas are often reflected in designs, designs include a multitude of ideas of varying scale, many of which are not explicitly expressed. Ideas may or may not be expressed as factors about a design or the test related to the performance of a design. Students mentioned ideas that were not coded as factors and many ideas that were not incorporated into their designs. Here, I am looking at the factors students referred to, rather than the designs themselves or changes to the designs or the ideas that could be derived from those designs.

From the testing markers, it is clear that students tested constantly during the design task and were rarely successful. Groups tested between 29 and 87 times, with an average of about 52 tests per groups. Of these tests, only 14 tests, or about 4%, of the tests were successes, and represent 9 unique designs (four successful designs were tested multiple times with no changes and were successful every time). Some tests, mostly at the beginning of the task, were tests of single materials—rather than tests of constructed designs—and were assumed to be conducted to see how a material behaved in the fan, as opposed to a test to see if a design is successful. The facilitators explicitly encouraged testing materials as a way to make the test station less intimidating. Students shifted to testing constructed designs very quickly—generally after only one or two tests of materials.

In terms of factors, this analysis revealed that first and consistently throughout the task, students in all groups (with the exception of Julian and Alex) mentioned weight as an influential factor. Then, generally later on in the task, most groups also referred to other factors, but most groups continued to mention weight. References to air flow and “other factors” typically occurred after many tests, the vast majority of which were failures.

This pattern seems to support the conjecture that failure and testing provides opportunities to think more about a problem, as seen by how students began attending to more factors of the design task and their designs after engaging in numerous tests and experiencing repetitive failure.
Part 2: Wind task analysis

Given the pattern revealed above, here I present and more deeply analyze representative excerpts from the wind design task as both examples and justifications of the codes shown in part 1. In particular, I seek to describe: initial references to weight, an example of the air-flow factor informing design changes, and an example of attending to other factors when evaluating tests.

**Initial factors:** As evident in the coding in Figure 3, weight was the most common factor mentioned. All groups except Julian and Alex referred to weight, from the beginning and generally throughout task, and they continued mentioning weight even after they began mentioning other factors. The following excerpts exemplify some of the ways this weight factor was articulated. Each of these turns were coded as “weight” factors to create Figure 3.

Weight was mentioned as informing and justifying changes to a design. Designs often seemed to be classified as too heavy or too light and materials were added or removed in response:

Aeden: We had to cut this in half so we’re putting tin foil on top of that for weight.

Vincenzo: Well, I added more tin foil ‘cause it was too heavy.

Abby: …and then the balloon and tin foil sort of makes it really really light so I added a popsicle stick, to make it like a little heavier.

Weight was also often used to explain test results, as in this exchange with a facilitator (facilitators are designated by first initial and an asterisk):

R*: Okay, why do you think it sunk?
Sarah: Umm, probably the weight of the tin foil.

Finally, weight was also invoked in giving advice to fellow students:

Abby: Add a popsicle stick to yours if it was too light.

Some students articulated why weight is important, including the role of the fan, as in this exchange between the workshop leader and two girls in the same group:

L*: So why do you think weight is even important in this activity?
Cecelia: Because if it's too light it'll just like, make it go [motioning upwards], it'll like the wind will just push it very easily.
L*: It'll fly out.
Ruth: Too light too high, too heavy too low.

Here, Cecelia explains that if the design is too light it will fly out the top, because “the wind will just push it very easily.” This line was coded for both weight (“if it’s too light”) and air pushing (“the wind will just push it”). Cecelia seems to recognize that, at a basic level, to succeed in the task they need to balance the upward pushing from the wind (a drag force, although the students do not use that language) and a downward force from the weight of the object.
It is unsurprising that weight is the first and most common factor mentioned by the students. Weight is an intuitive factor and is easy for the students to judge with proprioceptive feedback. Weight is certainly a critical aspect of this design task, and in certain designs, especially those built around balloons, weight may be the only factor students need to consider. In fact, a few groups were successful with their first designs (after many iterations) after explicitly referring only to weight. Finally, the facilitators were supportive of weight explanations, which would likely reinforce weight as an important factor to consider.

After some design time, many groups expressed surprise and often frustration when adjusting the weight of their designs did not result in the expected outcome. For example, in this excerpt Abby approaches the testing station and describes the change she made to her design to the facilitator:

Abby: I added another popsicle stick
K*: You added another popsicle stick.
Abby: Two popsicle sticks.
K*: Two?
Abby: [Tests design, flies out the top] And it’s still too light!

Here, Abby is surprised (and from her tone of voice, frustrated) that after adding two additional popsicle sticks her design still flies out the top.

While weight is important, and while it may be possible for some designs to succeed while only attending to weight, weight is certainly not the only factor at play in this design task. For example, stability is also crucially important—many students create flat designs that hover until they become unstable and rotate and fall to the bottom. The frustration students experience after many failed tests while attending solely to weight may prompt students to consider other factors.

Later patterns. As shown in the coding analysis (Figure 3), most groups also mentioned factors other than weight during the design task. For many groups, these factors were mentioned after many failures and after they had attended to weight for some time.

