OPENING THE BLACK BOX: INVESTIGATING THE IMPACT OF ENGINEERING DESIGN ON MECHANISTIC PROBLEM SOLVING AND MECHANISTIC UNDERSTANDING

A DISSERTATION
SUBMITTED TO THE GRADUATE SCHOOL OF EDUCATION
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Richard Lee Davis

August 2020

This dissertation is online at: http://purl.stanford.edu/ph667qp0822

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Roy Pea, Primary Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Paulo Blikstein, Co-Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Victor Lee

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Dan Schwartz

Approved for the Stanford University Committee on Graduate Studies.

Stacey F. Bent, Vice Provost for Graduate Education

This signature page was generated electronically upon submission of this dissertation in electronic format. An original signed hard copy of the signature page is on file in University Archives.

Abstract

The goal of this dissertation was to better understand how working on projects in makerspaces might affect the development of students' skills and knowledge. However, because makerspaces support an enormous variety of activities, it was necessary to restrict the focus to a specific type of activity and a specific set of outcomes. The activity was building mechanistic systems through the engineering design process, and the outcomes of interest were the development of mechanistic problem-solving skills and the construction of mechanistic understanding.

In the first study, high-school students took part in a year-long digital fabrication course in a makerspace. Over the course of the study, they worked on two long-term projects using the engineering design process. Both of the projects involved designing a single object that consisted of many interacting parts, and completing the projects required progressive refinement over the course of multiple iterations of the engineering design cycle. In the second project in the course, the students worked together to design and build a mechanistic system, the Rube Goldberg machine. After taking part in the course, the students received significantly higher scores on a set a hands-on problems involving mechanisms, and a fine-grained analysis of their problem-solving approaches indicated that they had become significantly more like expert mechanical engineers. This study suggested that designing and developing complex systems over multiple iterations of the engineering design process supported the development of mechanistic problem-solving skills in the high-school students.

The second study was designed to learn more about how building a mechanistic system through the engineering design process might support the construction of mechanistic knowledge. This study used a 2x2 factorial design to understand the impact of different aspects of the engineering design process—problem-centered making and reflection—on specific dimensions of mechanistic understanding. Students in this study were tasked with building a LEGO pendulum clock, and they all had access to a webpage containing text, diagrams, and videos about how pendulum clocks worked. The two factors in this study were problem-centered making [building the clock with step-by-step instructions vs. having to figure out how to build the clock using information on the webpage] and reflection [self-explaining the text on the webpage vs. not self-explaining]. Crossing these factors resulted in four groups: Read, Build, Explain, and Build+Explain. By comparing these groups on a set of posttest questions designed to measure different dimensions of mechanistic understanding, it was possible to learn more about how each of these activities supported learning about the pendulum-clock mechanism.

The first finding was that problem-centered making and self-explaining each supported learning about distinct dimensions of mechanistic knowledge. The students who self-explained learned more about what the purposes of the parts in the clock, and scored higher on a set of questions designed to assess declarative knowledge about the clock. The students who worked on the building problem (as opposed to the students who simply rebuilt the clock using instructions) learned more about the structural organization of the parts in the clock. However, neither of these activities was superior at supporting the construction of causal knowledge about the parts in the clock. The second finding was that the students in the Build+Explain group who engaged in both activities learned significantly more about the causal relations between components in the clock than any of the other groups. Additionally, this group of students were significantly more likely to transfer their knowledge to infer how an analogous, novel mechanism worked. These findings suggested that the combination of problem-centered making and reflection was most effective at supporting the construction of complete, transferable knowledge about a mechanism.

Together, the studies in this dissertation suggest that building mechanisms through a problem-centered, reflective process may support the development of mechanistic problem-solving skills as well as the construction of mechanistic understanding. This work has implications for researchers who are interested in how people learn about mechanisms, and for practitioners who are interested in supporting students in learning about how things work. Future work is needed to better understand the mechanisms through which skill and knowledge develop during the process of problem-centered, reflective making. These findings point to future directions for research, including the investigation of educational engineering design processes and the use of virtual, problem-centered, reflective making activities to support learning about a larger number of mechanistic systems.

Acknowledgments

I would like to thank my advisor, Paulo Blikstein, for all of his support, guidance, and encouragement throughout the PhD, particularly in the final stretch. He has provided me with innumerable opportunities to develop as a scholar, while also giving me the space and autonomy to find my own voice. I am grateful for his unwavering trust and confidence in me through the ups and downs of the PhD.

I owe a debt of gratitude to my dissertation advisor, Roy Pea, for his generosity and feedback. I have learned a great deal from him over the course of my PhD, including how important it is to seek out perspectives from other disciplines and historical eras in order to better understand the complexity of learning.

I would also like to thank Dan Schwartz and Victor Lee. From the beginning of my PhD, Dan has consistently pushed me to think harder about my work. His insights, critiques, and suggestions have dramatically improved my research, and he has had a lasting impact on my thinking. Victor has been a positive and supportive force throughout the dissertation, and I greatly appreciate his encouragement during the writing process. His feedback and support for this work has been invaluable.

I am grateful for all my colleagues and collaborators at Stanford and in the TLTL, including Bertrand Schneider, Chris Proctor, Veronica Lin, Allison Okamura, and Melisa Orta. I am especially grateful for Engin Bumbacher and Chris Piech, who have been incredibly generous with their time and support. Meeting them and working alongside them has been one of the most positive things that happened during my PhD, and I am a better researcher and person for knowing them.

My friends and family have been a source of strength and joy throughout the PhD. They have helped carry me through the difficult times, and have celebrated with me during the good times. I owe special thanks to my parents, Lee and Claire, and my siblings Nathan and Katie. You have all helped me get to this point, and I am so grateful for your love and support.

It is impossible to put into work how grateful I am for my wife, K-Fai Steele. She has been the most understanding, compassionate, and loving partner that I could ever hope for. She has been a consistent source of happiness and light, and her support over the past seven years is the main reason I have made it to the end of the PhD.

Finally, thanks to Derek Ang, Carolyn Ybarra, all of the students who took part in the studies in this dissertation, the GSE staff, and the administrative staff behind the Stanford Interdisciplinary Graduate Fellowship. There are so many hands that put in invisible labor so that PhD students can write and thrive and meet milestones; thank you for your work so that I, in turn, could write and complete my PhD.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Over the past 15 years, there has been a dramatic rise in interest and excitement about the potential of learning by making things in education. A number of organizations have emerged with the goal of promoting making in education, offering curricula, training, and support for educators who are interested in bringing making into their schools, libraries, and homes. These organizations have received funding and support from a broad coalition of corporations, foundations, governmental agencies, and philanthropists, representing the broad appeal and enthusiasm for this endeavor ("About the Fab Foundation", n.d.; "FabLearn Supporters", n.d.; "Maker Ed's Sponsors", n.d.). This development has helped fuel the rise of makerspaces in education, and today there are hundreds, if not thousands, of makerspaces in educational institutions across the United States (Melo & Rabkin, 2019). The result is that millions of students across the United States now have the opportunity to work in makerspaces over the course of their K-12 education (Peppler et al., 2015).

In recent years, researchers have begun to uncover a number of benefits for students working in makerspaces. Taking part in maker programs and working on maker activities have been found to positively impact students' attitudes towards technology and science (Kafai, Lee, et al., 2014; Kafai et al., 2013) and to improve their self-efficacy towards programming and building circuits (Qiu et al., 2013). Students working on maker projects that involve electrical circuits and computer programming have demonstrated improved understanding of electrical and computational concepts (Litts, Kafai, Lui, et al., 2017; Peppler & Glosson, 2013). And students who work in makerspaces report feeling more confidence about their ability to solve problems, attributing this to their experiences working on open-ended problems of their own design (Galaleldin et al., 2016; Harnett et al., 2014; Hartry et al., 2018).

The goal of this work, broadly speaking, has been to search for and identify educational benefits of working in makerspaces. This work has taken an invaluable first step down the path towards improving our understanding of the educational benefits of making in education. However, there is still much to learn about how these changes occur. This is because the prevailing approach to studying educational makerspaces has been to treat them as educational black boxes. Students go into a makerspace with one set of skills, attitudes, and knowledge; and students come out with a different set of attitudes, skills, and knowledge. The causes of these changes remain largely untested and unverified.

This would be an acceptable state of affairs if there were no variations in the activities that take place in makerspaces, since this would imply that whatever experiences a student had in one makerspace would be similar to the experiences they would encounter in another. However, this is not the case; makerspaces support an enormous variety of activities and forms of social interaction, which is reflected in the literature on making. Some groups of students work on projects with specific, technical requirements (Litts, Kafai, Lui, et al., 2017; Worsley & Blikstein, 2014), while others work on open-ended projects of their own design (Blikstein, 2013). Some students work in pre-assigned groups (Buchholz et al., 2014), while other students work alone or in groups of their own choosing (Davis et al., 2013; Sheridan et al., 2014). Because makerspaces are capable of supporting so many types of activities and forms of social interaction it is imperative to stop treating them like black boxes, and to begin searching for principles that describe how specific aspects of maker activities are related to outcomes of interest.

