Physical Testing and Failure in an Elementary Engineering Camp

A qualifying paper submitted by

Chelsea Joy Andrews

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In

Engineering Education

Tufts University

May 2016

Adviser: Kristen Wendell

Abstract

Testing and iterating on designs is a fundamental feature of the discipline of engineering. These tests often result in failure; within the professional world, engineers interpret this failure as feedback about their designs and alter their designs accordingly. At the elementary classroom level, where failure often takes the form of design constructions failing a physical test, it is unclear what to expect from students, as little research has looked closely at iteration in engineering with children.

This motivates the question: How do elementary students engage in failure-prone physical testing cycles in engineering design tasks? To answer this question, I analyze a case study of a pair of fourth grade students, focusing on how they evaluate failed tests and what they change in their design construction in preparation for the next test. I found that these students do see testing as a source of feedback and make intentional changes to their design construction in response to the test results. These findings imply that students are able to productively engage in iteration and see failure-as-feedback with minimal externally-imposed structure. These results have both pedagogical implications, in terms of how to create and facilitate design tasks, and methodological implications, namely the particular insights afforded by a case study approach for analyzing design tasks.

Introduction

Testing and iterating on designs is a fundamental feature of the discipline of engineering. 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 about the current design to inform design changes. As engineering is increasingly integrated into the elementary school classroom, it is critical to understand how students respond to this inevitable failure. One way to begin such an investigation is to look at what happens currently in classrooms.

Engineering design has become increasingly present at the K-12 level over the last few decades, spurred by multiple objectives, including: to help in learning concepts other than engineering (Kanter, 2010), to increase awareness of engineering as a career path, and to improve mathematics and science learning (American Society for Engineering Education, 1987; National Academy of Engineering [NAE], 2005). With engineering design explicitly included in the Next Generation Science Standards (NGSS Lead States, 2013), this trend is likely to accelerate. Despite this trend, there is a notable lack of basic research on how students engage with engineering design, especially at the elementary level (Brophy, Klein, Portsmore, & Rogers, 2008; NAE and National Research Council [NRC], 2009).

When engineering is included in elementary classrooms, it is often in the form of design tasks (Brophy et al., 2008; 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[1] (e.g., NAE and NRC, 2009). The design constructions often initially fail the physical test; that is, they do not meet the

[1] Following Wendell (2013), I will use the term "design constructions" to refer to the physical objects that students create during an engineering design task. Wendell built on descriptions by Dym (1994) and Roth (1996) to arrive at a definition of design constructions as "tangible, three-dimensional artifacts that result from some kind of engineering design process and that are created by children to perform specific functions or solve specific problems" (Wendell, 2013, p. 190). These constructions are typically rough functional prototypes that can be subjected to a physical test to evaluate whether they meet the requirements of the design problem (Benenson, 2001).

design criteria. Even though this failure is quite common, there is little literature that directly addresses what students do in this situation.

This motivates the question: How do elementary students engage in failure-prone physical testing cycles in engineering design tasks? To explore this, I investigated a pair of fourth grade students who engaged in a short design task and experienced repeated failure during the activity. While conducting the case study analysis, I attended to how they evaluated each physical test (each one was a failure) and to how they decided what to change in their design. This analysis uncovered two main aspects of their engagement: the pair see testing as a source of feedback on their design construction, and they make intentional changes to their design construction in response to the test results. These results have implications for how engineering design lessons are devised and structured: if students can productively engage in the testing and revision process on their own, then they may not require additional teacher-imposed structure.

Engineering at the elementary level: A classroom view

In elementary school classrooms, engineering tasks are implemented with various objectives in mind, including improved performance in mathematics or science (e.g., Apedoe, Reynolds, Ellefson, & Schunn, 2008; Silk, Schunn, & Strand-Cary, 2009) and increasing the number and diversity of students interested in engineering (NAE and NRC, 2009). Even so, classroom observations, descriptions from published research (e.g., Sadler, Coyle, & Schwartz, 2000), as well as curricula descriptions (e.g., Engineering is Elementary [Hester & Cunningham, 2007], Learning by Design [Kolodner et al., 2003], and others [see NAE and NRC, 2009]), reveal that the overarching organization of classroom design tasks are generally quite similar.

In these classrooms, a teacher presents students with a pre-determined problem and charges them with solving it by designing and creating a functional engineering solution. Very often, teachers also dictate the criteria and constraints of the problem and the physical test(s) used to evaluate student solutions[2]. Typically, pairs or teams of students then work together to plan and construct their designs from available materials. The groups of students then test their design constructions, and if time allows, iterate to improve their design. In this way, the classrooms loosely follow an engineering design process, with the teacher often performing the initial steps of problem and test definition (e.g., Dym & Little, 2004; French, 1985; Pahl & Beitz, 1996; Ulrich & Eppinger, 2008).

An example of a common engineering design task that follows this pattern is the egg-drop challenge (e.g., Egg Drop Competition, 2015), where students are tasked with creating an object out of craft materials that will allow an egg to survive a large fall. In this task, students build a container to hold and protect the egg; common designs use padding, suspend the egg with rubber bands, and/or attach parachutes to the container holding the egg. For the physical test, the object holding the egg is dropped from a large height and after it hits the ground the egg is checked for cracks. As may be anticipated, many initial designs (and often the next iterations, if students are given time to iterate) do not pass the physical test—the eggs break when they hit the ground. There is a lack of literature investigating how students handle this failure—what do students do in this situation?

Like in the egg drop task, failure is likely to occur in classrooms, given how engineering design tasks are often enacted. As I will review below, this is not necessarily unwelcome—failure is a fundamental part of engineering. And while there is currently not much literature focusing on failure with young learners in engineering education, research in other disciplines

[2] Some classrooms are beginning to include students in this process, e.g., McCormick & Hynes, 2012.

suggests that failure and struggle may actually contribute to learning.

Failure in engineering and STEM education

Engineering involves failure, both in finished products and during design. I take "failure" to mean that the design, design construction, prototype, or final product does not meet some intended design criteria under intended constraints. For finished engineering products, failure could thus range from unsatisfied users (e.g., a smart phone not meeting expectations), to an inefficient but still functional design (e.g., poor intersection signal timing causing traffic jams), all the way to catastrophic physical failure (e.g., a bridge collapse). 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. Importantly, failures during the design process are not mistakes; rather, they are a feature of working on complex problems incorporating many parts with "interrelationships…[that] can be difficult to analyze and predict" (Brophy et al., 2008, p. 371). For practicing engineers, failure during the design stage could range from a computer simulation showing excessive beam deflection, a focus group finding bugs in software, a concrete sample that crushes before the weight limit is reached, or, when a design involves a prototype, a physical failure during a physical test.

In their "Framework for Quality K-12 Engineering Education," Moore et al. (2014) include "Test and Evaluate" as a key indicator and a defining process in engineering. They describe the test and evaluate indicator as such:

> Once a prototype or model is created it must be tested. This likely involves generating testable hypotheses or questions and designing experiments to evaluate them.

> Students may conduct experiments and collect data (and/or be provided with data) to analyze graphically, numerically, or tabularly. The data should be used to evaluate the prototype or solution, to identify strengths and weakness of the solution, and to use this feedback in redesign. Because of the iterative nature of design, students should be encouraged to consider all aspects of a design process multiple times in order to improve the solution or product until it meets the design criteria. (p. 5)

As described above, “improv[ing] a solution or product until it meets the design criteria” likely entails encountering failure of a design construction—it is often this failure that demonstrates a design does not meet the criteria, which prompts redesign. The Framework and the state and national standards it is aligned with are grounded in theories of learning that suggest students should engage in authentic disciplinary practices. Indeed, failure could prompt students to revise and build on their ideas about why their designs are working or not (Kolodner et al., 2003). However, failure can only be productively implemented in classroom tasks if educators know what to look for in failure and how to support it. Identifying how students engage with failure in elementary classrooms can inform teaching and the design of curriculum.

At the K-12 level, and particularly in elementary classrooms, the engineering that is practiced is typically closest to civil or mechanical engineering, where the artifacts are products or scale models of infrastructure, as opposed to, for example, processes or materials. Thus, when students are experiencing failure during design it is similar to a physical failure of a prototype. Thus, from here on I will limit the discussion to that kind of designing and testing. For example, students may design bridges and test them by loading them with weights—the bridge breaking apart or collapsing before the goal weight is reached would be a physical failure.

For this kind of designing (creating physical products or infrastructure, rather than processes), practicing engineers 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)[3]. Practicing engineers create and test these constructions, then iterate using the previous test results as feedback. Inspired by this professional model, 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 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 beginning designers. They claim that novice designers tend 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), although specific studies that led to these findings are not referenced. These beginner designer patterns are described in contrast to those of informed designers, who "focus attention on problematic areas and subsystems when troubleshooting devices" (p. 749). Similarly, Sadler, Coyle, and Schwartz (2000) found that elementary students have difficulty evaluating feedback from tests, particularly in distinguishing between experimental error and legitimate feedback on their designs.