Here I give two examples of these later factors. In this first example, Regan describes making “slits” in his design, which provides an example of the “air flow” coding in Figure 3. In the second example, Cecelia attends to multiple factors during the same testing period, some of which are examples of the “other factors” coding in Figure 3.

Regan: air flow code. This episode, which takes place a little over 40 minutes into the design task, provides an example of what was coded as “air flow” ideas.

Like all of the groups in this task, Regan and his partner Nicky decided to split up and work on different designs. Regan first worked on a balloon-based design (like many others in the workshop) and was successful after 18 tests, about 30 minutes into the task. After this success, the workshop leader challenged Regan to create a design solution that did not include a balloon. Regan’s partner, Nicky, had been working on a series of non-balloon designs the entire task and had not been successful. In Regan’s second design, which he worked on until the end of the task, he taped a coffee filter to a small section of a cardboard egg carton and then taped “weight”, mostly popsicle sticks and lollipop sticks, on top of the coffee filter. After five failed tests of this
design, in response to which he adds or takes away weight, he decided to “add more parachute”—he taped another coffee filter on top of the popsicle sticks and lollipop sticks. After four more tests where the design fails by flying up and Regan responds by adding weight, Regan thinks of another possible solution: cutting “slits.” This excerpt begins with the test and failure just before Regan first cuts slits.

1 Regan: Don't fly away [tests, flies out top] More weight! I added enough weight for you thing-a-ma-bob.
2 Regan returns to table, seems about to add more foil
3 Regan: That's an idea. Where are the scissors?
4 Nicky: There.
5 Regan cuts slits in coffee filters and tapes on another lollipop stick [Figure 4]. Goes to testing station
6 Regan: [To K*] I slit, I made slits in the thing. To let some of the air out.
7 Regan releases design in tube, it hovers at the top
8 Regan: One, two. One, two, three, oh! [starts counting to see how long it hovers] [it flies out the top]
9 K*: I think that was a good idea though, you slit, you um, cut the coffee filters so the wind gets out?
10 Regan gets design and goes back to table, still very close to K*
11 Regan: I put little slits
12 K*: [Speaking from testing station] Maybe you could talk to Liam about your design.
13 Regan cuts more slits at table, then goes back to testing station
14 Regan: Made tiny little slits. More slits. [tests, flies out before he can start counting; Figure 4] No! More slits.

Regan goes through two more cycles of cutting slits at his table and then testing his design (it continues to fail) until it hovers for two counts and falls, at which point he decides to remove weight. Through this time, he only mentions slits in phrases like: “Let’s see if the slits work” and “More slits;” he does not again describe how he believes the slits will function.

Figure 4: Pictures of Regan cutting slits in his design, left (line 5), testing his design, center (line 14), and an artist’s rendition of the design, right.
In this episode, we first see Regan expressing frustration that adding weight was not effective in preventing the design construction from flying out the top of the tube (line 1). Regan then decides to cut slits in the coffee filters of his design. He seems to be talking to either himself, his partner, or the camera when he says, “That’s an idea” (line 3). When he has finished altering his design and goes to test it, he tells Kerrianne, the facilitator at the testing station, “I slit, I made slits in the thing. To let some of the air out” (line 6). After he tests, Kerrianne clarifies, “You slit, you um, cut the coffee filters so the wind gets out?” and in apparent agreement, from his table Regan says, “I put little slits.” Regan does not explain his thinking behind the slits again. From his brief description and the fact that he is apparently cutting slits as an alternative to adding more weight, I believe Regan thinks that with slits in the coffee filter, some of the air will flow through the slits and that this air will not be pushing up on the design, so the design will have comparatively less air pushing up on it.

In his first, ultimately successful balloon design, Regan simply added (with tape) and removed “weight” (small objects) from his balloon until it hovered in the tube for ten seconds. With that balloon design, Regan did not need to consider the air from the fan as anything other than an upwards force on his design that needed to be balanced with a downwards force from the weight of his design. This description is not intended to imply Regan explicitly thinks of the air in this way (he never uses words like “force” or even mentions air directly before this transcript), but rather to point out that there is no impetus for thinking more complexly about the air before this point. To understand why slits could be effective, one must be able to conceive of the air as something that can be divided—some of the air goes through the slits and some does not. This is a different way of conceptualizing the problem and the test, since it is more than just balancing weight. The fact that Regan cuts slits implies that he is now thinking about more dimensions of the problem. Of course, weight is still a very important dimension of this design task and Regan certainly does not forget about weight once he starts thinking about air flowing through slits. It appears that Regan has added on the factor of air flow and is now attending to this additional concept of air flow. It is also noteworthy that Regan continues to add slits three more times even though they do not seem immediately effective. After he first cuts slits, the design hovers for a count of three before flying up, which is longer than it hovered previously. But, after that the design only hovers for a maximum of a count of two. It seems reasonable that Regan is confident that cutting slits is an
effective design decision or he would not have continued trying it. Regan is also receiving positive feedback about the slits from Kerrianne, although the vast majority of Kerrianne’s feedback was positive over the entire design task. His persistence with the slits does eventually pay off—after cutting slits a total of four times the design construction falls to the bottom, thus solving his flying-out-the-top problem.