Research that uncovers these principles will not only provide a better theoretical understanding of the relationships between making and learning, it will also result in knowledge that is more useful for educators and educational designers. To see why this is the case, let's first look at the usefulness of research that takes a black-box approach. The main thing that can be extrapolated from research that treats making as a black box is that if a similar group of people were to work on the same activities in the same way, they would be likely to experience the same benefits. In some studies, these activities are not described at all, which severely limits the usefulness of the findings. In other studies where the activities are described in detail, the findings have some use, since they describe a link between a particular activity and a specific outcome. As long as it is possible to carry out the same activity in the same way, it should be possible to achieve similar outcomes.

However, research that looks inside the black box is capable of uncovering more-general principles that apply to a wide variety of activities, not just the specific interventions used in the studies. To date, there has been little research on making in education that has taken this approach. However, there are a handful of exceptions. For example, Schneider et al. (2015) found that following step-by-step instructions to build a working model of the human hearing system resulted in lower learning gains than students who had to figure out how to build the system on their own, suggesting that if the goal is to learn about something through building it, step-by-step instructions should be avoided. Another example is provided by Worsley and Blikstein (2014), who found that asking students to generate principles before working on a design problem resulted in a higher quality of design than asking them to solve general examples, suggesting that students should be encouraged to use principle-based reasoning during the design process in makerspaces. What makes these findings useful is their generality; they can be used to improve learning outcomes for a wide variety of maker activities.

The goal of this dissertation is to open the black box and more closely examine learning that occurs through making. The studies in this dissertation focus on a specific activity called engineering design. Engineering design has been defined as "an iterative process used to identify problems and develop and improve solutions" ("What Is Engineering Design?", n.d.). The first study explores how working on long-term engineering design projects in a makerspace affected students' problem-solving skills. The second study is focused on how different aspects of the engineering design process—problem-solving and reflection—each contributed to learning about mechanisms. Summaries of each of these studies are provided in the sections below.

1.1 Study 1: Building Mechanistic Problem-Solving Skills through Engineering Design

The first study, described in Chapter 3, was designed to examine how working through multiple cycles of the engineering design process affected high-school students' mechanistic problem-solving skills. A group of high-school seniors took part in a year-long course on digital fabrication where they worked on two long-term engineering-design projects. Before and after taking part in the course the students worked on a set of hands-on, mechanistic problems. A group of expert mechanical engineers also worked on the same set of problems. Problem-solving skill was measured in two ways. First, each participant was given a numeric score that represented how close they got to the solution. Second, each action or state visited as the participant worked through the problem was coded, which produced a temporal sequence of codes that represented the problem-solving process. The main finding in this study was that participation in the course had a positive effect on the high-school students' mechanistic problem-solving abilities. By comparing the expert mechanical engineers to the high-school students, it was possible to show that participation in the year-long course made the high-school students more like experts. Furthermore, by examining the differences in problem-solving processes between the pre-course high-school students, the post-course high-school students, and the experts, it was possible to hone in on the nature of this change: Experts and post-course students were more likely to adopt a mechanistic approach than pre-course students on both hands-on tasks. These results indicated that working on long-term engineering design projects during the year-long course positively impacted the high-school students' mechanistic problem-solving skills.

1.2 Study 2: Engineering Design and Mechanistic Understanding

The second study, described in Chapter 4, was designed to investigate the ways in which engaging in the engineering design process supports learning about mechanisms. This study focused on two aspects of the engineering design process—problem-solving and reflection—and attempted to tease apart the impact that each of these activities had on specific dimensions of mechanistic understanding. All of the students in this study had the same learning objective (to learn about how a pendulum clock worked) and worked with the same materials (a webpage containing information about pendulum clocks and a partially-disassembled LEGO model of the clock). However, the activities that students engaged in during the study differed between conditions, which were determined using a 2x2 factorial design. The first factor compared a problem-centered making activity (figuring out how to reassemble a pendulum clock using information from a webpage) to a making activity with no problem (assembling a pendulum clock with step-by-step instructions). The second factor compared self-explaining the text on the webpage to reading it twice. Crossing these factors resulted in four different treatments, and students in the study were randomly assigned into one of four groups that received one of these treatments (Figure 1.1).

Mechanistic understanding is theorized to involve four related dimensions of knowledge: (a) a phenomenon or phenomena that can be explained or understood by (b) decomposing it into parts or components that (c) are organized in such a way that (d) they give rise to the phenomenon through their causal interactions (Illari & Williamson, 2012). The posttest

		Build	
		With Instructions	With Information
Self-Explain	No	Read Twice	Build
	Yes	Explain	Build+Explain

Figure 1.1: 2x2 factorial design used in Study 2.

questions were designed to measure each of these four dimensions of knowledge about the pendulum-clock mechanism separately, and to measure the students' ability to transfer their knowledge to infer how a novel, analogous mechanism worked. The first finding was that problem-solving and self-explaining each supported learning about different dimensions of mechanistic knowledge. Problem-solving was effective at supporting students in learning about the structural organization of the parts within the clock, and self-explaining was found to support learning about the purposes of the parts in the clock (e.g., that the weight was the source of power). Additionally, self-explaining was more effective at supporting the construction of declarative knowledge about the pendulum clock.

The second finding was that the group who received both treatments—first working on the building problem, then self-explaining the text on the webpage—learned significantly more about the direct, causal relations between the parts in the clock than any of the other groups. This group also received dramatically higher scores on the transfer question than any of the other groups. This suggests that the ability to transfer knowledge between mechanisms—an important marker of expertise in science and engineering—requires full mechanistic understanding of the parts, purpose, organization, and causal relations. Additionally, the findings suggest that the ideal engineering design process, which involves both problem-solving and reflection, supports the construction of mechanistic understanding.

Chapter 2

Background

2.1 The History of Making in Education

2.1.1 Making in The Progressive Education Movement

Throughout the past 150 years, there have been numerous movements to bring making into schools in the United States. From the end of the 19th century to the first half of the 20th, these endeavors were a central part of the progressive education movement. This movement was united in its rejection of educational formalism, with its screwed-down seats and rote memorization, and sought to radically transform school by re-centering the educational process on experiences and activities, rather than lectures and recitations.

Two of the earliest educators in this tradition to integrate making into education were John D. Runkle, President of the Massachusetts Institute of Technology, and Calvin M. Woodward, a faculty member of Washington University. Both were proponents of manual training, which was the idea that a full education required both mental and manual activities. In 1879, Woodward launched a three-year secondary school called the Manual Training School of Washington University where traditional disciplines such as mathematics, science, and literature were combined with manual trades such as carpentry and metal work. Woodward conceived of this as a liberal education, not a vocational one, writing "[The system I advocate] aims to elevate, to dignify, to liberalize, all the essential elements of society" (Woodward, 1887, p. 239).

The manual training program soon spread to schools across the country. Some examples include the Gary schools, redesigned by superintendent William Wirt from the ground-up in 1907, which contained gardens, machine shops, and laboratories; the Play School, established by Caroline Pratt in 1914, which viewed children as artists with their own unique perceptions of reality, and supplied them with myriad materials that they could use to express themselves; and the Laboratory School, where students famously worked on carpentry, gardening, and sewing. These schools were all dedicated to the principle of learning by doing, and viewed doing and making as an essential part of a complete education (Dewey, 1901).

In the early 1950s, the progressive movement came under attack from all corners. A number of books critical of progressive education were published during this time period with titles like *Crisis in Education*, *And Madly Teach*, *Educational Wastelands*, *Quackery in the Public Schools*, and *The Diminished Mind*. Some blamed the progressive movement for the moral decay of American youth, caused by the focus on educating the whole child which over-reached into affairs better taught in the home or the church (Bell, 1949). The remedy offered by Bell for this malady was to incorporate religious education into public schooling. Others argued that progressive education had committed educational malpractice for failing to properly teach the basic academic disciplines (i.e., the three R's) (Bestor, 1955). Again, this failure was blamed on the over-reach of progressive educators, and the solution offered was to strip down school to the essential academic disciplines. These attacks were not always coherent or consistent, which indicated they were symptoms of deeper cultural causes. Cremin argued that a more general swing towards conservatism in American society was one of the root causes of this discontent.

Then, in 1957, Russia launched Sputnik, making them the first nation to put an artificial satellite into orbit, and inadvertently dealing a death-blow to progressive education in the United States. To the American public, Sputnik symbolized the scientific and technological superiority of the Soviet Union over the United States, and the progressive education movement was blamed for failing to produce mathematicians and scientists of equal quality to the USSR. Sputnik was the main catalyst for the National Defense Education Act (NDEA), a large federal investment in science and mathematics education that reorganized the curriculum around science and mathematics education. This re-imagining of public school curriculum as centered on science, technology, engineering, and mathematics (STEM) had long-term effects on formal schooling that continue to this day (Wissehr et al., 2011). While this had the positive effect of dramatically improving the quality and scope of STEM education in K-12 education, it also had the effect of displacing progressive ideas and pushing hands-on, creative activities like making to the periphery of the school.