Undergraduate students have also been characterized as idea fixated, that is, they do not explore the entire design space (Crismond & Adams, 2012; Cross, 2000; Douglas, Koro-Ljungberg, McNeill, Malcolm, & Therriault, 2012; Newstetter & McCracken, 2001). The

[3] Of course, professionals must consider many other criteria and constraints that that are not physically testable, like cost, materials, and usability. Proxies for these other kinds of criteria are sometimes included in K-12 engineering.

concern is that this could be problematic for redesign after failed tests; if students are idea fixated, they may be unwilling or unable to deviate from their initial plan to use feedback from the test to create an improved design construction. On the other hand, others have noted that experts do not typically consider and choose between a wide range of design alternatives either (Atman et al., 2007; Cross, 2004).

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 (Crismond & Adams, 2012; Kolodner et al., 2003; Sadler et al., 2000; Schauble, Klopfer, & Raghavan, 1990), effectively avoiding failure, at least initially. In another approach, many published curricula feature a substantial amount of scaffolding in the form of, for example, standardized fill-in worksheets, pre-determined and limited building materials, and employing a strict step-by-step process with all students on the same step at the same time (e.g., Cunningham, 2009; 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, cautions: "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).

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).

Findings, such as those above describing the ineffective practices of beginning designers, often change depending on the methods used and the interpretations applied in engineering education research. Much of the work identifying potential engineering problem areas for students has relied on simplistic criteria, which may obscure students' productive beginning resources in these activities. Recent research by Watkins, Spencer, and Hammer (2014) on problem framing found that using an in-depth qualitative approach led to different conclusions, namely that students "demonstrate greater abilities than suggested by previous research" (p. 11). Thus, different research methods can lead to very different conclusions and recommendations.

In addition to the differences in conclusions stemming from different research approaches, how the data is interpreted also naturally plays a substantial role in characterizing design abilities. For example, Schauble, Klopfer, and Raghavan (1991) found that elementary students run confounded tests, in agreement with others (e.g., Crismond & Adams, 2012; Kuhn et al., 2000; Sadler et al., 2000), but, they still concluded that an engineering model of experimentation was more natural for children than a science model. In their view, because engineering is goal driven, knowing the effect of every variable is unnecessary, and thus changing multiple variables with each test was deemed appropriate for the practical goal of the engineering approach. Fundamentally, when the focus is on differences between students and experts, students tend to be characterized as deficient. In contrast, in studies that look closely at young students' productive beginnings and describe students without comparing their approaches to adults', a more positive view emerges and students' actions are seen as reasonable and productive.

Another reason this topic warrants further research despite the findings previously reported is the existence of a substantial body of literature in science and mathematics education that has found it productive to allow students to struggle on their own without much structure. 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). Subjecting design constructions to a physical test that often results in failure is a clear way of allowing students to "bump up against the world" and test out their ideas. The thinking that students are testing in the test and redesign process is likely both content and process related: they are testing their ideas about how the world works and their engineering design process ideas. 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, Siler, Murray, Yamauchi, & Baggett, 2003). Kapur and colleagues (Kapur, 2008; Kapur & Bielaczyc, 2012) found that students who were allowed to struggle with complex, ill-structured mathematics and science problems without support structures later outperformed the students taught with a direct instruction approach. This body of work advocates for allowing students to struggle with new problems without initial scaffolding. However, none of these investigations considered physical failure, such as that encountered during an engineering design task.

Research questions and contributions of the current study

There are few detailed descriptions in published literature considering how young students interpret failure in engineering design tasks, leaving the questions of how young students handle physical failure and how much scaffolding they require unsettled. Do students use ineffective approaches to testing and redesign that require teachers to employ didactic methods and scaffolding worksheets, as suggested by much of the engineering education literature? Does failure lead students to productively revise and refine their ideas about a phenomenon, as suggested by some of the mathematics and science education literature? Do students think failure during a physical test reflects their failure as an individual, rather than failure of their design, as teachers are concerned about?

If students are to be engaging in authentic engineering disciplinary practices, then they need to be engaging in testing and iteration and getting experience with interpreting failure as feedback about their designs. And if students are going to be engaging in these activities, then it is crucial for educators to know how to support them. The first step toward the goal of helping teachers to recognize, notice, and encourage productive engineering practice is to determine what children do "naturally" in these situations. Toward this end, in this study I focused on a pair of fourth grade students designing, creating, testing, and iterating design constructions with little adult intervention to investigate what elementary students are capable of with physical testing and failure. The guiding question in this work is: *How do fourth grade students engage in failure-prone physical testing cycles in engineering design tasks?* This broader question can be split into two related research questions:

RQ1) How do students evaluate and use the results of previous test(s)?

RQ2) What do students do in response to testing failures?

The study

Context of the study

The data for this analysis comes from a week-long engineering design summer camp for upper elementary students held at a university engineering education center. The summer camp was run as part of the Novel Engineering (NE) project[4], a research and professional development project that seeks to improve STEM education at the elementary level by using classroom literature as the context for engineering problems.

Participants

Participants were recruited by an email sent to a listserv of parents who had previously expressed interest in this type of out-of-school engineering design program. A total of 16 students, entering fourth to sixth graders, attended the camp; some students had previously attended similar camps and workshops and a few students knew each other outside of camp. The cost of the camp was $75 to offset the cost of materials (a comparable camp would cost about 5 times more). The camp facilitators, researchers affiliated with the NE project, included undergraduate engineering students, graduate education students, and university staff familiar with engineering outreach activities.

Task design

This paper focuses on a single design task from the first day of the camp. The activity was conceived to investigate the effect of having students purposefully select limited building materials in advance of a design task. It was also intended as a short ice-breaker task appropriate for the first day when students were still getting to know each other. Just before the task, students were given time to explore materials in preparation for a "mystery" design task. During this exploration phase, groups were encouraged to drop materials into a tube set over a fan (a "wind tunnel"), put materials in bins of water, and to investigate what materials were good for joining

[4] This project is funded by the National Science Foundation DRK-12 program, grant #DRL-1020243.

and for strength. Working in pairs and groups of three (for a total of six teams), students noted their observations on sticky notes (e.g., "Tape floats in water") and attached them to posters for public use. Then, the students were presented with two design task options; all but one group opted to do a boat task involving floating objects in water. The other offered task, which is the focus of this paper, is the wind tunnel task. The concurrent tasks lasted about 45 minutes (including 15 minutes of introduction and final testing time).

The goal of the wind tunnel task is to create an object out of craft materials that will hover in a plastic tube set above a vertically-oriented fan for about three seconds without flying out the top or falling to the bottom (Figure 1).

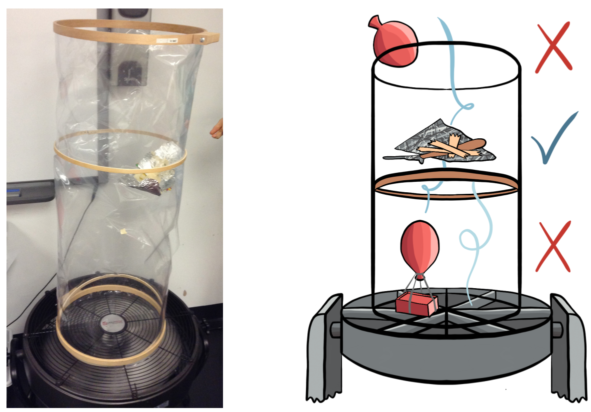


Figure 1: The test setup for the wind tunnel task, consisting of a clear plastic tube set atop an upwards-facing fan. The plastic tube is 14 inches in diameter and 40 inches tall. In the schematic on the right, the middle design is passing the test, while the two red balloon designs are failing.

For both tasks, students were instructed to choose three craft materials to use for their design construction before beginning to build, based on what they learned during the earlier exploration activity. The pair attempting the wind tunnel task chose aluminum foil, wooden popsicle sticks, and masking tape (other options included straws, balloons, cardboard, plastic wrap, and duct tape). They were allowed to use only these three materials, but they were permitted as much of these materials as they desired.

Data selection

As mentioned above, only one pair of students opted to attempt the wind tunnel task[5]. This is the main reason this data was selected for a case study: because they were the only group trying this task, they had nearly exclusive access to the testing station, which made it much easier to follow their design trajectory. The other major reason for choosing to focus on this group is that the two boys, Marco and Vincenzo[6], both entering fourth graders, are cousins and very close to each other. As a result, even on the first day of camp, when many members of other groups were still shy with their partners, Marco and Vincenzo freely conversed and were comfortable enough with each other to disagree and debate design decisions. As a result, their relationship might more closely mirror the kind of relationship between close friends in a classroom setting, who may be used to working together over several years in school. Also, their relatively constant talk offered a window into their thinking about the design and testing.

Data analysis

As a reminder, the guiding question in this work is: *How do fourth grade students engage in failure-prone physical testing cycles in engineering design tasks?,* and I answered this question by attending to two aspects of their design cycle decision making: how they evaluate and make use of previous test(s) and what they do in response to testing failures.

To investigate how young students engage in physical testing cycles during engineering design tasks I used an in-depth case study approach. Case studies are relevant to answering "'how' or 'why' some social phenomenon works" (Yin, 2013, p. 4). Employing a case study

---

[5] One other pair briefly worked on this task, but quickly switched to the other task. One other student, working independently from her team, tested a design in the wind tunnel at the beginning of the task.