After the design falls, Regan goes back to adding and removing weight for the seven remaining tests until the design task is over (Regan was never successful with this second design). Like most groups, even after he began using different ideas, like air flow, in his design, weight remained an important parameter, as it arguably should be. It is easier to perceive why weight works, it is more intuitive to judge how much adjusting to do with weight, it is easy to add and remove weight (whereas it is hard to undo cutting slits), and adjusting weight worked for Regan’s first, successful design.

**Cecelia: Other factors code.** This episode, which takes place about 45 minutes into the design task, provides an example of what was coded as “other factors.” As explained above, the “other factors” code was created because aside from the weight and air flow ideas, the factors mentioned by groups were so diverse that it was not possible to create reasonable codes that crossed groups.

Cecelia is part of the single three-person group. Like all of the groups, they decided at some point to split up and work on separate designs, but unlike most groups they continued actively helping each other build their designs and asked for and gave each other advice. All three of the girls worked on balloons designs (without success) until about 30 minutes into the task, when Lija, the workshop leader, challenged everyone to create designs without balloons. (This challenge was issued after a few groups had been successful with balloons.) This prompts Cecelia to put her balloon design aside and begin working on a new design using a coffee filter. She begins with a coffee filter, pokes a pipe cleaner through each side of the filter like a handle filter, and adds popsicle sticks and foil after testing and having out-the-top failures (see Figure 5). She has tested this design at least six times before this transcript (the testing camera ran out of battery and was off line while it was replaced, so some tests may be missing); all of these tests resulted in failures. At this point, Cecelia has been working on this design for about 15 minutes. She approaches Riley, a facilitator, at the testing station:

![Figure 5: Picture and an artist’s rendition of Cecelia’s design](image)

1 R*: So what did you do?
2 Cecelia: I wrapped this because I didn't want [unclear] popsicle sticks
Cecelia tests—it very slowly sinks—still floating low (bowl orientation)

R*: One, two [begins counting to 10 to see if design passes] [design drops below middle band and hovers just above the fan]. You really got it to float though. [2.5 second pause] This is cool. Why, do you think?

Cecelia: I don't know.

R*: Here, can I see it one more time?

Cecelia: Because the paper is being pushed. [Gestures with palms coming together horizontally]

R*: The paper is being pushed down? [Lifts up tube and Cecelia picks up design]

Cecelia: Yeah. So I'm gonna try it like this way [Flips it upside down—parachute orientation].

Cecelia tests upside down from before—parachute orientation—it flips over

Cecelia: It just goes right back upside down [to bowl orientation]

R*: Why do you think it goes right back upside down?

Cecelia: Because, there's space there to—[interrupted]

R*: Count it, one, [Cecelia joins in] two, three, four, five, six, seven, eight, nine, ten. I think that's ten.

Cecelia: Yay! [High fives Lija]

R*: Whoa. So why do you think that's work—it's it's hovering?

Cecelia: Umm, because maybe the wind is pushing up [gesturing upwards] on the paper [likely means the coffee filter].

R*: The what is?

Cecelia: The wind.

R*: [Lifts up tube and Cecelia gets design] It's pushing up on the paper, but so why is it working now when it didn't work other times?

Cecelia: I don't know. [Holding design in hands and rotating it, looking at it] Maybe because I started it [holding it up—parachute orientation] right side up and then it flipped itself over [gestures flipping—to bowl orientation].

R*: Do you want to test it one more time? I know we have a line going but I want to see if we can figure it out.

Cecelia tests in parachute orientation [Figure 6, left]—flies out before it goes in.

Cecelia tests again in bowl orientation [Figure 6, right]—it hovers successfully. Both watch for a few seconds.

R*: It really likes to be this way huh?

Cecelia: That is that [unclear] how it works, yeah.

R*: What do you think is special about this way that—[picks up tube and collects design] what's special about it that way that lets it float?

Cecelia: I guess the weight on the bottom. [Holds it up to show popsicle sticks on bottom]

R*: The weight being on the bottom. Okay.
In this episode, Cecelia tests her design four times; two of these tests (at line 14 and line 24) are successful. Importantly, she tests it in two orientations: what I call “parachute” orientation (Figure 6, left) and “bowl” orientation (Figure 6, right). The first time, in bowl orientation, it hovers for a count of two, then she tests it upside down, in parachute orientation, and it flips itself over to bowl orientation and is successful. In an attempt to get a better idea of why it is working, she tests it again, first in parachute orientation, and it flies out (line 23) and then finally once more in bowl orientation where it hovers successfully again. When Cecelia first tests the design (well before this transcript), she describes her design as “my little parachute thing”, so we can assume she pictured it working in parachute orientation and she may be surprised that it is actually successful in bowl orientation.