2.1.2 Logo and Constructionism

Making did not remain out of schools for long. When it first re-emerged in the mid-1960s, it was in a form that was more-closely aligned with the STEM-focused educational mission of the day: computer programming. In 1965, Wally Feurzeig formed an Educational Technology Department at Bolt, Beranek, and Newman, Inc. (BBN) with the express purpose of researching programming languages as educational environments. Soon after he was joined by Seymour Papert, and together they laid the groundwork for Logo, the first programming language designed for children. It was initially designed with three basic requirements:

1. Third-graders with very little preparation should be able to use it for simple tasks.
2. Its structure should embody mathematically important concepts with minimal interference from programming conventions.
3. It should permit the expression of mathematically rich non-numerical algorithms, as well as numerical ones.

(Feurzeig, 2006, p. 26)

Together with Daniel Bobrow and Cynthia Solomon, Papert and Feurzeig implemented the first working version of Logo in 1966, which was a language for playing with words and sentences.

The first trial of Logo took place in a classroom in Lexington, Massachusetts, where seventh-grade students learned and worked with Logo as part of a year-long computer math course. This course, taught by Solomon and Papert, set out to teach mathematics by having students work on a series of programming projects such as converting text into Pig Latin, teaching math, and telling stories. The next iteration of Logo, inspired by the robotic tortoises of William Grey Walter, included the ability to control a “turtle”. The original Logo turtle was robotic, and students could write programs that controlled where and how they moved on the floor. Each of the turtles contained a pen that could be raised and lowered on command, making it possible to create drawings by writing programs. Soon after, Logo incorporated virtual turtles, which were represented by small triangles on a computer screen and could also be programmed to move around and make drawings.

Over the following decades, Papert developed a theory of learning called constructionism that was based on his observations and experiences with students using Logo. An early version of this theory appeared in *Mindstorms* (Papert, 1980). Though the word “constructionism” did not appear in the book, and there was little discussion about the relationship between making and learning (a connection that became central in later writings), this work contained a number of concepts and claims central to future constructionist research and writing. Papert provided his own summary of the central ideas in the book:

> Two fundamental ideas run through this book. The first is that it is possible to design computers so that learning to communicate with them can be a natural process, more like learning French by living in France than like trying to learn it through the unnatural process of American foreign-language instruction in classrooms. Second, learning to communicate with a computer may change the way other learning takes place… We are learning how to make computers with

> which children love to communicate. When this communication occurs, children learn mathematics as a living language... The idea of "talking mathematics" to a computer can be generalized to a view of learning mathematics in "Mathland"; that is to say, in a context which is to learning mathematics what living in France is to learning French." (Papert, 1980, p. 6)

Papert used the term "microworld" to describe these sorts of computational learning environments. Logo was presented as the first example of a microworld, offering a proof of concept that heralded the dawn of a new educational future. Papert wrote:

> The computer presence will enable us to so modify the learning environment outside the classrooms that much if not all the knowledge schools presently try to teach with such pain and expense and such limited success will be learned, as the child learns to talk, painlessly, successfully, and without organized instruction. This obviously implies that schools as we know them today will have no place in the future. But it is an open question whether they will adapt by transforming themselves into something new or wither away and be replaced. (Papert, 1980, p. 9)

Interest in Logo microworlds exploded after the publication of *Mindstorms*. In the decades that followed, a number of Logo-based programming environments were developed to allow students to learn other topics "painlessly, successfully, and without organized instruction." Microworlds based on Logo were developed to allow the exploration of a number of domains, such as physics (DiSessa, 2001), music (Bamberger & Schön, 1983), and emergent systems (Wilensky & Resnick, 1999).

Papert was optimistic about the transformative power of learning with microworlds. They would make it possible for students to "learn to transfer habits of exploration from their personal lives to the formal domain of scientific theory construction" (p. 117), and they would allow students to "become the active, constructing architects of their own learning" (p. 122). In the process of developing computer programs, students would "reflect on [their] own actions and thinking... [making] more complex decisions and finding themselves engaged in reflecting on more complex aspects of their own thinking" (p. 28).

These claims rightfully attracted interest from educational researchers. Some of the earliest controlled studies on Logo were conducted by Pea and colleagues in the 1980s. They used a series of experiments to investigate changes in students' knowledge of programming (Pea, 1987) as well as changes in their planning skills (Pea et al., 1985). Students in these studies spent long periods of time working on self-directed programming projects "under the supervision of experienced classroom teachers knowledgeable in the Logo language, who followed the 'discovery' Logo pedagogy set out by Papert" (Pea et al., 1985, p. 238). Surprisingly, these studies found little evidence that working with and learning Logo in self-directed, project-based classrooms had cognitive or metacognitive benefits for students.

The conclusion that Pea and colleagues came to was not that learning to program in Logo had no educational value, but that learning to program required significant instructional support that was not provided in the Logo instructional environment. Pea and Kurland wrote:

> Learning thinking skills and how to plan well is not intrinsically guaranteed by the LOGO programming environment; it must be supported by teachers who, tacitly or explicitly, know how to foster the development of such skills through ajudicious use of examples, student projects, and direct instruction. But the LOGO instructional environment that Papert (1980) currently offers to educators is devoid of curriculum, and lacks an account of how the technology can be used as a tool to stimulate students' thinking about such powerful ideas as planning and problem decomposition. Teachers are told not to teach, but are not told what to substitute for teaching. (Pea & Kurland, 1983, p. 212)

This conclusion directly challenged Papert's theories about learning in microworlds, which he described as "a process that takes place without deliberate and organized teaching" (Papert, 1980, p. 8).

Subsequent work supported the conclusion that instructional support was key to unlocking the educational benefits of learning Logo. For example, Clements and Gullo (1984) found evidence that students who learned Logo scored higher on measures of reflectivity, two measures of divergent thinking, one measure of metacognitive ability, and a measure of the ability to describe directions than a control group. Students in this study received instructional support in the form of a sequenced, 24-lesson curriculum, and researchers fluent in Logo individually led groups of two to three students through this curriculum. Another study by Clements and Battista (1989) found that students who learned Logo scored significantly higher than a control group on a suite of questions about angles (e.g., recognition of right angles) as well as a suite of questions about shape and motion (e.g., charting a path through a map). The instructional support provided in this study was a 78-hour curriculum on geometry that involved learning Logo. In addition to Logo programming, the teachers in this study also engaged students in exercises and discussions about geometric concepts such as line segment, rotation, angle, right angle, congruence, and similarity. This additional support was provided because "It was not expected that Logo programming alone would produce changes in children's geometric schemata" (Clements & Battista, 1989, p. 452). These and other studies demonstrated that when appropriate forms of instructional support were provided, learning Logo provided a number of cognitive and metacognitive benefits[1].

Sometime after these issues were raised, the theory of constructionism shifted away from a focus on Logo-based microworlds and towards a more-general focus on making and maker activities, significantly widening the theory's scope. Papert wrote:

> The word constructionism is a mnemonic for two aspects of the theory of science education underlying this project. From constructivist theories of psychology we take a view of learning as a reconstruction rather than as a transmission of knowledge. Then we extend the idea of manipulative materials to the idea that learning is most effective when part of an activity the learner experiences as constructing a meaningful product. (Papert, 1986).

Constructionism held that learning was most likely to happen "in a context where the learner

[1] Today, the importance of teacher preparation is recognized by the constructionist community. In a history of Logo published in 2020, Cynthia Solomon and colleagues wrote, "If there are to be consistent positive outcomes in educational settings, significant investment in teacher preparation is sure to be crucial. The central importance of teacher preparation remains a crucial lesson today, as schools attempt to bring computer science into the classroom" (Solomon et al., 2020, p. 79:52).

is consciously engaged in constructing a public entity, whether it's a sand castle on the beach or a theory of the universe" (Papert & Harel, 1991). In this form, constructionism arguably became the first theory of learning to explicitly center the learning process on making, in contrast to theories that included making as part of a broader class of experiences held to be at the heart of learning. The original progressive spirit, expelled from schools in the wake of Sputnik, had returned in full force under the guise of constructionism.

This broadened focus is reflected in the work that has been carried forward by a new generation of constructionist researchers that is largely comprised of Papert's former students and postdocs. This group has taken advantage of the dramatic reduction in size and cost of computing technology to develop a broad array of programming languages and computational construction kits. Some examples of these include the Scratch programming language (Resnick et al., 2009), the LilyPad Arduino electronic-textiles construction kit (Buechley, 2006), and the GoGo Board robotics kit (Sipitakiat et al., 2002). The primary thing that all of these technologies and kits have in common is that they make it easier for novices to create computational technologies that would otherwise require too much prior knowledge and technological expertise. Resnick and Silverman described these technologies as ideally having low floors, high ceilings, and wide walls; meaning these tools have few barriers to entry, but are powerful enough to allow for the construction of a wide range of sophisticated artifacts (Resnick & Silverman, 2005).