[6] IRB approval was granted and students consented to the use of their real first names in research documents.

approach allows researchers to focus on and closely examine the discussions and actions of a single group over the course of an entire design task. Privileging depth (concentrating on a single or few cases) over breadth (looking at a larger number of cases) provides the space to deeply consider student thinking on a moment-by-moment basis (Flyvberg, 2006).

The entire wind tunnel design task, a 45-minute activity, was chosen as the case, as it was deemed to be short enough to examine in detail while still containing many design cycles. To prepare Marco and Vincenzo's case for study, I first transcribed the video data.

The second step involved dividing the transcript into building and testing time sections, to facilitate answering the two research questions above. Testing time sections were defined as periods of time when either or both students moved with the design construction from their table, where they were building, to the testing station and tested their design construction. Each time they went to conduct a test and then came back to their table without making any changes to their design construction was considered a single test; sometimes they dropped the design construction into the fan twice during the same test period, but both attempts always gave the same result, so they were considered a single test. Unfortunately, there was no microphone at the testing station, and the fan was quite loud, so there is only video and no audio during testing time sections. Any time students were not at the testing station was considered a building time section.

Once the building and testing segments were identified, I examined the video to collect pictures of the current state of each design construction immediately prior to each test. The testing video was then analyzed to determine the result of each testing period and the building video and pictures of each design were analyzed to determine the changes in the design that occurred between each test. To give an overall picture of the design task, I summarized the result of each test and the changes made to the design construction during each building segment. This basic overview is presented in Table 1 below.

In less than 30 minutes of build and testing time, Marco and Vincenzo conducted seven tests of their design constructions, making changes or completely rebuilding after each test (Table 1). Their design construction failed the test in every attempt until the final, public test, or test eight. As the goal of the design task is for the design construction to hover within the plastic tube, a design construction can fail the test in two ways: by flying out of the top of the tube ("flies up") or by dropping to the bottom of the tube and landing on the fan cage ("falls") (see the 'Result' column in Table 1).

Table 1: Overview of test results and responses

Test	Object	Result	Change to object following test
1 [6:00]		Flies up	Add 4 popsicle sticks (taped to foil)
2 [8:45]		Falls	Add additional set of foil wings
3 [12:00]		Falls	Start over
4 [15:30]		Falls	Continue building; add foil wings (Note: Just popsicle sticks taped together—did not really expect to work, see discussion below)

5 [21:00]		Flies up	Add 6 popsicle sticks
6 [24:30]		Falls	Add tin foil “patch” (mentioned starting over again, but not enough time)
7 [26:30]		Falls	Add extra set of foil wings
8 [28:45] (done building)		Success	

For the seven tests the pair conducted while building, all of which were failures, twice their design construction failed by flying out of the top of the tube (test 1 and test 5) and during the other five tests the design construction fell to the bottom of the tube. Both times after their object flew out of the top, the pair altered the design construction by taping on additional popsicle sticks. In response to four of the five tests that resulted in the object falling to the bottom (tests 2, 4, 6, and 7), the pair added foil onto their object. The remaining response (to test 3) was to discard their current design construction and build a new object.

Because I am interested in how Marco and Vincenzo engaged in failure-prone physical testing cycles, I was particularly attuned to their responses to each test. Therefore, after constructing the summary of the task (Table 1), I closely examined each building segment with the research questions in mind, that is, how they evaluated the preceding test and the changes they made to their design. To accomplish this, I drew on tools from discourse analysis (Gee,

2005) to analyze the video data, including both verbal and non-verbal data as evidence. By looking closely at their interactions, including their conversations, constructions, and responses to tests, I was able to explore their design decisions and to gain insight into how they were engaging in the physical testing cycles.

As mentioned above, while analyzing the data, I attended to: (1) how Marco and Vincenzo evaluate and use the previous test and (2) the changes they make to the design construction. For RQ1, I looked for references (explicit or implicit) to previous tests. For example, if Vincenzo said, "Too light, too light" as he left the testing station; I would interpret this as a direct reference to the fly-out-the-top failure in the previous test (which may or may not refer to the actual mass or weight of the object). For RQ2, I looked at the physical changes they made to the design construction and their discussions about those changes, including the scale of changes (a small addition vs. a complete redesign), whether they considered alternatives, whether they "undid" old changes, and whether they seemed confident in their decisions.

Results

For this case study approach, I present the response to each test in turn and provide excerpts of transcript for analysis. To make it easier to follow the design process of these boys, the data are presented chronologically, rather than organized thematically. The data is divided into segments beginning with one test and ending just before the next test, describing what took place during their building time at their table. This first section describes what occurred as Marco and Vincenzo were creating their first design, prior to the first test.

Before test 1: The initial design

Prior to test 1, Marco and Vincenzo create their initial design construction without too much disagreement. They cut out two foil squares and connect them together with popsicle sticks in between. They then tape popsicle sticks across the squares (perpendicular to the other center popsicle sticks) to "stabilize it." Marco decides to add two popsicle sticks to the bottom in a V-shape and then tapes a triangle of foil between those popsicle sticks (Figure 2).

After they add those V-angled popsicle sticks, Marco expresses concern that they are adding too much tape:

Marco: Dude, I think we're putting a little too much tape on this. Because, like, remember—
Vincenzo: Tape floats! Tape floats.
Marco: But it'll also add more weight.

Vincenzo's declaration of "Tape floats!" is likely referring to a discovery the pair made during the materials testing activity before the design task began. There is no video of this activity, but one of the post-it notes Marco and Vincenzo recorded about tape was: "Tape can fly on a fan," which I take to mean they tested a piece of tape over the fan and found that it either hovered or flew up. Thus, at the beginning of this design task before they have tested their first design, Marco and Vincenzo are referencing tests they conducted before the actual task started.

They then question whether they should add anything else and both answer uncertainly ("I dunno"); they make no other changes. Marco tosses the design construction into the air at their table, and it quickly drops back to the table, which seems to surprise Marco:

Marco: What?
Vincenzo: Well there's gonna be a big fan pushing on it [picks up the design construction and demonstrates air pushing up on the bottom of it].
Marco: Oh yeah. Let's test this guy.

Vincenzo's response reminds Marco that throwing the design construction in the air is not the same as the actual test, because in the actual test there is "a big fan pushing on it." At this point they get up and bring their design construction (Figure 2) to the testing station.

Test 1: The first failure response

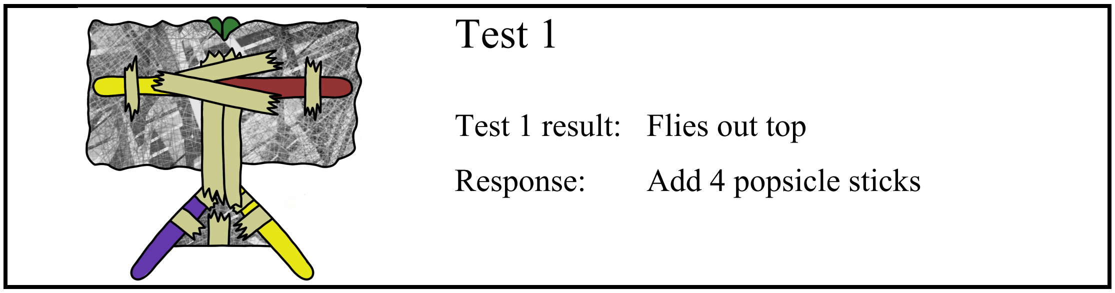


Figure 2: Schematic of design construction prior to test 1.

When Marco and Vincenzo test their first design construction (Figure 2) over the fan, it fails, flying out the top. In response, Marco and Vincenzo immediately decide to add more weight. While leaving the testing station, Marco declares "We have to put some more popsicle sticks and make it—add more weight" and Vincenzo agrees: "More weight? OK." They then debate where to add the extra popsicle sticks, and end up adding four in an "X" pattern on the foil wings of their design (Figure 3).

Thus far, the pair has engaged in an entire design cycle: they planned, created, and tested a design construction, evaluated the test result, and changed the design based on that evaluation.

Test 2: The first falling response

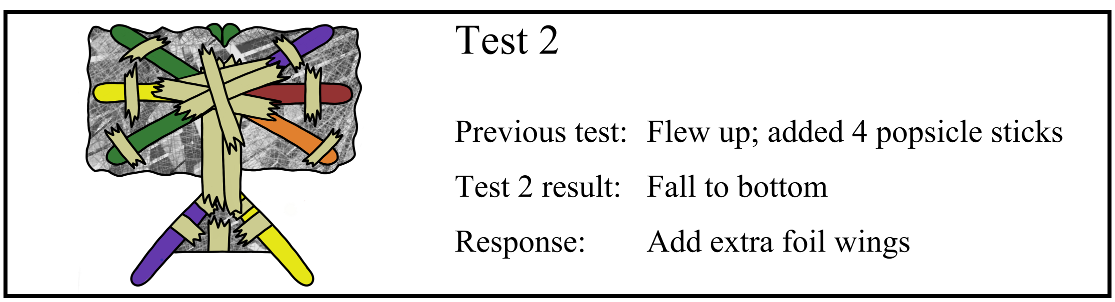
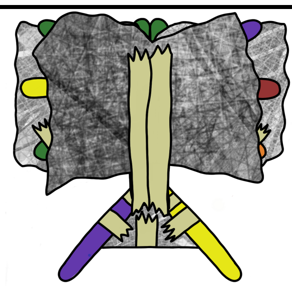


Figure 3: Schematic of design construction prior to test 2. The purple, orange, and two green popsicle sticks (forming an "X") were added during the previous build period.