While Cecelia is conducting these tests, she is encouraged by Riley, the facilitator, to think about “why” the design is working (line 4, 12, 16, 20) and she comes up with multiple possible factors that may be influential. We see Cecelia mention at least four possible factors: (1) that the “paper” (likely she means the coffee filter) is being pushed (lines 7 and 17), (2) that the orientation matters, as she tries it upside down (line 9), (3) that “it flipped itself over” (line 21), and (4) the “weight being on the bottom” (line 28). The first factor, that the wind pushes on the coffee filter, is similar to how she talked about the earlier designs and tests; a few times when she mentioned weight it was expressed as countering the wind pushing up on the weight (this line was coded as “air pushing”). Riley likely recognizes this as an idea she’s already expressed and an idea that is not specific to this situation, and he asks: “But so why is it working now when it didn't work other times? (line 20), which prompts her to give the other three factors above. These factors are not given by her as a list, she mentions each one after examining the design either while testing or in her hands, and after being questioned by Riley. It seems from the way Cecelia pauses and looks carefully at the design and/or the test before responding that she is thinking about what is going on and trying to think of explanations in the moment. With the exception of the paper being pushed by the wind idea, it seems that she did not come into this testing period considering these aspects as factors, or at least expecting these factors to be influential, but that they only became potentially important factors as a result of what she witnessed in the test and how she interpreted it.
In most of the preceding tests Cecelia attends solely to weight and air pushing (which seem to be closely linked for her) as factors. Here, she attends to more factors, likely in addition to weight, which is a more sophisticated approach to design. Cecelia shows that she is looking closely at the design and comparing different tests to determine what might be causing success in some cases and not in others. This seems different from earlier tests, when the failed designs were considered either “too heavy” or “too light” and responses consisted of adding or subtracting material that functioned solely as “weight.”

Cecelia definitely does not seem certain that the factors she presents are responsible for the success or failure of her design. In addition, the ideas she puts forward may not be accurate (although it is difficult to judge accuracy with such a difficult test and with her ideas not well described). I consider certainty and accuracy to be not as important as thinking about different factors and how they might interact. I think at this stage, where she is closely examining the test, most designers would be brainstorming reasons for the results, which they would more completely articulate and attempt to defend later. Cecelia is also certainly not now an expert at this task nor would she likely be able to easily create another successful design. However, we do see the beginnings of Cecelia thinking about more factors related to the task, test, and her design, and how these factors might interact to lead to a design passing or failing the test.

This excerpt is an example of the more sophisticated reasoning that occurred later in the design challenge and after the students experienced many tests and failures (both their own and of others). I hypothesize that the repetitive testing and failure (which go hand-in-hand, as students only test so many times because they are experiencing failure) provides opportunities for students to think more deeply about a situation and their designs, which may be expressed as attending to different and multiple factors while designing. Here, Cecelia is encountering both failure and success, and trying to differentiate between the two situations. The success here is likely conducive to considering more factors because the hovering design gives her more time to examine the design during testing—possibly the biggest issue with the wind task is that most failed tests are so fast is it difficult to see what caused the failure. However, I still argue that the many failures Cecelia has experienced until now may have prepared her to think more deeply about the task and the test at this point. Because she has experienced so many failures, she does not interpret this success as the design just working, but rather she understands that it must be working for a reason.

Finally, the role of the facilitator, Riley, is clearly critical in this episode; it is unlikely Cecelia would have articulated these ideas if Riley had not prompted her. The influence of Riley should not be perceived as somehow discrediting Cecelia’s ideas or lessening how sophisticated we perceive her designing to be, because Riley is not telling her what he thinks is important, but rather is encouraging her to figure out what she thinks is important. The way he models how to look at the test prompts her to think more carefully about the test and her design in a more sophisticated way. However, the facilitators were at the testing stations the entire task asking similar questions and only at this point did Cecelia mention so many factors, so her responses cannot be solely attributed to Riley’s prompting. While the facilitators are clearly essential in helping students progress through design tasks, their utility can be limited by the task. As evidenced in the next data section, when success is readily achieved without much struggle, it is hard for facilitators to push students to think carefully about the test and designs in the way we see Riley doing here.
Part 3: Water task counter-example

To clarify further how failure provides opportunities to think more deeply about a problem, here I present data from the water task, which did not have much failure, as a contrasting example. This task immediately followed the wind task in the same workshop with the participants working in the same groups.

As described above, for this task the students had to create an object to float an egg passenger down a tub of water, powered by a fan pointing down the length of the tub. After the task was presented to students and they began to work at their group tables, one of the facilitators saw egg carton tops leftover from another task at the materials table and handed them out to the groups. This unintended extra material made the task much easier for the students, as the egg carton tops made ideal hulls. As a result, groups succeeded much faster in this task and with far fewer failures than expected.

For this task, every time a design was placed in the water or a design was brought back to the fan end of the tub for another attempt was counted as a test. Counting tests for the water task was more difficult than for the wind task because the water task criteria can be broken down into three parts: floating, travelling, and carrying an egg. Sometimes, groups only checked to see if their design floated but did not turn on the fan to see if it traveled to the end of the tub, and sometimes groups tested with the fan but without the egg. Tests of this type that did not result in sinking are considered partial-tests and are counted separately, as they did not fulfill all of the criteria of the task and thus should not be grouped with the full successes. Any test that resulted in sinking are considered failures, because if a design cannot float it cannot fulfill the other criteria of the task.