2.1.3 Digital Fabrication and The Rise of the Maker Movement

Another class of technology that has come to be associated with constructionism is digital fabrication. Digital fabrication has been called "the third industrial revolution" (Anderson, 2012; "The Third Industrial Revolution", 2012; Troxler, 2013). In the first industrial revolution, manufacturing was transformed by the introduction of efficient steam engines which made possible new inventions such as the power loom and the cotton gin. In the second industrial revolution, electric power transformed manufacturing, giving rise to new technologies such as the assembly line and the telegraph. The computer has given rise to the third industrial revolution. Engineers can create, modify, analyze, and optimize their designs using computer-aided design (CAD) software, and the final, digitized designs are sent to computer numerical control (CNC) machines for production. This new type of computer-enabled manufacturing is known as digital fabrication.

The same phenomenon that made it possible for constructionist researchers to make construction kits by embedding computers into everyday objects (Moore's Law) also directly contributed to a dramatic reduction in price and size of the tools used in digital fabrication, making them more accessible to hobbyists and amateurs. Open-source and low-cost CAD software has made it possible for individuals to create, test, and refine digital designs, and low-cost and easy-to-use hardware such as 3D printers, laser cutters, milling machines, vinyl cutters, and microcontrollers have made it possible to rapidly explore, produce, and iterate on creative artifacts, objects, and inventions.

In the early 2000s, a group of constructionist researchers began to explore the educational potential of digital fabrication. The first of these explorations was called Learn2Teach: Teach2Learn (2004-2006), which brought students in a community center fablab in Massachusetts (Millner & Daily, 2008). Soon after, the first organized initiative to bring fablabs into formal K-12 educational settings around the world, FabLab@School, was launched in 2008 by Paulo Blikstein (Blikstein & Krannich, 2013). Blikstein explicitly framed fablabs as constructionist learning environments, and introduced technologies such as computational construction kits to the collection of digital-fabrication tools such as 3D printers and laser cutters. He wrote "What Logo did for geometry and programming—bring complex mathematics within the reach of school children——fabrication labs can do for design and engineering. Digital fabrication is Logo for atoms" (Blikstein, 2013, p. 3). The design of Blikstein's educational fablab, with its combination of construction kits, digital-fabrication tools, and low-tech tools and materials, provided the template for the majority of educational makerspaces and fablabs that followed.

Around the same time, the maker movement was born. As high-quality, professional-grade manufacturing technologies became more accessible, a grassroots culture of hobbyists interested in digital fabrication began to take shape. In 2005, this culture emerged into the mainstream with the publication of Make magazine, and the members of this group of like-minded individuals took on a common identity as “makers”. A year later, Make:Magazine organized the first gathering of makers at Bay Area Maker Faire. Make:Media described the Maker Faire as “the Greatest Show (and Tell) on Earth—a family-friendly festival of invention, creativity and resourcefulness, and a celebration of the maker movement” (“About Maker Faire”, n.d.). These gatherings provided makers with an opportunity to meet one another; to showcase and share projects; and to learn about novel tools, techniques, and ideas.

The maker movement started strong and grew rapidly. The first Maker Faire in 2006 attracted over 20,000 people (“Looking Back at the Launch of the Maker Movement”, 2015), and less than a decade later over 200,000 people were attending the gatherings. As making grew into a cultural phenomenon, there was a growing interest in engaging students in making. In 2012, Make:Magazine launched the Maker Education Initiative built on the idea that “every child is a maker” with the goal of developing “places and programs that provide hands-on learning experiences for children and teens”. This initiative was initially supported with funding from industry (Intel, O’Reilly Media, and Cognizant), whose support was intended to “usher in the next wave of economic growth for our country” and to improve “21st century skills such as critical thinking, collaboration, creativity, and innovation” (“Maker Education Initiative Launches at Maker Faire”, n.d.). That same year, Maker Education received a large federal grant from DARPA to install 1000 makerspaces in schools over the next five years (“O’Reilly’s Make and Otherlab Win DARPA MENTOR Award to Bring Making to Education”, n.d.).

The Maker Education Initiative’s goals were to transform the educational experience for every child through making, and “to shift learning to a balanced eco-system that provides multiple pathways for learners to develop their own agency and problem-solving dispositions, and to collaborate and learn from each other" ("About Us", n.d.). Making was said to have a wide array of benefits for students:

> When students are engaged in making, they embrace creativity, innovation, and discovery, and have meaningful opportunities to collaborate, solve problems, and imagine a vibrant future... When students and adults are collaborating to build the future together, there is a ripple effect of excitement that engages parents, shifts the culture of learning, and builds more connected communities. ("About Us", n.d.)

Additionally, making was framed as a way to bring joy back into an overly-formalized school system:

> We must try to bring the youthful magic of play into schools, hard as it may be. Formal education has become such a serious business, defining success with abstract thinking and high-stakes testing, that there is no time and no context for play. "Makerspace Playbook", p. 3

Today, there are hundreds, if not thousands, of makerspaces in educational institutions in the United States (Melo & Rabkin, 2019), and millions of students across the United States now have the opportunity to work in makerspaces over the course of their K-12 education (Peppler et al., 2015).

2.1.4 The Need for Research on Making and Learning

The rise of the maker movement fits a familiar pattern within the history of education: Whenever making comes to school, it is ushered in with unbounded optimism, enthusiastic support from industry, promises of transforming an overly-formalized school system by restoring joy and creativity, promises that learning will be unlocked by restoring a balance of manual and intellectual activity, and promises that students will become more curious, motivated, and improve at skills such as problem solving and creative thinking.

Historically, there has been little effort to test these claims, perhaps because proponents of making have held such strong beliefs about its effectiveness. Cremin describes the "conversion experience" of Caroline Pratt, who created the Play School in 1917:

> Miss Pratt describes the situation as follows: "[The child] was building with blocks, toys, odd paper boxes, and any other material he could find… It seemed to me that this child had discovered an activity far more satisfying to him than anything I had ever seen offered to children." (Cremin, 1961, p. 203-204)

Almost 100 years later, Gabrielson echoed an almost identical sentiment about the maker movement in education:

> Heck yes they're learning something, and it may be the most valuable thing they've learned all week, and it may raise all sorts of questions in their minds that inspire them to learn more about what they're tinkering with, and it may start them on a path to a satisfying career. (Gabrielson, 2013, p. xi)

He goes on to explicitly argue that it is so intuitively obvious that children learn by making that research is unnecessary:

> Now, you can get a PhD trying to show, incontrovertibly, that learning is happening in a tinkering environment, or attempting to work out exactly how it is happening. I'll certainly not stand in your way… [But] I'm comfortable with my gut instinct. (Gabrielson, 2013, p. xii)

However clear and distinct these truths may be to proponents of making, they can not be accepted as any sort of meaningful evidence for the educational impact that making has on students. The lack of scientific evidence for the effectiveness of making may be one reason that its position in mainstream education has been somewhat precarious. There are deeply-ingrained cultural beliefs that view making and other hands-on activities as being "lower" than other intellectual activities, in that they are incapable of producing abstract knowledge. Dewey argued in *Democracy and Education* that this viewpoint could be traced back to the original distinction made by Aristotle between *episteme* (abstract, scientific knowledge) and *techne* (practical knowledge, knowhow):

> Much as [the Greeks] differed in many respects, they agreed in identifying experience with purely practical concerns; and hence with material interests as to its purpose and with the body as to its organ. Knowledge, on the other hand, existed for its own sake free from practical reference, and found its source and organ in a purely immaterial mind; it had to do with spiritual or ideal interests. (Dewey, 1916/2004, p. 270)

There is historical evidence that this sentiment has played a role in preventing making from occupying a stable position in educational practice. For example, the National Education Association (NEA) originally argued that the manual training program was flawed because "it failed to differentiate between higher and lower forms of knowledge" (Cremin, 1961, p. 31). Later, in a criticism of the entire progressive education movement, Bestor wrote that education had become detached from its true purpose, which was "the deliberate cultivation of the ability to think", and argued that the way forward was the restoration of traditional academic disciplines. What these criticisms demonstrate is that the burden of proof is on proponents of making in education to demonstrate its educational value. Otherwise, making is likely to continue to be an occasional guest in schools, but never a permanent occupant.

There is a need for more research on the relation between making and learning. Some of the questions whose answers are largely unknown include:

- What are the types of things that people learn when they make things? What forms of knowledge do they construct?
- What are the boundary conditions of making to learn? That is, what are the forms of knowledge that people are *unlikely* to learn from making?
- How can making activities be designed and scaffolded so that people learn? Are there specific events that must occur during making for learning to occur?
- What forms of instructional support can help ensure that learning occurs when students make things?

- How does making compare to other pedagogical methods? Are there other methods that offer the same educational benefits, or is making an activity that builds a specific type of knowledge?

Without the answers to these questions, the role that making can play in education is unclear, leaving educators to rely on their own habits, experience, and intuition when designing or choosing making activities for their students. However, over the past decade research that attempts to answer these (and similar) questions has begun to appear. A summary of this work is provided in the next section.