In test 2, the altered design construction (Figure 3) clings to the side of the tube, then falls to the bottom of the tube. Because the design construction does not fly out of the top of the tube in the second test as happened in the first test, one can conclude that adding popsicle sticks (the response to test 1) was effective at solving that specific problem. As he collects the construction from the bottom of the tube, Marco comments, “Too much weight [pause] this happened before,” and when they return to their table Marco states his plan: “We have to make more wings, like, we’re gonna make more tin foil wings. Quick quick make squares.” Vincenzo doesn’t reply directly to these comments, and they begin looking for the materials to make more squares. It is unclear what Marco is referring to when he says, “this happened before,” because the first test flew out the top. It could be that he is referencing the activity before the design task began when the students were given free rein to test out materials in the fan, or the time Marco tested the first design construction at the table not over the fan. It does appear that he is referring to something in their shared history, considering that Vincenzo does not question or debate the phrase.

The pair then adds another layer of two wings above the original wings, with the new wings bent upwards such that from the end looking down the length of the center popsicle sticks, the wings form an ‘X’ shape (Figure 4).

Test 3: Starting over



Test 3

Previous test: Fell to bottom; added extra foil wings

Test 3 result: Fall to bottom

Response: Rip up design and start over

Figure 4: Schematic of design construction prior to test 3; the top, darker layer of foil wings is new. [Note: there is a gap between the two wing layers that is difficult to see in this plan view.]

The design construction (Figure 4) fails in test 3 when it sticks to the side of the tube near the bottom and then falls to the bottom again, the same way it failed in test 2. In response, Marco and Vincenzo return to their seats and silently contemplate their object for a few seconds. Then

Marco states, “OK, then I know exactly what to [most likely “do”; he does not finish the thought]” and begins tearing off the foil, which surprises Vincenzo:

Vincenzo: What are you doing, dude?
Marco: Let’s start over it’s too big. Take off the materials and [unclear: possibly ‘rip it’ or ‘wreck it’]. The chopsticks are the only thing that [likely means popsicle sticks; does not finish phrase].
Vincenzo: OK, we need a new idea.
Marco: You mean more like the hang glider idea?
Vincenzo: It’s a good idea [likely the “hang glider idea”] you think we should change it or keep it? [Likely means change/keep the “hang glider idea.”]
Marco: Keep it. Like this time [pause] I know what we’re gonna do [12 second pause]. Should we do foil?
Vincenzo: Huh?
Marco: Should we do the foil one? [pause] Nah. That one’s gonna be even harder.
Vincenzo: Yeah.

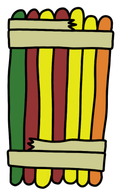
It may seem surprising that Marco and Vincenzo are willing to tear apart their object and begin again after only three (failed) tests. Vincenzo was certainly surprised at first (and sounded a bit annoyed, possibly because Marco began tearing it up without checking with him first), but he quickly calmed down when Marco responded (or he at least accepted it quickly, since once it was ripped their only option was to start again).

There are many possible reasons why Marco felt the need to rip up the design construction and start again. Marco states that “it’s too big”, but it is not clear what he means by this, as the design construction still easily fit in the plastic tube, which was the only size restriction for the task. Marco may have used “too big” meaning too heavy, as if once the object was too large and heavy it was no longer possible to get it to hover. It could have been that Marco felt it was too difficult to alter the design construction they had, a reasonable concern as it would have been difficult to remove all of the tape without destroying the object. It is interesting that Marco wrecked the design construction after it had failed twice in the same way (by falling)—perhaps he could only think of one response to that type of failure (adding foil wings), and when that response did not work, he thought there was no way to make that type of design construction work and it would be better to start again. It is also possible that Marco was worried about the imposed time constraint and thought it would be faster to build something new than to fix what they had. Because they do not discuss the reasons for starting over (other than the cryptic "it's too big"), it is impossible to know Marco's full justification. Regardless, it is noteworthy that Marco decides not to invest more time and energy into that design construction, and that Vincenzo so readily goes along with him.

After they decide to start again, the pair discusses what they should do for their next plan. If we look ahead to the object they end up with after test 4, it is perhaps surprising that their new design construction (in tests 5–8) is so similar to the original design construction (in tests 1–3)—both have popsicle sticks in the center with foil "wings" on either side. However, we see in their dialogue above that they do consider at least one other option, "the foil one" (although there is little explanation as to what that option entails), but decide that it would be "even harder." This is evidence that they at least discuss other designs, even if they do not attempt to build them. Considering that they felt they were under time pressure, it makes sense that they might stick to ideas which they think are easier to build, rather than attempt riskier designs.

Marco then tapes together a row of popsicle sticks (Figure 5) without any more explicit talk about what the new design is going to be. As he adds more tape and they discuss ripping versus cutting tape from the roll, Marco suddenly questions: "Wait, does this hover?," and gets up to go to the testing station. At the station before he tests, Marco tells Vincenzo, who has a doubtful expression on his face, "If this hovers I'll just be like" in an excited fashion. Both of these comments imply that Marco thinks it would be exciting if the popsicle sticks could hover, but that it is not to be expected, and thus he has to justify the test to Vincenzo. It thus seems that test 4 is different from the other tests, as it appears that Marco and Vincenzo did not actually expect the design construction at this point to pass the test. The comments show that it was more of a test of opportunity.

Test 4: An opportunistic test



Test 4

Previous test: Fell to bottom; ripped up

Test 4 result: Fall to bottom

Response: Add foil wings

Figure 5: Schematic of design construction prior to test 4.

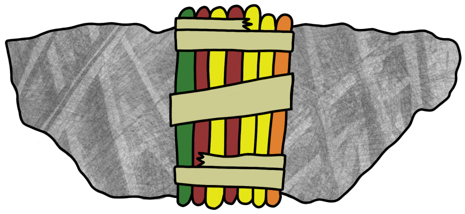
Test 4 was a very quick test, it took about 20 seconds to get up from the table, test the popsicle sticks (Figure 5)—which dropped straight to the bottom—and return to the table. They do not seem surprised and simply continue building, by agreeing to add wings:

Marco: Now we just put wings.
Vincenzo: Yeah, now we just put wings.

They do not seem upset about having to add wings and begin singing and whistling while building. The way they continue working is further evidence that test 4 was more of an unexpected interruption to their building than a test of whether their design construction would work. The test station was described to the students as a place for testing out design constructions or pieces of design constructions in order to gain information to iterate and improve their design constructions. It is noteworthy that Marco also saw the test station as providing an opportunity to try out ideas that he did not necessarily expect to work.

Once the wings are added, the next design construction is ready to test (Figure 6).

Test 5: "Didn't you see that thing fly?"



Test 5

Previous test: Fell to bottom; added wings

Test 5 result: Fly out the top

Response: Add 6 popsicle sticks

Figure 6: Schematic of design construction prior to test 5.

After they add wings, Marco and Vincenzo test the new design construction (Figure 6), and it sticks to the side of the tube and then flies out of the top. The only other time the design flies out of the top of the tube was during test 1. As they catch the design, Vincenzo remarks: "Too light, too light" and Marco agrees: "Put more weight on it." They agree to add popsicle sticks, which was the same reaction they had to test 1. Since that response worked after test 1, in the sense that the object no longer had the problem of flying out of the top, it is reasonable that they employ that option again. However, they disagree on how many popsicle sticks are needed to fix their design. As Marco attempts to tape a pile of popsicle sticks to the middle of their design, Vincenzo tries to pull the tape away:

Vincenzo: Now don't put too many. No no no I don't think you should do that [trying to grab tape away]. (See Figure 7, left panel.)

Marco: We'll just put a mountain of tape. Put more right [pointing] stack stack. Dude we need a lot more tape. [Vincenzo is trying to cut tape, Marco moves his hand away.] Didn't you see that thing [the design construction] fly? (See Figure 7, right panel.)

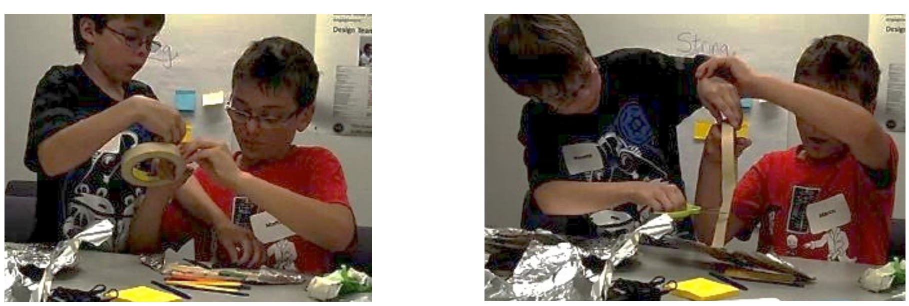


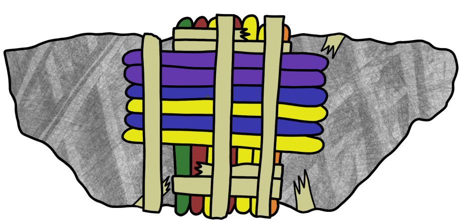
Figure 7: After test 5, Marco [red shirt] attempts to add many more popsicle sticks, Vincenzo [black shirt] disagrees and tries to pull them off (left panel), and Marco appeals to the previous test result and moves Vincenzo's hand away as Vincenzo tries to cut less tape (right panel).