Groups tested between 1 and 11 times, including partial-tests (tests without the fan and/or without an egg), for a total of 44 tests among all groups. There were 29 tests that fully succeeded, 9 tests that were partial-successes (either did not use the fan or did not have an egg), and 6 failures (where the design sunk or did not move). Thus, while the success rate in the wind task was less than 4%, the success rate in this task was 83% (86% including partial successes). Only two groups failed on their first attempt, and they both succeeded in their second test.

Some students also explicitly mentioned that they thought the task was easy. After their respective first tests, which were successful, Marco stated, “That was really easy” and Alex questioned the facilitators, “What kind of challenge is that?” It seems from these comments that after struggling with the wind task for so long, the students were surprised to succeed so quickly in the water task.

With these numbers and comments in mind, I now present a testing episode from the water task to demonstrate how different the conversations around testing were when success was readily achieved.

Sarah and Abby. Like many groups, Sarah and Abby were successful with their first design on the first test. Their design consisted of a hull created out of an egg carton top that they covered in foil, a sail attached to a straw mast, and a spot for the egg that would prevent it from rolling.
Abby and Sarah also taped straws on the bottom of the hull with the ends taped off (Figure 7, left); they focus on this aspect of their design when describing it to the facilitators.

Figure 7: Pictures of Sarah and Abby’s design. Close up of the straws on the bottom during construction, left, and the entire design during the final test, right.

The following episode is of Sarah and Abby’s first test (which is also their only test before the final share out), about 30 minutes into the design task. When they arrive at the testing station, they show their design to Lija (L*), the workshop leader:

1 Abby: We put straws [pause]. Show her the bottom!
2 L*: Show me the bottom? What’s cool about the bottom? What did you?
3 Abby: We put straws and then we taped the ends.
4 L*: Can you take it out for a second? [asking to take the design out of the water]
5 Abby: We taped the ends.
6 L*: We want the camera to see [lifts up design to show camera].
7 Abby: We taped the ends so it would float in water.
8 L*: And then water wouldn’t get inside.
9 Abby: Yeah, because if the water like went inside it would sink.
10 L*: But why is this important to your design?
11 Abby: Because then it will keep it floating.
12 L*: You think it will help with the floating? Almost like a catamaran sort of has like [gestures with her hands out the two hulls of a catamaran]. OK.
13 Sarah: So it’s balanced.
14 L*: OK, ready here you go. So you might want to come on this side Abby. Are you ready?
15 Sarah: Yup.
16 L*: OK. [Turns on fan, design travels to the end and succeeds] Nice work!
17 Abby: That went fast!
18 L*: And did you try this without the straws?
19 Sarah: No.
20 J*: This is their first test.
21 L*: This is your first test? I love that.

At this point, Abby and Sarah talk to the facilitators about how they anchored their sail and decide to challenge themselves now that they have been successful. However, while they discuss other ideas they run out of time before they build anything else, and when it’s time for the final
public test they use their first design, which is the only one they tested. At the final public test, when Lija asks them to “Tell us about your design,” Abby responds:

Abby: Well, Sarah wanted to do like a pirate boat so we have a wheel and then at the bottom [shows bottom] we have straws and the ends are covered so the water doesn’t si—come into them and so it won’t sink. And then we have our majestic sail.

In this episode, Abby excitedly volunteers to show off the straws on the bottom of the egg carton hull to Lija, the workshop leader. Lija holds up the design to get it on camera (line 6) and they discuss the straw idea. Abby explains that they taped the ends (she repeats it three times until Lija explicitly acknowledges it in line 8) so “it would float in water” (line 7). When Lija asks why the straws are important to their design, Abby answers “it will keep it floating” and Sarah adds “so it’s balanced.” When they test the design (line 16) by turning on the fan it is successful (it travels to the opposite end of the bin). Lija asks if they tried it without the straws and they say they did not, but Lija does not directly ask them if they believe it would work without the straws and they do not offer that information, although Abby does mention earlier that “if the water like went inside it would sink” (line 9).

Like many groups in this task, Abby and Sarah were successful with their first design on their first test. Because they were successful, the test did not provide them with much feedback, other than that the entire design construction worked. It is difficult to determine which aspects of a design are influential when it works. This seems to be the case with Abby and Sarah. Both when they are testing and when they describe their final design, Abby and Sarah focus on the straws they taped to the bottom of their hull. The straws distinguish their design from other groups’ (many final designs looked quite similar), so it is reasonable that they are eager to bring it up. Every time they discuss the straws, from the beginning of the task, to the first test with Lija, to the final description, they explain that the straws will help the boat “float in water” and that they taped the ends to keep out water “because if the water like went inside it would sink” (e.g., line 9). These are reasonable and interesting ideas, and having reservoirs of air is an effective floating mechanism in many cases. However, the air-filled straws were not likely a necessary part of this design. At the final public test, three other groups had nearly identical hull designs without straws and they were all successful as well (some with additional eggs). Because Abby and Sarah were successful from the start with this design, the effectiveness of the straws was not challenged, and they never had any reason to doubt that this was an essential aspect of their design. Their immediate success meant that they did not have to dissect their design and tease apart which features led to their design’s success; they knew that the whole thing worked, but they avoided having to discover why it worked.