2.2 A Review of Research on Making in Education

We are entering the second decade of research on digital fabrication and making in education. Over the past decade, there have been three main areas of research on the educational benefits of making: changes in attitudes and perceptions, changes in knowledge, and changes in skills.

The effects of making on attitudes and perceptions have been found to be fairly consistent: Taking part in courses and workshops involving making and digital fabrication positively influences students' attitudes and intentions around science and technology. Based on review of 43 empirical studies selected as "rigorous, credible, and relevant", Papavlasopoulou et al. reported that "no matter what the age of the group and which tool was used, making proved to be a successful process in all the different areas of interest... Almost none of the studies reported negative effects in the research" (2017, p. 60). Outcomes that were present across multiple studies included changes in perspectives towards technology and computing (Kafai, Lee, et al., 2014; Kafai et al., 2013), changes in self-efficacy towards building and making sense of technology (Chu et al., 2015; Qiu et al., 2013), and changes in attitudes towards technology, making, and science class (Garneli et al., 2013; Harnett et al., 2014; Searle et al., 2014; Togou et al., 2020).

Research in the second category of learning in makerspaces has mostly focused on technical concepts such as programming, electricity, and electronics, though there is some research on other domains such as biology (Schneider et al., 2015), art-making (Patton & Knochel, 2017), and history (Blikstein, 2013) as well. On the whole, students in these studies learn by working on well-scaffolded projects designed to teach specific concepts. A number of qualitative and ethnographic studies have found evidence of positive changes in students' understanding of electricity and electronics (Kafai, Fields, et al., 2014; Kafai, Lee, et al., 2014; Searle et al., 2014; Telhan et al., 2014) as well as in their understanding of programming concepts (Davis et al., 2013; Franklin et al., 2013; Qiu et al., 2013). A smaller number of quantitative studies have found evidence for positive changes in students' understanding of electricity and programming (Litts, Kafai, Lui, et al., 2017; Peppler & Glosson, 2013), though small sample sizes are sometimes blamed for mixed results (Bers et al., 2014).

The third category of skill acquisition brings together research on a broad array of topics, including problem solving, collaboration, computational thinking, design thinking, and technical skills. When compared to the other two categories, the research on skill acquisition is less mature. While there is little debate over whether skills such as creativity, computational thinking, and problem solving are desirable; there is less agreement on how to assess or teach these skills. One method is to use self-reported assessments in interview and survey data. This has been the approach in most of the research on changes in problem-solving skills. Harnett et al. (2014) found that some university students who spent a semester working in a community hackerspace reported increased confidence in their problem-solving and project-planning abilities, and Galaleldin et al. (2016) reported that 60% of university engineering students reported feeling "more confident in their engineering knowledge and skills to solve a complex engineering problem". At the K-12 level, Hartry et al. (2018) found that K-12 students working as interns in a museum makerspace self-reported increases in problem solving skills, attributing this to their experience working on open-ended problems during their internship. An emerging methodology involves analyzing the artifacts students create, such as programs, projects, and project documentation. For example, Houchins et al. (2020) recently introduced a method for identifying computational thinking in students project documentation, and demonstrated its effectiveness by identifying changes in computational thinking practices that depend on the characteristics of projects and tools used. Bers et al. (2014) and Kafai and Vasudevan (2015) analyzed students' programs directly, using the number, type, and sequence of expressions used in a program as a proxy for computational thinking skills.

2.3 A Critique of Research on Making and Education

The majority of published work has taken a qualitative approach to understanding learning in fablabs and makerspaces. Papavlasopoulou et al. (2017) reported that 85% of the empirical studies in their literature review were either purely qualitative or utilized mixed methods. A separate literature review of 57 empirical studies conducted by Timotheou and Ioannou reported that 88% used either qualitative or mixed methods. Given the complexity and variety of activities that occur in digital fabrication labs, the use of qualitative and mixed methods is both necessary and justified. In many cases, these studies offer a proof-of-concept that students are able to understand and properly work with some newly-designed toolkit or successfully take part in a novel workshop design, and have highlighted a number of promising activities and outcomes of interest. However, there is still much to learn about the specific relationships between the interventions (e.g., toolkits, activities) and processes of learning and development.

This issue is at the heart of a critique that Pea made over 30 years ago when writing about Logo:

> For a developmentalist, there is a major problem pervading each of these characterizations of the effects on higher thinking skills expected from learning to program. Programming serves as a "black box," an unanalyzed activity, whose

> effects are presumed to irradiate those exposed to it... [But] they require especially serious consideration of the developmental roles played by the contexts interpenetrating the black box: the programming environment, the instructional environment, and the relevant understandings and performances of the learner. (Pea & Kurland, 1984, p. 144)

To date, much of the research on making is vulnerable to this criticism. Making is treated as a black box whose effects can be perceived, but whose inner workings remain untheorized and unanalyzed.

In the conclusion of their literature review on making, Timotheou and Ioannou suggest that a shift to other research approaches is warranted, writing "We now have enough evidence of the value of computational making, allowing for scaling-up the impact and measurement via quantitative studies" (2019, p. 227). However, a cursory review of the quantitative studies listed in Timotheou and Ioannou (2019) and Papavlasopoulou et al. (2017) demonstrates that simply shifting to quantitative research methods does not, by itself, address these criticisms. Out of the 43 studies reviewed in Papavlasopoulou et al. (2017), 12 were categorized as using quantitative methods. Of these 12, only 8 reported their results, and only 2 out of the 8 studies that reported their results performed statistical analyses of their data. The rest reported percentages, or provided written summaries of the results. With few exceptions, these studies treated making as a black box, and provided little insight into the mechanisms of learning by making. For example, Galaleldin et al. (2016) reported that 60% of students who worked in a makerspace reported feeling more confident in their ability to solve a complex problem using their engineering knowledge and skills. However, the authors reported no details about what these students actually did in the makerspace, and did not attempt to assess the changes in problem-solving skills. Without these details, there is no way of conducting research that builds upon these findings, nor is it possible to use them to improve students' experiences in other makerspaces.

Research that looks inside the black box is more useful because it is capable of uncovering general principles that link specific aspects of maker activities to specific outcomes of interest. To date, there is little research on making in education that has taken this approach. However, there are a handful of exceptions. Schneider et al. (2015) found that following step-by-step instructions to build a working model of the human hearing system resulted in lower learning gains than students who had to figure out how to build the system on their own, suggesting that if the goal is to learn about something through building it, step-by-step instructions should be avoided. Lane et al. (2013) found that enthusiastic support from a pedagogical agent increased students' self-efficacy more than an unenthusiastic one, suggesting a link between enthusiastic support and self-efficacy. Garneli et al. (2013) compared students who learned math concepts with a game to those who learned using a textbook, and found "the storytelling element in an educational game does not seem to affect the improvement of students' performance" (p. 82). Finally, Worsley and Blikstein (2014) found that asking students to generate principles before working on a design problem resulted in a higher quality of design than asking them to general examples, suggesting that students should be encouraged to use principle-based reasoning during the design process in makerspaces. These are useful findings because they are relevant to a wide variety of activities, not just the specific interventions used in the studies.

There is no point in switching from qualitative to quantitative methods, as suggested by Timotheou and Ioannou (2019), without also adopting a new perspective on research in maker education. The goal for the next decade of research should be to open up the black box, peer inside, and search for general principles that apply to many types of making and learning.

2.4 A Closer Look at Engineering Design

Since the goal in this dissertation is to explore the mechanisms through which students learn by making, there is a need to specify a testable theory that links making and learning. In both of these studies in this dissertation, I chose to focus on a specific type of activity called engineering design. Engineering design has been defined as “an iterative process used to identify problems and develop and improve solutions” (“What Is Engineering Design?”, n.d.). The problem-centered nature of engineering design suggests that it may be an effective method for the construction of knowledge about mechanisms. My focus on engineering design is the result of an analysis of the similarities between Piagetian constructivism and Deweyan pragmatism, and the focus on mechanisms is based on the similarities between Dewey’s theory of reflective inquiry and the iterative process of engineering design.

2.4.1 Engineering Design as a Form of Problem-Centered Making

Both Dewey and Piaget frame learning as a process with its genesis in a state of uncertainty or confusion. Dewey called this an indeterminate situation, and Piaget called this a state of disequilibrium. For both, the cause of this state of confusion is the inability to understand some aspect of the situation (i.e., a gap in knowledge). In many cases, this moment of confusion can be ignored, or else the situation will change and the moment will pass. However, in some cases a person will actively set out to locate the source of confusion and attempt to resolve it. When this occurs, the indeterminate situation transforms into a “problematic situation.” For Dewey, learning occurs through the process of reflective inquiry, and “All reflective inquiry starts from a ”problematic situation,“ and no such situation can be settled in its own terms” (Dewey, 1930, p. 181). This commits Dewey to an extended meaning of the word problem: “[a problem is] whatever—no matter how slight and commonplace in character—perplexes and challenges the mind so that it makes belief at all uncertain” (Dewey, 1910/1997, p. 9).