In this exchange, Marco appeals to the preceding test as evidence to support his proposal to add more popsicle sticks: “Didn’t you see that thing fly?” While they never explicitly discuss what occurs during each test, all of their design changes seem to be in direct response to the test results, and this comment reflects that shared understanding. After a bit more work, the same disagreement arises again when Marco wants to tape on an additional “pack” of popsicle sticks:

Marco: Tape that on.
Vincenzo: Wait, dude, let’s just test it with that first.
Marco: Two packs.
Vincenzo: No, dude, just test it.
Marco: Yeah, we’re going to test it.
Vincenzo: Test it and THEN if it’s too light we’ll put this [the “pack” of popsicle sticks] on.

Here, Vincenzo is still trying to stop Marco from adding more to their design construction, but instead of trying to convince Marco with an argument or physically prevent him from adding on (as in Figure 7), Vincenzo instead makes a bid to test the design. It takes him three verbal attempts, but Vincenzo does convince Marco to test before adding more popsicle sticks. Vincenzo wants the test to serve as the arbiter for their dispute, and it is hard for Marco to argue with Vincenzo’s logic of testing first, given their unlimited access to the testing station, how little time testing takes, and the fact that the test does not adversely affect their design construction (as it may in other design tasks, e.g., when testing bridges with weights, where a bridge may break if it fails). Through this exchange, Vincenzo convinces Marco to test the new design (Figure 8) with only one “pack” of six popsicle sticks.

Test 6: A new foil response—a tin foil patch



Test 6

Previous test: Flew up; added popsicle sticks
Test 6 result: Fall to bottom
Response: Add tin foil patch

Figure 8: Schematic of design construction prior to test 6.

After the design construction falls in test 6 (it flew up in test 5), Marco immediately makes a bid for more foil:

Marco: Dude, put more tin foil on.
Vincenzo: No.
Marco: We're not wrecking this because we don't have time to make a—
Vincenzo: —Well let's put a tin foil patch on the bottom.
Marco: Yup. Patch tin foil on bottom. Interesting.

Here again Marco immediately wants to add foil, while Vincenzo is initially resistant. However, when Marco responds "We're not wrecking this because we don't have time to make a—", and we can assume he was about to say something like "new one," Vincenzo then puts forth a new option: to "put a tin foil patch on the bottom," which Marco agrees to go along with.

One interpretation of this exchange is that Vincenzo hesitates to "put more tin foil on" because he remembers that adding the foil wings earlier (before test 3) did not stop the object from falling again in that test. Thus, this hesitation on Vincenzo's part could be seen as evidence that they are using the results of earlier tests to inform current design decisions. (Vincenzo also may be beginning to doubt Marco's ideas in general, since it was Marco who wanted to add more wings after test 2, who wanted to test just the popsicle sticks in test 4, and who wanted to add even more popsicle sticks after test 5, and none of these suggestions panned out.)

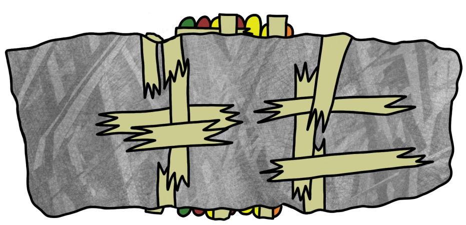
However, when Marco eliminates the other response they tried before—starting over—Vincenzo may then feel trapped by the time limit to find a relatively quick fix. It is not clear where the idea for the tin foil patch comes from, but it makes sense that Vincenzo would not want to add wings, because that did not work earlier, after test 2. There is no evidence Vincenzo recalls each test they have conducted and the result, but it is likely Vincenzo recalls adding foil as a test response, and it is also quite likely that he remembers that every test thus far has failed, so it is reasonable to believe that he remembers adding foil has not been successful.

After they have taped on the foil patch they are almost ready to test the new iteration of the design construction (Figure 9):

Marco: Think it's good?
Vincenzo: Not yet, one more thing of tape.
Marco: If it's too light we got the patch [holding the popsicle sticks he made before test 6].

Marco's comment here about the popsicle sticks he taped together before test 6 ("If it's too light we got the patch"), may imply that Marco feels confident that the added tin foil patch will at least stop the object from falling to the bottom, and could even have too much of an effect, so that it may now be too light. Alternatively, Marco may just be commenting that in the case of a fly-out-the-top failure, they already have a solution constructed. However, when they test it (test 7) it falls again.

Test 7: Last minute extra wings



Test 7

Previous test: Fell to bottom; added foil patch
Test 7 result: Fall to bottom
Response: Add extra pair of foil wings

Figure 9: Schematic of design construction prior to test 7.

As their design construction floats to the bottom, Vincenzo immediately states:

Vincenzo: Too heavy.
Marco: More tin foil.

Just like after test 6 ("Dude, put more tin foil on"), Marco again immediately makes a bid for adding more tin foil. After test 2, Marco specifically states that they should add more wings ("We have to make more wings, like, we're gonna make more tin foil wings"); it is therefore unclear whether Marco is using "more tin foil" as a shorthand for saying "more tin foil wings" or whether he thinks tin foil in any configuration would make the design construction not fall to the bottom. Either way, even though adding foil wings did not work after test 2 and adding the foil patch after the last test (test 7) did not work, Marco still wants to try adding more foil (in the shape of wings or something else) this time.

A few seconds after they get back from the test the camp director gives a one-minute warning. Marco is now taking charge of the building and mostly ignores Vincenzo, and Vincenzo does not make any attempt to interfere with Marco's building. Marco flips the object over and adds another set of wings over the original set, on either side of the patch, and bends them so they are pointing up (Figure 10).

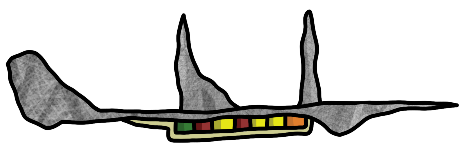
The fact that Marco and Vincenzo add wings again here, after that response failed after test 2, could be seen as evidence that they are approaching each test as a unique event and not chaining together the results from different tests. On the other hand, Marco bends these wings upwards, so it is possible that he was attempting to add a different kind of wing here than they used before. Their next conversation can be seen as further evidence that they are recalling previous tests.

They decide to test the newest iteration (Figure 10), but as Marco stands up the camp director gives the announcement to stop building so the final tests can begin. Marco and Vincenzo seem very concerned that they will not be able to test their latest iteration before the final public test:

Marco: [Quietly] What the hell.
Vincenzo: I don't know, dude, I dunno.
Marco: You think we lost?
Vincenzo: I dunno, dude, I dunno.
Marco: Me neither.

In this exchange, Marco and Vincenzo seem very distressed and worried ("What the hell" is a strong expression for these young boys). After test 7, they had just failed twice in the same way (falling) and were running out of time. When Marco then adds wings, it could be that he is doing so as a last resort and because there is no time to think of something else, and not because he truly believes wings are the best option. Their nervousness about not being able to test their final design supports this conclusion. Thus, it is still possible that they remember that adding extra foil wings did not solve their design problem the first time they employed that response (after test 2), but given the context in which they are working, where they only have three materials available, are now down to very little time, and already know how to build wings, adding more foil wings seems reasonable, even if it did not work earlier.

Test 8 (Final, public test): Unexpected success



Test 8 (Final test)

Previous test: Fell to bottom; added extra wings

Test 8 result: Success

Figure 10: Schematic of design construction prior to the final test (test 8). [Schematic is a front view, flying out of the page, the others were plan view.]

In the final public test, Marco and Vincenzo's design construction surprisingly (to them and the facilitators) passes, by hovering for three seconds. As they return to their seats, they congratulate each other and brag to a facilitator about how many versions they had:

Vincenzo: Dude, we're awesome! We did both challenges! [They were successful in both the hover challenge and a bridge challenge that came before it.]
Facilitator: That was really cool. I love that I saw yours kinda progress and make some changes.
Marco: Yeah, we made like—I think we made like 5, 4, 3, um like 6 [design constructions].

It is noteworthy that Marco seems proud about how many different versions they had to make and test to arrive at a successful design construction.

Summary of tests and changes

The following schematic (Figure 11) gives a succinct overview of the eight tests conducted by Marco and Vincenzo and the changes the pair made to their design construction between tests. The information in the schematic is similar to that provided in Table 1, with more detail about the design changes, based on evidence from their dialogue provided in the results above. In particular, the schematic distinguishes between small and large-scale changes and changes that were enacted and those that were only considered (shapes in dashed lines show considered changes). Starting over (which occurred once and was considered one other time) is the only change considered to be a large-scale change.