In contrast to the wind task, where groups mentioned different and multiple factors later on in the task, there is no evidence that Sarah and Abby considered different factors related to floating later on. Only the few groups who had failed tests consider other ideas related to floating, like hull size, stability, and weight. These groups who experienced failure, who notably tried designs without an egg carton hull, had the opportunity to consider what went wrong and what needed to be altered to create a successful design. However, because the egg carton provided a ready solution, even the groups that experienced failure with one design quickly created (or had already created) a design using the egg carton and did not further grapple with floating ideas.
In addition to not having to struggle with what factors influenced a design’s ability to float, with the ready success in the water task the students also avoided engaging in many design practices. They did not need to closely examine a test to find the source of the problem and use the test result as feedback to improve their design. The students did not have the opportunity to tease apart the different components of their designs to see why something a certain effect.

The immediate success also made it difficult for the facilitators to push students to think more deeply about their designs and the task. For example, Lija probably suspected straws weren’t necessary to prevent sinking, but because Sarah and Abby were successful on their first attempt, it was difficult for Lija to direct their attention to specific parts of their design or to model for them how to compare results of different tests. In contrast to Cecelia’s case with the wind task, it is not possible for Lija to ask questions like “But why is it working now?” that were effective for Cecelia, who had experienced repetitive failure before her success.

**Discussion**

This study explored the conjecture that repetitive failure during physical tests of designs in engineering design tasks provides the opportunity for students to think more deeply about their designs and the test and the factors that influence success or failure. My research questions were: (1) In what ways did the ideas and factors used to inform designs evolve as groups engaged in failure-prone physical testing cycles? and (2) In what ways did groups’ responses to ready success differ from those seen after repetitive failure?

To answer the first research question, I analyzed the wind task, attending to the ideas students mentioned concerning the factors they believed influenced how their design functioned in the physical test. The results from the coding of transcript from the wind task (Figure 3), which featured repetitive failure, showed that over time most groups attended to new factors and multiple factors simultaneously. Importantly, students did not drop their initial ideas, such as weight, when they began considering other ideas, such as how air might flow through holes in a design. The later factors are not in any way more valid than initial factors, and the later factors may not be more sophisticated, but it is likely that these factors are less obvious, as the students did not mention them initially. Also, while the later factors themselves are not necessarily more complex, I do consider attending to multiple factors at once to be a more sophisticated way of thinking about the problem.

As shown in the coding results in Figure 3, different groups showed very different patterns in terms of the factors they attended to over the course of the wind task. For example, while Liam and Aeden cycle between different factors over the course of the task and stop mentioning weight half way through the task, Marco and Vincenzo attend to weight over the entire task and have a single 8-minute segment where they mention three other factor categories. This variation is to be expected: even the most ideal design tasks will only provide opportunities to develop one’s understanding of a phenomenon, and like all instruction do not in any way guarantee that all learners will engage in them in similar ways or arrive at the same place by the end. From a constructivist learning perspective, all learners build on their existing knowledge and experiences, so while design tasks can be carefully designed to ensure that most students will succeed within a reasonable amount of time, we cannot create tasks that will impart the same knowledge to all students. A reasonable goal for an effective design task is for all students to
progress to a more sophisticated understanding than what they arrived with. The coding method employed here, tracing ideas of groups over time, recognizes that students begin in different places and reveals whether groups are over time attending to more factors and multiple factors concurrently.

I believe that in this case, as predicted by the literature, failure provided opportunities to think more deeply about the situation, and, as shown by the data, after many failures students mentioned other factors. However, it may not be failure in and of itself that prompted students to describe more ideas about their designs and the test, but rather it could have been the many tests they conducted or simply the many interactions they had with the fan. Because failures, testing, and time at the testing station are inextricably intertwined, it is impossible to tease apart which of these may have more of an impact on student ideas. However, in this case, and with many design tasks, the only reason students tested so many times and spent so much time at the fan is precisely because their designs kept failing the tests. As seen in the water task, even when students are interested and engaged in a task, if they are not experiencing failure, they will likely conduct far fewer tests.

To answer the second research question, I analyzed an episode from the water task, which provides a contrasting case. In the water task, designs passed the test readily; all groups were successful within two tests. In this task, the students did not seem to deeply discuss why their designs worked, but instead focused on describing the features of their designs. Like the straws in Sarah and Abby’s boat, there are many features employed in the water task that may be unnecessary, but if the entire design is successful, individual features of designs are not challenged. 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.