In a synthesis of Piaget's writings, Fosnot and Perry write

> These progressive experiences sometimes foster contradictions to our present understandings, making them insufficient, thus perturbing and disequilibrating the structure and causing accommodations to reconstitute efficient functioning. (Fosnot & Perry, 1996, p. 10)

In other words, a person sometimes encounters a situation that is confusing, and the reason it is confusing is because there is something about the situation that the person does not understand. This confusing state is one of disequilibrium, and there is a deeply felt need to restore equilibrium. In order to restore equilibrium and "restore functioning", learning (i.e., assimilation or accommodation) occurs: Fosnot continues:

> Disequilibrium facilitates learning. 'Errors' need to be perceived as a result of learners' conceptions, and therefore not minimized or avoided. Challenging, open-ended investigations in realistic, meaningful contexts need to be offered which allow learners to explore and generate many possibilities, both affirming and contradictory. Contradictions, in particular, need to be illuminated, explored, and discussed. (Fosnot & Perry, 1996, p. 27)

For both Dewey and Piaget, learning—the construction of knowledge—is a process through which equilibrium is restored. This does not mean that knowledge is a necessary result of encountering a confusing situation. For example, Chinn and Brewer (1993) identified seven ways that students respond to anomalous data, of which only three involve any sort of learning. However, both Piaget and Dewey argued that when learning does occur, it is the outcome of resolving a "problematic situation" and restoring equilibrium.

This conclusion implies that some forms of making—those that involve encountering and solving problems—are more likely to result in learning than others. There are many types of activities that are lumped together under the category of making in education. For example, a definition of making put forward by Martin in "The Promise of the Maker Movement for Education" was "a class of activities focused on designing, building, modifying, and/or repurposing material objects, for playful or useful ends, oriented toward making a 'product' of some sort that can be used, interacted with, or demonstrated" (2015, p. 31). This definition encompasses a wide range of activities, some of which are unlikely to bring students into "problematic situations." Some examples of making activities that are unlikely to involve problem-solving include downloading design files from the Internet and sending them to a 3D printer to be manufactured, following step-by-step instructions to assemble a device, and producing multiple copies of a design with small variations. Though it is possible to encounter problems while engaging in these activities, these problems are rarely rooted in lack of understanding about the object or system being assembled or manufactured, and solving these problems rarely involves learning about the workings of the object or system itself.

One type of making that is specifically focused on identifying and solving problems is engineering design. Engineering design describes the act of identifying and solving problems that arise during the process of turning an idea into a final product (Sheppard et al., 2006), and has been defined as "identifying the problem; specifying requirements of the solution; decomposing the system; generating a solution; testing the solution; sketching and visualizing the solution; modeling and analyzing the solution; evaluating alternative solutions, as necessary; and optimizing the final design" (*Engineering in K-12 Education*, 2009). Each iteration in the engineering design process provides students with a new opportunity to practice identifying problems and hypothesizing, testing, and evaluating solutions. The process begins with identifying a problem or objective, and then proceeds through the stages of hypothesizing a solution to the problem, implementing the possible solution in a prototype, testing the prototype to determine whether the solution was effective, identifying new problems, and repeating the process. The cycle begins anew each time a new problem is identified or the nature of an existing problem is clarified, and each iteration of the design cycle is centered on the problem that was surfaced in a previous cycle.

2.4.2 Engineering Design, Reflective Inquiry, and Learning about Mechanisms

There are a number of parallels between the engineering design process and Dewey's formulation of the reflective inquiry process. In fact, the engineering design process can be almost perfectly mapped onto the process of reflective inquiry (Figure 2.1). This suggests that Dewey's theory of reflective inquiry can be used to make predictions about the types of knowledge that students who engage in engineering design and other forms of problem-centered making will construct.

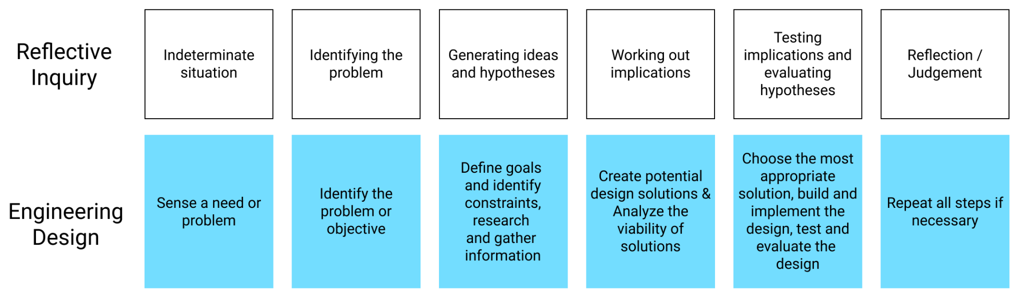


Figure 2.1: Parallels between the process of reflective inquiry (Dewey, 1910/1997) and the engineering design process.

Dewey described the reflective inquiry process as follows:

> "Upon examination, each instance reveals, more or less clearly, five logically distinct steps [in reflection]: (i) a felt difficulty (ii) its location and definition; (iii) suggestion of possible solution; (iv) development by reasoning of the bearings of the suggestion; (v) further observation and experiment leading to its acceptance or rejection; that is, the conclusion of belief or disbelief" (Dewey, 1910/1997, p. 72).

This described the process through which scientific knowledge, "the most perfect type of knowledge" (Dewey, 1910/1997, p. 134), was constructed. Dewey described scientific knowledge as being about the causal relations between things in the world. He wrote:

> This scientific definition is founded, not on directly perceived qualities nor on

> directly useful properties, but on the way in which certain things are causally related to other things; i.e. it denotes a relation. (Dewey, 1910/1997, p. 134)

For Dewey, "doing something" (i.e., directly interacting with the elements of the "problematic situation") was a requirement for the construction of scientific knowledge through reflective inquiry. He wrote:

> The most direct blow at the traditional separation of doing and knowing and at the traditional prestige of purely "intellectual" studies, however, has been given by the progress of experimental science. If this progress has demonstrated anything, it is that there is no such thing as genuine knowledge and fruitful understanding except as the offspring of doing. The analysis and rearrangement of facts which is indispensable to the growth of knowledge and power of explanation and right classification cannot be attained purely mentally—just inside the head. Men have to do something to the things when they wish to find out something; they have to alter conditions. This is the lesson of the laboratory method, and the lesson which all education has to learn. (Dewey, 1916/2004, p. 283)

The link between interaction and learning is a consequence of Dewey's theory of knowledge. For Dewey (and other pragmatists) all knowledge about the world is practical, in that it allows for prediction and control. This is, in fact, the pragmatic criterion for what it meant to know something at all (Peirce, 1878). This is the reason for Dewey's insistence that experimentation—trying things and observing the consequences to resolve a "problematic situation"—is necessary for learning. Dewey wrote that the "common character of all such scientific operations... [is that] they disclose relationships" (Dewey, 1930, p. 120-121). Discovery of these causal relations "[places] in our hands an instrument of control. When one change is given, and we know with measured accuracy its connection with another change, we have the potential means of producing or averting that other event" (Dewey, 1930, p. 98).

This conception of knowledge places natural phenomena on the same level as machines.

Natural phenomena and machines both "work" through the causal interaction of components, and knowledge of each means understanding these causal relations in such a way that prediction and control are possible. Dewey made this connection explicitly in *The Quest for Certainty*:

> Nature has mechanism. This mechanism forms the content of the objects of physical science, for it fulfills the instrumental office to be performed by knowledge. If the interactions and connections involved in natural occurrences were not sufficiently like one another, sufficiently constant and uniform, so that inference and prediction from one to another were possible, control and purpose would be non-existent. Since constant relations among changes are the subject-matter of scientific thought, that subject-matter is the mechanism of events. (1930, p. 236)

Under this conception of knowledge and learning, a wide array of activities and professions are placed on equal footing with respect to their relation to knowing. "Were we to define science not in the usual technical way, but as a knowledge that accrues when methods are employed which deal competently with problems that present themselves, the physician, engineer, artist, craftsman, lay claim to scientific knowing" (Dewey, 1930, p. 191). Types of making designed to bring students into "problematic situations" such as engineering design fit this description. The conclusion is that engineering design should support students in the construction of knowledge about the causal relations between components in the systems they are constructing. In other words, engineering design should facilitate the construction of mechanistic knowledge.

In recent years, the idea that scientific knowing involves understanding and explaining mechanisms has received newfound attention within the philosophy of science. This work has helped to refine the concept of mechanism, and has converged on a description of what it means to possess mechanistic knowledge about a phenomenon. The word "mechanism" is commonly used to refer to machines or parts within machines, such as gear or pulley mechanisms in an elevator. However, the definition that has emerged in recent years applies to a wider range of phenomena across both engineering and science.