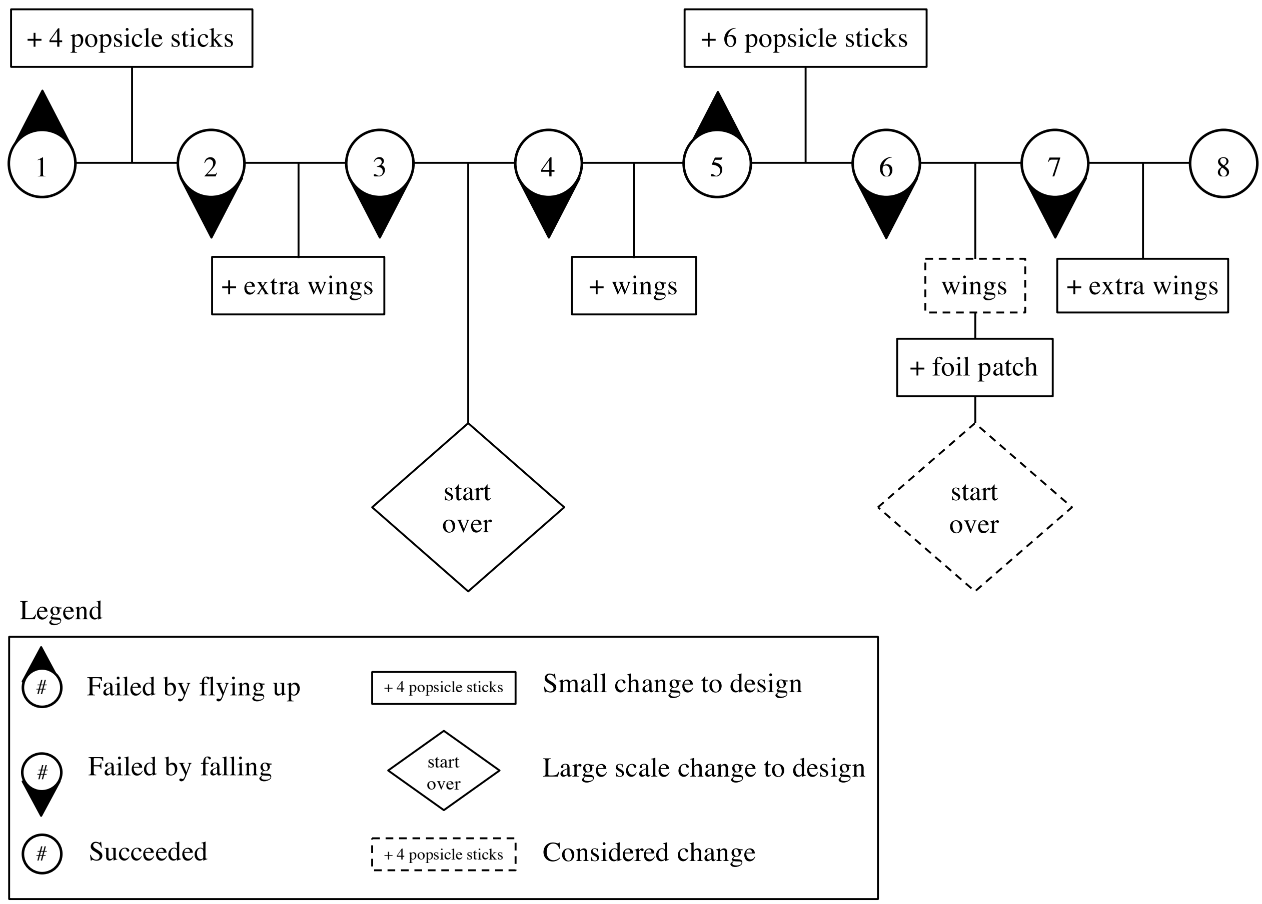


Figure 11: Schematic of all eight tests and changes to design construction.

Discussion

When professional engineers get to the point where they have an idea for a physical design to solve a problem, they often create physical prototypes or computer simulations to test those designs. They test and iterate on these designs using the results of previous tests to inform the changes they make to their subsequent designs. When students are encouraged to test their design constructions multiple times during building, it is with the assumption that they will gain insight into how their designs work from those tests and apply the test results as they continue designing—that they will use the physical failures as feedback about their designs. Recently, researchers have become concerned that this assumption is not warranted with novice designers, that they need more support to effectively test and iterate on their designs. However, these studies have not looked in detail at how young students handle the test and redesign cycle. This study intends to help fill that gap in the literature.

Before answering the research question posed in this study: *How do fourth grade students engage in failure-prone physical testing cycles in engineering design tasks?*, I will first address whether fourth grade students are *able* to engage in these testing cycles; the short answer to that question is yes. Even the most cursory analysis of the summary of their design cycle, given in Figure 11 (with pictures in Table 1), makes it clear that Marco and Vincenzo are indeed iterating in the most basic sense of the word: they test their design construction and make changes to it (or create a new object) before testing again. They do not create a new object for every test, continue to test the same design repeatedly, or abandon the task. With this basic question answered in the affirmative, the remaining question is: how do they engage in these cycles?

In analyzing this case study in terms of "how" they engaged, I attended to two aspects of their design cycle decision making: how they made use of previous test(s) and what they did in response to testing failures. Attending to those two aspects revealed that Marco and Vincenzo (1) see testing as a source of feedback on their design construction, and (2) make intentional changes to their design construction in response to the test results. Below I review the evidence for these findings, and derive both pedagogical and research implications from the results.

RQ1 finding: Testing-as-feedback

A necessary but not sufficient condition for effective iteration is to see testing as a source of feedback about a design construction. This requires an appreciation of the need for empirical evidence about a design's functioning and a view of the test as a source of that evidence (e.g., Moore et al., 2014). While in many cases there may be other non-empirical objectives in justifiable competition with physical test results (e.g., aesthetics when designing a phone), in this task the physical test was the sole arbiter of a successful design. Engineering requires a way to compare a design construction against the intended criteria and constraints, which in many cases is accomplished by an empirical test.

The first time Marco and Vincenzo demonstrate that they see a need for empirical evidence occurs just before test 1. When Marco is satisfied with the first iteration of the design, he throws it up in the air to try it out. At this point, they have a design construction that, based on their intuitions about how the world works, they think may succeed, and they recognize the need to try it. While the need to try it out may seem obvious, and probably was obvious to them, it may not be so for everyone—we can imagine that very young children would be satisfied if their construction resembles something that works and not see a need to test it. Indeed, Crismond and Adams (2012) claim that "beginning designers run few or no tests on their design prototypes" (p. 765). When the design construction falls back down, Marco seems surprised, saying, "What?" Vincenzo argues that the test in the actual wind tunnel will be different: "Well there's gonna be a big fan pushing on it," and Marco quickly agrees: "Oh yeah. Let's test this guy." Vincenzo recognizes that the test in the wind tunnel above the fan is not replicated by throwing the design in the air—he demonstrates as he speaks that there will be an upward force on the design construction from the "big fan." Here we see that Marco and Vincenzo both see the need to test their design, and recognize that the test setup is critical, or at least is not equivalent to throwing the design in the air.

Other evidence that Marco and Vincenzo see the test as a source of feedback is found in their dispute after test 5, when they both appeal to the test to make their cases. In test 5 the construction fails by flying up and they agree that they need to "Put more weight on it." As Marco tries to convince Vincenzo to add more popsicle sticks and tape to the design construction, he exclaims in a frustrated tone: "Didn't you see that thing fly?" This rhetorical question clearly demonstrates that Marco is basing his design decisions on the test results and his phrasing shows that he expects Vincenzo to have the same understanding. Soon after this, Vincenzo makes a bid to test the construction before making more changes. When Marco attempts to add even more onto the design, Vincenzo argues three times to test it first, before making more changes, with a final appeal of: "Test it and THEN if it's too light we'll put this [a 'pack' of popsicle sticks] on." Instead of trying to convince Marco using verbal arguments, Vincenzo proposes they obtain more empirical evidence, from the test.

In this building period after test 5, both students appeal to the test in different ways to argue for opposite ends: Marco claims that the previous test proves they need to add even more onto their construction, and Vincenzo argues that they should just test again before adding any more, that they can always add more on.

Another thing to consider is what kind of feedback Marco and Vincenzo get from the test. At the most basic level, the test has three clear results: the object can hover (a success), or it can fail by either flying up or falling down. After each of the tests, Marco and Vincenzo do not explicitly state the test result, but rather they jump straight to commenting on either an interpreted fault in the current design ("too much weight" after test 2 or "too light, too light" after test 5), or a change they want to make to the design ("…add more weight" after test 1 or "put more tin foil on" after test 6). Most of these comments point to a discrete or Boolean evaluation of the test results: either fail-up or fail-down (or pass, but that does not occur until the final test). This type of reading of the test would not be surprising in such a clear-cut test, especially considering how fast each test is. However, as mentioned above, while arguing over how much tape and popsicle sticks to add, Marco questions, "Didn't you see that thing fly?" This rhetorical question implies that Marco and Vincenzo are evaluating the test result as more than just a Boolean up/down—they interpret degrees of failure.

Of course, it is not too surprising that Marco and Vincenzo choose to test their design construction, because the task was explicitly set up with passing the test as the sole goal, and they were encouraged to test often. However, even with the understanding that the goal was to pass the test, they could have viewed the test as an exam, with a pass/fail grade, and not as a source of feedback. If students did not see the test as a source of feedback, they might exhibit different behaviors, such as continuing to test the same thing over and over, possibly blaming the test for the failures (if they were convinced their design was good), they might test a completely different design every time (if they had less faith in their constructions), or they may disengage after a failure and stop working on the task.