It is interesting to directly compare the episode with Cecelia and the wind task with Sarah and Abby in the water task as both episodes captured the first success for each of them. Cecelia had experienced many failures at the wind tube before this point, including at least six failures with the current design, while Sarah and Abby were testing for the first time. When Cecelia’s design finally works, she is able to consider the question posed by Riley, “Why is it working now when it didn’t work before?” Because she has experienced many tests, Cecelia is in a position to reflect back on the different versions of her designs and the test results of those versions and compare them to see what is different now. This reflection may help her to differentiate factors and break down her design into its constituent parts. The many tests and failures may prime Cecelia to know what to look for (or what not to look for, in terms of what did not change between tests). Bransford and Schwartz (1999) describe how contrasting cases are important for preparation for future learning, and can lead to well-differentiated knowledge structures. The tests and failures Cecelia experienced could function as a set of contrasting cases, helping her determine what factors are influential in her design. Because Cecelia has faced failure, her current design does not simply “work,” rather, it works for a reason. On the other hand, Abby and Sarah only mention the straws on their design—the piece they find the most interesting and the aspect which sets their design apart, which they also seem to believe is important to its success. There is no evidence that their understanding of the task, their design, or the underlying mechanics of floating is refined in any way over the course of the task. They explain their design in the same way every time and do not mention any additional factors after they test.
The notion that it would be harder to gain insight into a phenomenon when success is readily achieved compared to when a learner has to struggle to arrive at a successful design is consistent with the literature looking at failure in other situations. Van Lehn (1988) went as far as to argue that “if there is no impasse, there is no learning” (p. 32). That is a very strong claim and this data can in no way confirm or refute it, but it does seem that additional work needs to be done around immediately successful tasks to ensure students progress in their understandings of the phenomenon. Importantly, what Van Lehn (1988) refers to as an impasse, or what Piaget (1977) refers to as a perturbation are quite possible for learners to experience even if their designs do not fail a physical test. According to Van Lehn, an impasse “occurs when a student realizes that he or she lacks a complete understanding of a specific piece of knowledge” as evidenced by when a “student gets stuck, detects an error, or does an action correctly but expresses uncertainty about it” (Van Lehn et al., 2003, p. 220). While it is possible for students to reach this point of uncertainty without their design failing a test, physical failure provides a clear indication to learners that they do not have a complete understanding of a phenomenon.

Role of facilitation

As mentioned repeatedly, failure is theorized to provide opportunities for deeper thinking, but it certainly does not guarantee that students will look for, articulate, and design based on these other factors. Learners need opportunities to reflect on their experiences and consolidate new knowledge, and they need support in this process (Kapur, 2008; Van Lehn et al., 2003; Bransford & Schwartz, 1999).

The main way students were supported in this workshop was through interaction with the facilitators, particularly at the testing stations. The facilitators helped with the logistical aspects of the test, like lifting the tube to retrieve designs and promoting respectful queuing behavior, celebrated and commiserated with students over test results, and, most importantly, asked students many questions about their designs and the test. Many of the ideas coded in Figure 3 were expressed at testing stations while talking to facilitators. It is unlikely that all of these ideas occurred to the students during questioning, but they were articulated at these times.

Many of the questions the facilitators asked were to help students reflect on their design processes. As others have found, many students do not spontaneously engage in reflection (Schauble, Klopfer, & Raghavan, 1991; Crismond & Adams, 2012). The facilitators at the testing stations encouraged students to reflect on their design processes and decisions, and specifically how their decisions affect the designs’ performance, by asking questions about:

- the changes they made to their designs
- the results from their last test
- why they think the changes they made will lead to the desired result
- what they expect will happen in the current test
- what just happened (after a test)
- why they think a certain result occurred
- what they are going to change for the next test
- why they thought a design was working
These questions made students think about their current design and how their current design fit into their design history. Importantly, these questions were framed as wanting to know students’ ideas, rather than as co-creating knowledge with the facilitators or trying to get students to guess the answers the facilitators had already decided were correct. This was made explicit to the students in an announcement during the wind design task:

Lija: So we're noticing awesome awesome ideas and work and we're interested in knowing about what you're thinking before you test. So we've got two cameras on the testing stations and we're hoping if you feel comfortable if you could say, quickly, what it is that you hope to see happen. OK? And if you make a change if you could say why you made that change and Kerrianne and Riley and Chelsea and I will kind of prompt you, so don't worry about it, we'll help you with that piece, 'cause we'd like to get at your thinking and we want you to be noticing your thinking as well. So how, see how it can help you design something.

Another major way the facilitators supported students’ designing was by modeling how to look at the tests and what to attend to, as this can be hard for novice designers (Crismond & Adams, 2012; Sadler, Coyle, & Schwartz, 2000). For example, when talking with Cecelia in the second episode, Riley modeled a number of messages through the questions he asked her. First, by asking her to test multiple times, Riley modeled that multiple tests are sometimes necessary to really see what is happening. By asking “why” questions and specifically, “Why is it working now when it didn’t work other times?”, he showed her that, as a designer, these are questions she should be asking herself while testing. Riley also models how to turn observations into questions: when Cecelia points out “It just goes right back upside down,” Riley responds with, “Why do you think it goes right back upside down?” By flipping her observation into a question and by continually questioning her, he implicitly sends the messages: (1) that this is something that we should want to figure out, (2) that it is likely possible to figure out by closely observing the test, and (3) that he believes Cecelia is capable of figuring it out by herself. Crucially, the facilitators positioned the students as competent designers, not as students needing to be taught what to do.

While the facilitators played an important role in the design tasks and were sometimes, as with Cecelia, crucial for helping students to look closely at their designs and the test, there is more opportunity for facilitators to engage when students are struggling with getting their designs to work. As we saw with Sarah and Abby, when success comes readily, it is much more difficult for a facilitator to push students to keep grappling with a problem they have already solved.