One such definition is provided by Glennan, who writes that a mechanism "consists of entities (or parts) whose activities and interactions are organized so as to be responsible for the phenomenon" (2017, p. 17). Another definition is offered by (Miłkowski, 2013): "While mechanisms are defined variously, the core idea is that they are organized systems, comprising causally relevant component parts and operations (or activities) thereof. Parts of the mechanism interact and their orchestrated operation contributes to the capacity of the mechanism" (Miłkowski, 2013). Competing technical definitions of the term provided by others (e.g., Machamer et al. (2000)) differ in the details, but all definitions share the same core elements: (a) a phenomenon or phenomena that can be explained or understood by (b) decomposing it into parts or components that (c) are organized in such a way that (d) they give rise to the phenomenon through their causal interactions (Illari & Williamson, 2012).

These concepts are best illustrated through an example. The cardiovascular system can be understood as a mechanism. Its phenomenon, or purpose, is to supply the body with oxygen and nutrients and to remove waste. Its entities, or parts, include the heart, blood vessels, valves, and blood. These entities are organized both structurally (e.g., the heart is divided into chambers separated by valves and tissue, the blood vessels contain valves and vary in size and thickness) and causally (e.g., valves open and close to keep blood moving in one direction, the chambers of the heart work together to keep blood circulating). In this literature, the causal relations are called activities; they are "the things that entities do... Activities are the causal components of mechanisms" (2006, p. 371). Mechanisms defined in this way are of central concern across most science and engineering disciplines. As Illari and Williamson writes:

> Even astrophysical mechanisms are grouped by the phenomena they produce. Scientists aim to give a detailed account of how the phenomenon is produced by entities and activities“ (Illari & Williamson, 2012, p. 4).

Expert mechanistic understanding of a phenomenon is theorized to involve all four of these elements: the phenomenon, the parts or entities, the structural organization, and the activities (i.e., the direct, causal interactions between the parts) (Craver, 2006). Investigations into the actual differences between experts and novices agree with this theory: novices possess partial knowledge of mechanisms while experts possess knowledge of all four of these elements. This distinction between experts and novices has been found in virology (Jee et al., 2015), ecology (Hmelo-Silver et al., 2007), and chemistry (Chin & Brown, 2000). More specifically, novices have been found to lack knowledge about the direct, causal relations (i.e., the activities) between parts in a mechanism. Hmelo-Silver et al. (2007) found that "there were minimal differences between the expert and novice groups on structures, but that the locus of the difference was on understanding causal behaviors and functions, the least salient elements of the systems" (p. 307); and Chin and Brown (2000) found that high-performing students produced explanations that "described nonobservable theoretical entities and cause–effect relationships", while lower-performing students produced "black box" explanations that did not discuss causal mechanisms at all.

The type of knowledge that novices lack is precisely the type theorized by Dewey to be constructed through the process of reflective inquiry. This implies that if engineering design is a valid form of reflective inquiry, then it is an activity that is capable of fostering the construction of knowledge about the causal relations between entities in a system.

2.4.3 Investigating Engineering Design

Makerspaces are an ideal environment for students to work on engineering design projects. The engineering design process is super-charged through the use of digital fabrication tools present in makerspaces. Students are able to make changes to digital design files using CAD/CAM software and send them to a 3D printer, laser cutter, or some other fabrication device to be physically realized. This makes it possible for students to quickly test and refine their designs, sometimes making multiple iterations within a single class period. As each iteration of the design cycle provides students with a new problem to solve, makerspaces and fablabs can be viewed as a kind of "engineering design gym" where students can rapidly gain experience working on ill-defined problems. This suggests that working on engineering-design projects in makerspaces may support the development of problem-solving skills in non-engineers. The first study in this dissertation was designed to learn more about the specific types of problem-solving skills that students working in makerspaces developed. Because of the theorized link between engineering design and learning about mechanisms, the choice was made to focus on changes in the students' ability to make sense of and solve problems involving mechanistic systems.

The second study in this dissertation was designed to more-directly test the theory that engineering design supports the construction of mechanistic understanding. In this study, two aspects of the engineering design process were isolated in order to investigate their role in the construction of mechanistic knowledge: problem-solving and reflection. One of the objectives in this study was to determine whether the problem-centered aspect of engineering design was necessary for learning about a mechanism, or whether any form of making was sufficient. Some types of making do not involve working through this process, such as assembling a kit with instructions, 3D printing a model that was downloaded from the Internet, or producing trinkets on a laser cutter. All of these are common activities that occur in Makerspaces in and out of schools (e.g., see Blikstein (2013)). By comparing problem-centered making to assembling with step-by-step instructions, it was possible to learn more about how these different forms of making support learning about mechanisms.

Chapter 3

Study 1: Building Mechanistic Problem-Solving Skills Through Engineering Design

> As a matter of fact, the development of an unconscious logical attitude and habit must come first. A conscious setting forth of the method logically adapted for reaching an end is possible only after the result has first been reached by more unconscious and tentative methods... (Dewey, 1910/1997, p. 108).

This study is an investigation into how long-term engagement in the engineering design process might foster the development of problem-solving skills in high-school students. The hypothesis being tested is that through working on long-term engineering design projects during a year-long course in digital fabrication, high-school students will become more skilled in solving problems involving mechanisms. This hypothesis is restricted to mechanisms of the type that students gained experience with during the course: simple, engineered systems. The reason for limiting the scope is based on research that shows that problem-solving skills are situated (Kirsh, 2009), and that the way a person frames a problem has an effect on their problem-solving approach. The way a problem is framed is dependent on the features of the problem and the environment, which means that problem-solving skills are unlikely to transfer from one context to others. This means that it would be unreasonable to predict that students who gained experience working through problems encountered in a course on digital fabrication would become better at solving problems with mechanisms in other domains such as biology or astronomy. However, evidence that students improved at solving problems like those encountered during the course would suggest that students engaging in similar activities in other domains might see similar improvements in solving problems characteristic of those domains.

There is evidence in the literature that students who work in digital-fabrication spaces experience an improvement in problem-solving skills. However, these studies have relied almost entirely on self-reported assessments in interview and survey data. Harnett et al. (2014) found that some university students who spent a semester working in a community hackerspace reported increased confidence in their problem-solving and project-planning abilities, and Galaleldin et al. (2016) reported that 60% of university engineering students reported feeling "more confident in their engineering knowledge and skills to solve a complex engineering problem". At the K-12 level, (Hartry et al., 2018) found that K-12 students working as interns in a museum Makerspace self-reported increases in problem solving skills, attributing this to their experience working on open-ended problems during their internship. Students were observed to be more capable of reasoning about and debugging complicated mechanisms after participating in a maker workshop (Blikstein, 2013). Fields et al. (2012) found that after a four-week electronic textiles workshop, students were more able to fix faulty designs like short circuits, poor crafting, and incorrect code.

There are a few reasons to believe that working on engineering design projects in digital-fabrication labs might be a particularly effective way of improving problem-solving skills.

The first reason is that the speed at which students can iterate through the design process is dramatically increased through the use of digital fabrication tools. Students are able to make changes to digital design files using CAD/CAM software and send them to a 3D printer, laser cutter, or some other fabrication device to be physically produced. This makes it possible for students to quickly test and refine their designs, sometimes making multiple iterations within a single class period. Because each iteration of the design cycle provides students with a new problem to solve, working on projects in these spaces provides students with opportunities to gain more practice in encountering and solving problems. Whether students have these types of experiences depends on many factors, including the design of the activities and the types of support provided by teachers, facilitators, and peers. But the potential is there, and this was one of the factors that influenced the choice to investigate changes in students' problem-solving abilities after participating in the digital-fabrication course.

The second reason is that the types of problems that students encounter are different from those that they typically encounter in school settings. The problems students encounter in Makerspaces and fablabs are often ill-defined, in the sense that it is obvious that something is wrong, but understanding the precise nature of the problem requires further investigation and testing (Robertson, 2003). The failure to provide students with opportunities to work on ill-structured problems has been identified as a weakness in other problem-based pedagogies, since these pedagogical methods typically provide students with well-defined problems which bear little resemblance to those encountered in non-schooling situations (Jonassen et al., 2006). Solving ill-structured problems requires much more than simply finding the right answer. Students must work to identify and characterize the nature of the problem, to hypothesize solutions, to implement these possible solutions during prototyping, and to evaluate their solutions by testing and observing their prototypes. This type of problem solving extends over longer periods of time, and is capable of producing deeper and more nuanced understanding of the problem (Sheppard et al., 2006). The ability to tackle ill-structured problems is highly-valued in many professions, including engineering (Jonassen et al., 2006), and students working in Makerspaces are provided with many opportunities to encounter and work on these types of problems.