RQ2 finding: Intentional changes based on test results

With the above result that Marco and Vincenzo saw testing as a source of feedback on their design, the next question is how they used that feedback. Overall, we see that Marco and Vincenzo make changes to their design constructions in response to their interpretation of the test results. Below are the main findings on how Marco and Vincenzo went about deciding what to change in their design constructions.

Immediately make changes

Marco and Vincenzo immediately begin to make changes to their objects after returning from tests. They spend very little time (at most about 30 seconds and often less than 10 seconds) discussing what to change before they start collecting new materials, cutting, taping, or ripping up the previous design. Also, all they ever do before changing the design construction is talk—they do not demonstrate possible changes with their bodies or materials or draw ideas. Once they begin making changes they do occasionally ask each other questions about the alterations by demonstrating with materials.

Small and large-scale changes

Marco and Vincenzo consider and implement both small and large-scale changes to their design constructions. The small changes include taping on more popsicle sticks (after tests 1 and 5) or taping on more foil (after tests 2, 4, 6, and 7). The only large-scale change, after test 3, was to start over and create a new object from new materials. Starting over was also considered after test 6, but that option was rejected due to limited remaining time.

The large-scale change occurs after test 3, when Marco decides to destroy their original design construction and begin again. It may seem strange that Marco and Vincenzo are willing to tear apart their object and begin again after only three tests. Alternatively, it could be seen as remarkable that they started over at all, as other researchers have found this failure response to be rare: in a design task building bridges with fifth graders, Roth (1996) found that in over 50 projects, only one group chose to abandon their current design and begin again. This is consistent with the behavior of beginning designers described by Crismond and Adams (2012), some of whom "will not abandon their design ideas, even after running many tests and design iterations that clearly demonstrate a plan's ineffectiveness" (p. 767). They attributed this trend to ineffective diagnostic troubleshooting: because beginning designers do not actively look for "worrisome" patterns, they do not recognize "flawed performances," resulting in final designs that are "strikingly similar" to the initial plans (p. 767). Whether or not Marco and Vincenzo were engaging in effective diagnostic troubleshooting, they certainly recognized their tests as "flawed performances" and Marco at least did not hesitate to start over, even after only three failed attempts.

On the other hand, in a design task of constructing a tower out of paper, Welch (1999) found that many students abandoned their designs and started fresh, often multiple times in a single task. It seems reasonable that the complexity of the task, construction difficulty, materials, and time limit greatly influence whether students are likely to abandon a design construction and begin again.

While the literature characterizes students as idea-fixated, based in part on how infrequently they make large-scale changes to designs (Douglas et al., 2012; Cross, 2000; Newstetter & McCracken, 2001; Crismond & Adams, 2012), there are certainly occasions when both small and large-scale changes are warranted, so there should not be an automatic value judgment on the scale of the changes. It may be more productive to focus on whether students are making reasonable design decisions considering all of the constraints they are under.

Responding to previous test vs. chaining of tests

I also considered which test results Marco and Vincenzo took into account when making changes to their constructions: Do they make changes solely in response to the immediately preceding test, or do they consider the entire history of tests? If they only attend to the previous test result, it could imply that they see each test as a unique event, rather than as a single instance of a series of events. In order to effectively and efficiently create a design to solve a problem, it is critical to be able to chain together the results from multiple earlier tests in addition to incorporating feedback from the previous test. Some researchers have claimed that this is difficult for novice designers to do, that they may instead approach each test as a separate event (Crismond & Adams, 2012).

There is considerable evidence that Marco and Vincenzo attend to the previous test when making changes to their construction. For example, in response to test 1, an out-the-top failure, they add popsicle sticks to add weight to their design, which then falls in test 2. It therefore may seem counterintuitive that Marco and Vincenzo do not respond to this fall-to-bottom failure by removing some of the popsicle sticks they added before test 2. This could be seen as evidence that the boys are approaching each failure as a unique event and not thinking in terms of their past design trajectory. Because they do not mention it in their speech, it is impossible to know whether they silently considered and rejected the idea of removing some popsicle sticks. To be sure, with the amount of tape attaching the popsicle sticks to the foil, it may have been impossible to remove the popsicle sticks without destroying the object. The fact that Marco and Vincenzo only ever add and never subtract from their design may be considered evidence that they are only thinking in terms of their previous test, or it could be an artifact of how physically difficult it is to undo masking tape, or both.

However, the fact that it is easy to see how each change is informed by the previous test result is not necessarily evidence that they are not also considering earlier tests. Looking closely at their dialogue, and not just the actual design changes, reveals tentative evidence that Marco and Vincenzo are chaining together earlier test results.

For example, after test 6, Marco wants to add more foil (“Dude, put more tin foil on”), but Vincenzo resists (“No”). When Marco claims they do not have enough time to build a new construction, Vincenzo suggests adding a tin foil patch. It is possible that Vincenzo recalls that adding wings after test 2 did not work—in test 3 the design construction fell again—and instead tries to find a new solution to their falling problem. If the boys continually added wings every time the design construction failed by falling, that could be evidence that they are viewing each test as a unique event and trying to same solution every time. By trying different solutions in response to failure-by-falling, the boys seem to recall that their past attempts were unsuccessful.

Of course, they do add wings another time—after test 7, when the patch was unsuccessful and the design construction fell again. This is the point where they have run out of time and do not have a chance to test again before the final test. This time Marco bends the wings until they are almost vertical; this change in shape could be seen as an attempt to add a different kind of extra wings, which could be evidence that he recalls that extra wings did not work the first time (after test 2). Additionally, after they add wings this time Marco and Vincenzo seem very nervous, which may be a sign that they have not forgotten that wings did not work before, and that they added them only as a last resort. Certainly, if they had full faith in the extra wings, they would not have whispered to each other, "What the hell" (quite strong language for these young boys) and, "You think we lost?"

This tentative evidence that Marco and Vincenzo are chaining together the results from earlier tests is not intended as a contradiction to other studies that have found that students struggle with this aspect of iteration. There is certainly a need for more studies looking closely at this point before strong conclusions can be made.

Finally, while we certainly want students to view each test as one instance in a series of tests, and not approach each test as a singular event, it is also important to remember that as constructions change, old findings are unlikely to be directly applicable. This is particularly true when building with craft materials, as opposed to a standardized building set like Legos. For example, adding extra wings after test 2 did not stop the construction from falling again in test 3; however, that does not automatically mean that adding extra wings would never work—in fact, we see that the second version of the construction succeeded after adding extra wings (with a slightly different shape) after test 7. Evaluating what aspects of a changed design led to certain outcomes is not a simple straightforward task with such complex design constructions. Thus, researchers should be cautious when drawing conclusions about whether and to what extent students are chaining together test results.

Note on task design

The wind tunnel task presented here differs in a number of ways from other common design tasks: (1) the tests are very fast, making it difficult for students and facilitators to observe what factors led to failures; (2) the design constructions are not hurt or broken by the test (as opposed to, e.g., a weighted bridge task); (3) failure is exciting, especially when the design flies out the top of the tube; and, (4) the designs are judged by whether or not they pass but not by the degree to which they succeed (like building the tallest tower or fastest boat). These features correspond to both benefits and drawbacks of using this task to encourage iteration: on the negative side, per feature (1), it is more difficult to evaluate the test if the failure occurs so fast that it is difficult to discern what happened. On the other hand, the fact that designs are not hurt by testing may lower the barrier to testing and make fast iterations easier, the fun failures may make it less stressful to see a design fail, and the lack of ranking successes may make the task less competitive as everyone can succeed equally.

Certainly, task design is critical, both for pedagogical and research goals. Although this study was not conducted to evaluate task design, the most salient feature of this wind tunnel task—the quick iterations—is important to discuss. These results support the recommendation of Sadler, Coyle, and Swartz (2000) to have design tasks with many, fast iteration cycles. In this study, the boys were able to test their first design after about six minutes of building time, and it seems reasonable that this short build time is part of the reason that they were not overly disappointed with the first failure and continued building without interruption. The many, fast iteration cycles in this task stand in contrast both to tasks that never have students test, such as the playground design task used by Atman and Adams (e.g., Atman et al., 2007), and tasks that have students test only once at the end of a longer build time, like in many egg drop challenges. If students learn by "bump[ing] up against the world" (p. 82) as articulated by Bransford and Schwartz (1999), then many testing cycles allow them to bump up more often. Finally, having many iteration cycles may allow teachers or facilitators more access into the children's thoughts; here, the fact that Marco and Vincenzo continued trying to add wings, despite the early failures of wings, may suggest something about how they understand drag.

Conclusions and Implications

From an analysis of a case study of two fourth grade boys working on a design task in an engineering design summer camp, I have provided evidence that elementary students are able to engage in the test and redesign process. I found that Marco and Vincenzo saw testing as a source of feedback about their design. For example, both students appealed to the test to justify their design decisions. When Marco questions, "Didn't you see that thing fly?," he invokes a shared understanding that extreme results justify extreme changes (in this case, adding many additional popsicle sticks). I also found that the students made intentional changes to their design construction in response to the test results. Most of these changes appear to be directly attributable to the immediately preceding test, but there is tentative evidence that they are chaining together results from different tests over the course of the design task. For example, after repeated fall-to-the-bottom failures (4 out of the 6 failures at that point), a new design alteration was put forth: to add a tin foil patch to the bottom of the construction.