This study was intended to build on the literature around struggle and delayed instruction and explore whether and how it applies in the context of engineering education. Therefore, it is important to highlight some distinctions between the previous literature, which looked at failure in other domains, mostly mathematics and science education, and failure as it pertains to the physical failure of a design during engineering design tasks, as in this study. As referenced above, there are different definitions for what I am referring to as failure, and it may be possible for students to experience the benefits purported by the literature without experiencing physical failure on their designs. Further, while the previous literature often dealt with problem solving situations, for which there exists a correct answer, there is no single solution to engineering design problems, as seen by the diversity of successful designs in the wind task. Another piece of
this difference is that failure in many of the previous studies consistent of a student not arriving at the right answer to a math or science problem, which is likely a very different experience for a student than having a physical object they designed and built fail a physical test. Finally, the previous literature was principally concerned with individual failure instances, where in this design task students experienced repetitive failure, and it may have been critical for students to be able to reflect on multiple instances of a test. Thus, while this literature base sets up the expectation that, as in other fields, failure in engineering education will provide opportunities for learners to reflect on and revise their ideas about a phenomenon, much more work is needed to describe the particular role failure plays in engineering design.

Conclusions and implications

Building on literature outside of engineering that suggests failure and struggle can lead to a deeper understanding of a phenomenon, this study explored whether failure in engineering design tasks can be productive in similar ways. I found that students are able to engage and persist in failure-prone physical testing cycles. In the wind task, coding the ideas expressed by students showed that most groups attended to additional factors during the task, after experiencing many failures, but that they did not drop the initial factors. The different factors themselves were not considered as more or less complex, but the way students attended to more factors and multiple factors simultaneously later in the task was considered to be a more sophisticated approach to designing. In contrast, in the water task, which featured more immediate success, students did not describe factors related to why their designs worked, but instead described features of their designs. In this task, students seemed to interpret the success as a success of the entire design, and did not have a need to tease apart the many aspects of their designs and the test that may have led to success.

This study also illustrated the importance of facilitation in design tasks, particularly around interpreting and responding to test results. Students in this workshop benefitted from facilitators modeling how to engage in testing and design cycles.

These findings have implications both for classroom instruction and for the design and study of engineering design tasks for upper elementary classrooms. In particular, this study suggests that the tasks and materials given to students are critical in determining how students engage with testing and interpreting the feedback from test results. For example, when the materials given to students during the water task allowed them to succeed after one or two tests, there was little evidence that the ideas students had about their designs and the task evolved over the course of the task. The physical test used to evaluate designs is also naturally critically important. In the wind task, the test itself is fun, failures can be exciting, and the designs are not hurt in any way by the test. However, the test is so fast that it is often not clear what caused a failure. Having tests that are straightforward to interpret should be a major consideration in creating design tasks.

Acknowledgements

I would like to thank the Center for Engineering Education and Outreach for sponsoring the workshop during which this data was collected and for supporting this research.
References


Appendix

Table A1
*Descriptions and examples of codes used to create Figure 3*

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>Example(s)</th>
</tr>
</thead>
</table>
| Weight            | Weight of materials or the entire design as a reason for design changes or test result. Not just references to material weight. | Vincenzo: Well like, like, um, I'm just kind of like if it's too light, like I'm adding some popsicle sticks  
                 |                                                                             | *NOT: “Oh these things are heavy”*                                                                 |
| Size              | Size of design or parts of a design, related to design performance           | R*: This is the first balloon I've see not float. Why do you think your balloon doesn't float?         |
|                   | Note: does not include size as related to construction, like needing a long piece of tape | Sarah: Um, it's not big enough.                                                                     |
| Air pushing/      | Air/wind pushing on a design, references to catching air or trapping air in a design. “Air” refers specifically to air coming from the fan—thus, blowing up a balloon would not be trapping air. In contrast to air flow, this code considers air as a single unit or force. | Cecelia: No wait, because this isn't helium so it will just go down and the air will be pushing it up. |
| Catching air      |                                                                             | Marco: Uh, it's a bag so it's gonna catch air like this and then the weights in there are gonna make it fall [Also coded for weight] |
| Air flow          | Air flowing through designs; air can be split up into different streams. In contrast to ‘air pushing/catching air’, ‘air flow’ codes conceive of air as something that moves. | Liam: Too light, but I have an idea. Why don't you try using the scissors to poke a hole so wind goes through it? |
| Other factors     | Design or test-related factors mentioned by students that do not fit into any of the above categories. These factors were each only mentioned by one or two groups. | Sophia: Wow that's so weird! Probably because it's not on the center of the thing. [Test result is due to tube not centered on the fan] |

**Additional coding notes**
Only student turns were coded; no facilitator speech was coded, even if they used speech that would have been coded for a student. For example, a facilitator after a test asked: “Maybe add more weight?” and the student replied, “Yeah.” Nothing in this exchange would be coded.

In general, references to design features or modifications had to be explicitly linked to factor (so, “I added popsicle sticks to make it heavier” would be coded, but “I added two popsicle sticks” on its own would not be coded) and references to a factor absent a design change had to be explicitly linked to a design’s performance or a test result (so, “I need it to go down, so I’m adding weight” would be coded, but “Yours is way heavier than mine” would not be coded).