Many engineering design curricula do not allow students to fully engage with iteration, including failure. Some programs have students start with successful designs and thus potentially avoid failure; other tasks only include a single, final test, not allowing iteration. Curricula that do allow students to fail often include a substantial amount of scaffolding, under the assumption that students cannot productively engage in failure without these supports. If taken too far, these approaches may interfere with students' opportunity to engage in authentic disciplinary practices, which in engineering include testing and evaluating test results, of which interpreting failure-as-feedback is an integral aspect. The findings from this study show that elementary students are able to engage in the test and redesign process on their own, including handling repetitive failure.

These findings have both pedagogical and research implications. Pedagogically, there is tentative evidence that repeated failure motivated a need for new ideas. In addition, the repeated failure may provide teachers a window into students' ideas. In terms of research, this case study method uncovered benefits of failure that have been difficult to ascertain using more common approaches, which typically aggregate the work of many students and juxtapose student performance with that of experts. Finally, this detailed description of two students engaging in the testing and iteration cycles gives insight into how failure in engineering differs from failure in other disciplines.

References

ASEE (American Society for Engineering Education). 1987. A National Action Agenda for Engineering Education. Report of an ASEE Task Force. Washington, D.C.: ASEE.

Apedoe, X., Reynolds, B., Ellefson, M., & Schunn, C. (2008). Bringing engineering design into high school science classrooms: The heating/cooling unit. Journal of Science Education and Technology, 17, 454–465.

Atman, C., Adams, R., Cardella, M. E., Turns, J., Mosborg, S., & Saleem, J. (2007). Engineering design processes: A comparison of students and expert practitioners. Journal of Engineering Education, (October), 359–379.

Benenson, G. (2001). The unrealized potential of everyday technology as a context for learning. Journal of Research in Science Teaching, 38(7), 730–745. doi:10.1002/tea.1029

Bransford, J., & Schwartz, D. (1999). Rethinking Transfer : A Simple Proposal With Multiple Implications. Review of Research in Education, 24, 61–100.

Brophy, S., Klein, S., Portsmore, M., & Rogers, C. B. (2008). Advancing engineering education in P-12 classrooms. Journal of Engineering Education, 97(2), 1–19.

Crismond, D., & Adams, R. (2012). The informed design teaching and learning matrix. Journal of Engineering Education, 101(4), 738–797.

Cross, N. (2000). Engineering design methods: Strategies for product design (3rd ed.). New York: John Wiley & Sons.

Cross, N. (2004). Expertise in design: an overview. Design Studies, 25(5), pp. 427–441.

Cunningham, C. M., Knight, M. T., Carlsen, W. S., Kelly, G. (2007). Integrating engineering in middle and high school classrooms. International Journal of Engineering Education, 23(1), 3–8.

Cunningham, C. M. (2009). Engineering is Elementary. The Bridge, 30(3), 11-17.

Douglas, E. P., Koro-Ljungberg, M., McNeill, N. J., Malcolm, Z. T., & Therriault, D. J. (2012). Moving beyond formulas and fixations: solving open-ended engineering problems. European Journal of Engineering Education, 37(6), 627–651.

Douglas, J., Iverson, E. & Kalyandurg, C. (2004). Engineering in the K-12 classroom: An analysis of current practices & guidelines for the future. Washington, DC: American Society for Engineering Education.

Dym, C.L. (1994) Engineering Design: A Synthesis of Views. Cambridge, UK: Cambridge University Press.

Dym, C., and P. Little (2004). Engineering Design: A Project-Based Introduction. Hoboken, New Jersey, John Wiley & Sons, Inc.

Egg drop competition. (2015, October 1). In Wikipedia, The Free Encyclopedia. Retrieved 15:59, October 16, 2015, from https://en.wikipedia.org/w/index.php?title=Egg_drop_competition&oldid=683651386

French, M.J. (1985). Conceptual Design for Engineers, The Design Council, London, and Springer-Verlag, Berlin.

Flyvbjerg, B. (2006). Five Misunderstandings About Case-Study Research. Qualitative Inquiry, 12(2), 219–245.

Gee, J. P. (2005). An introduction to discourse analysis: Theory and method. (2nd edition) (pp. 1-19; 94-117). New York: Routledge.

Hester, K. and C. Cunningham (2007) Engineering is Elementary: An Engineering and Technology Curriculum for Children. American Society for Engineering Education Annual Conference and Exposition, Honolulu, Hawaii.

Kanter, D. (2010). Doing the project and learning the content: Designing project-based science curricula for meaningful understanding. Science Education, 94(3), 525–551.

Kapur, M. (2008). Productive Failure. Cognition and Instruction, 26(3), 379–424.

Kapur, M., & Bielaczyc, K. (2012). Designing for Productive Failure. Journal of the Learning Sciences, 21(1), 45–83.

Kolodner, J.L., P. Camp, D. Crismond, B. Fasse, J. Gray, J. Holbrook, S. Puntambekar, and M. Ryan (2003). Problem-based Learning Meets Case-Based Reasoning in the Middle-School Science Classroom: Putting Learning by Design into Practice. The Journal of the Learning Sciences, 12(4), 495-547.

Kuhn, D., J. Black, A. Keselman & D. Kaplan (2000). The Development of Cognitive Skills To Support Inquiry Learning, Cognition and Instruction, 18:4, 495-523.

Lottero-Perdue, P. S., & E. A. Parry (2014). Perspectives on Failure in the Classroom by Elementary Teachers New to Teaching Engineering. American Society for Engineering Education Annual Conference and Exposition, Indianapolis, IN.

Marulcu, I., & M. Barnett. (2012). Fifth Graders' Learning About Simple Machines Through Engineering Design-Based Instruction Using LEGO™ Materials. Research in Science Education, 43(5), 1825–1850.

McCormick, M. and M. Hynes. (2012). Engineering in a Fictional World: Early Findings from Integrating Engineering and Literacy. Conference Proceedings, Collection: Proceedings from the American Society of Engineering Education Annual Conference, San Antonio, Texas.

Moore, T. J., A. W. Glancy, K. M. Tank, J.A. Kersten, K.A. Smith, and M.S. Stohlmann. (2014) A Framework for Quality K-12 Engineering Education: Research and Development, Journal of Pre-College Engineering Education Research (J-PEER): Vol. 4:1, Article 2.

National Academy of Engineering. (2005). Educating the engineer of 2020: Adapting engineering education to the new century. Washington, DC: The National Academies Press.

NAE (National Academy of Engineering) and NRC (National Research Council). 2009. Engineering in K–12 Education: Understanding the Status and Improving the Prospects.

National Research Council. (2011). A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas. Washington, D.C.: National Academies Press.

Newstetter, W.C., and W.M. McCracken, “Novice Conceptions of Design: Implications for the Design of Learning Environments,” Design Knowing and Learning: Cognition in Design Education, Amsterdam, Netherlands; New York, NY, 2001, pp. 63–78.

NGSS Lead States. (2013). Next Generation Science Standards: For States, By States. Achieve, Inc.

Pahl, G. and W. Beitz. (1996). Engineering Design: A Systematic Approach, Second Edition, Springer, London

Petroski, H. (1992). To engineer is human: The role of failure in successful design. New York: Vintage Books.

Project Lead the Way. (2014). About us. Retrieved October 29, 2015, from http://www.pltw.org/about/about-us.html

Roth, W. -M. (1996). Art and artifact of children's designing: A situated cognition perspective. Journal of the Learning Sciences, 5(2), 129-166.

Schauble, L., Klopfer, L. E., & Raghavan, K. (1991). Students' transition from an engineering model to a science model of experimentation. Journal of Research in Science Teaching, 28(9), 859–882.

Silk, E. M., Schunn, C. D., & Strand-Cary, M. (2009). The impact of an engineering design curriculum on science reasoning in an urban setting. Journal of Science Education and Technology, 18(3), 209–223.

Sadler, P. M., Coyle, H. P., & Schwartz, M. (2000). Engineering competitions in the middle school classroom: Key elements in developing effective design challenges. Journal of the Learning Sciences. 9(3), 299–327.

Ulrich, K. and S. Eppinger (2008). Product Design and Development, Fourth Edition, Boston: McGraw-Hill Higher Education

Van Lehn, K., S. Siler, & C. Murray. (2003). Why do only some events cause learning during human tutoring? Cognition and Instruction, 21(3), 209–249.

Watkins, J., K. Spencer, & D. Hammer. (2014). Examining Young Students' Problem Scoping in Engineering Design Examining Young Students ' Problem Scoping in Engineering Design, 4(1).

Wendell, K. (2013). Children's Design Constructions as Representations of Science Ideas. In Brizuela, B. M., & Gravel, B. E. (2013). (Eds.) Show me what you know: Exploring representations across STEM disciplines. New York: Teachers College Press.

Welch, M. (1999). Analyzing the Tacit Strategies of Novice Designers. *Research in Science & Technological Education*, *17*(1), 19–34.