A third reason is that the ways that problems are framed, and the types of support and guidance that students receive during problem-solving, are also not typically encountered in schools. Mistakes and failures are framed as an important part of the design process, essential to achieving the goal of arriving at a satisfactory design, as opposed to something to be avoided in fear of receiving a bad grade (Vossoughi & Bevan, 2014). This has been described as one of the distinguishing features that separates Making from other problem- and project-based approaches. Martin writes:

> Failure is not a happy word in most educational circles, particularly when attached to schools, students, or initiatives. Yet within the maker mindset, failure is celebrated... Failures in a school setting can be "productive" as well, helping students to better understand the structures and constraints of problems, so that they can learn better when given another chance. (Martin, 2015, p. 35)

Additionally, because of the ill-defined nature of the problems, students are likely to encounter situations where the teacher or facilitator does not know the solution. Instead of providing the solutions, teachers and facilitators may make suggestions or engage in the problem-solving process with students, making the types of problem solving that occurs in these spaces more like an apprenticeship or collaboration than a typical student-teacher relationship. This provides students with multiple opportunities to practice solving problems in an apprenticeship model, which is an ideal environment for the development of tacit problem-solving knowledge (Polanyi, 2005).

The research questions in this study were as follows:

1. Do students improve at solving mechanistic problems after taking part in a year-long course in digital fabrication?
2. If so, are the problem-solving skills that students develop during the course similar to those of expert mechanical engineers?

To answer these questions we invited a group of high-school seniors to take a year-long course in digital fabrication. Before and after the course, the students worked on (a) a set of hands-on problems involving mechanisms and (b) a paper-based assessment deigned to assess mechanistic thinking. We also invited a group of engineering experts to work on the hands-on problems. The two hands-on problems were designed to be similar to those that the students might have encountered while working on their projects in the course. The first problem was assembling a 3D-printed mechanical system (a differential gearbox) with no directions. The second problem was to troubleshoot and repair a broken electrical system (a flashlight), with an identical working version provided as a resource.

The hands-on problems were analyzed in two ways. The first method was to score participants on how close to the solution they came, and to compare the experts, pre-course students, and post-course students on their scores. The second method was to code the participants' sequences of actions taken and states visited while working on the problems. This produced a sequence of codes that represented the problem-solving process of each participant. This process data was analyzed using clustering methods to identify groups of participants who adopted similar problem-solving approaches. Full details about the methods and findings are provided in the sections below.

3.1 Methods

3.1.1 Participants

We recruited 20 high-school seniors (16 females, 9 males) and 18 graduate students (9 females and 9 males, mean age=24.67, SD=2.13) in mechanical engineering from an R1 university to take part in this study. Three high-school students had to be excluded from our analyses because they dropped out of the course, leaving 13 girls and 4 boys. The graduate students received a $20 gift card for their participation.

3.1.2 Design

The high-school students were randomly split into two groups at the start of the study. Group A (N=7) worked on the Gearbox-Assembly Task before the course and the Flashlight-Repair Task after the course, while Group B (N=8) worked on the Flashlight-Repair Task before and the Gearbox-Assembly Task after (Figure 3.1). On the Gearbox task, students in Group A served as the control group for students in Group B, who worked on the Gearbox problem after taking part in the course. On the Flashlight-Repair task, students in Group B served as the control group for students in Group A, who worked on the Flashlight-Repair problem after the course. The mechanical engineering experts worked on both tasks in a single session and did not take part in the course on digital fabrication.

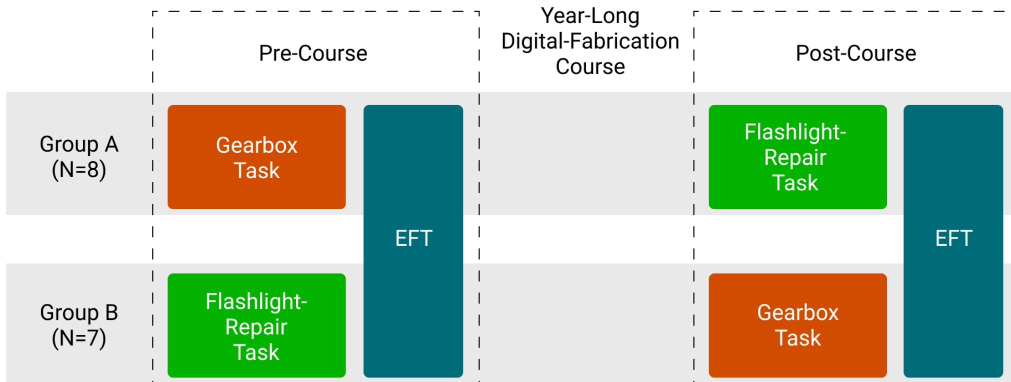
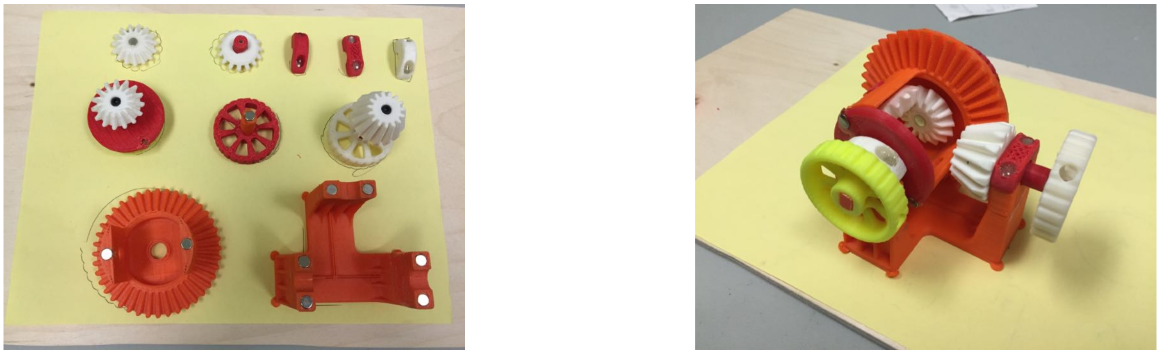


Figure 3.1: Study design. This does not show the experts, who worked on the Gearbox and Flashlight-Repair tasks in a single session.

3.1.3 Materials

3.1.3.1 The Gearbox-Assembly Task

The Gearbox-Assembly Task was designed to measure the participants' ability to assemble a complex geared mechanism with no instructions, images, or information about the completed object. The Gearbox was presented to the students in 10 pieces (Figure 3.2a). Each of the



(a) Unassembled Gearbox (b) Assembled Gearbox

Figure 3.2: The Gearbox-Assembly Task. Participants received the Gearbox in the unassembled state and had five minutes to assemble it with no instructions or images of the final assembly.

pieces contained magnets that stuck together when two pieces were correctly assembled. It is important to note that there were many ways of magnetically connecting the pieces that were not correct, while there was only one way of correctly assembling the Gearbox (Figure 3.2b).

3.1.3.2 The Flashlight-Repair Task

The Flashlight-Repair Task was designed to measure the participants' ability to troubleshoot and repair a faulty device. Two flashlights—one green, one red—were presented to each participant (Figure 3.3). The green flashlight was working and the red flashlight was broken. Three errors were present in the red flashlight: one of the batteries was reversed, the electrical contact in the base of the battery was inverted and disconnected, and the bulb was burnt out. The participants were not provided with any information about the broken flashlight, nor were they presented with any additional resources.

3.1.3.3 The Exploration and Fabrication Technologies Assessment

The students took the Exploration and Fabrication Technologies (EFT) assessment two times, once before and once after the course (Blikstein et al., 2017). The EFT assessment

Figure 3.3: The Flashlight-Repair Task

consisted of two main parts: a survey and an assessment of mechanistic thinking. The survey was designed to measure the students' post-high-school plans, their academic self-concept, and a battery of other attitudes. We did not analyze the entire survey, instead focusing on the mechanistic thinking questions. The assessment of mechanistic thinking consisted of two questions. The first question asked students to explain the steps they would take to repair a broken coffee maker (Figure 3.5). For this question, the students were provided with the following prompt:

> You come home one day and find that your electric coffee making machine (like the one in the picture) has stopped working. There are no visible problems with the coffee maker and no external buttons/switches are stuck. You check the filter only to find it's brand new with no coffee grounds in it at all. What steps would you take to fix it? For each step below please write a one to three sentences answers.

The second question asked students to identify the parts of a blender (Figure 3.4). For this question, the students were provided with a list of parts and asked whether the blender had the part as well as how confident they were in their answer. The parts the students had to choose from were motors, sensors, gears, batteries, LEDs, a program, light bulbs, screws, glue, and a camera.

	Does a blender have this part?	How sure are you? From 1 (you're not sure at all) to 10 (you're really sure)
Motors		
Sensors		
Gears		
Batteries		
LEDs		
Some kind of program		
Light bulbs		
Screws		
Glue		
Camera		

Figure 3.4: The Blender Question was designed to measure knowledge about the components in a device and confidence in that knowledge

3.1.4 Procedure

The study took place in four distinct phases: pre-course, course, post-course, and expert.

3.1.4.1 Pre-Course

On the first day of class, each of the students spent 15 minutes working on the EFT assessment. Additionally, each high-school student was randomly assigned to work on either the Gearbox-Assembly Task or the Flashlight-Repair Task. During the first week of the course, each student was asked to leave the class for a short period of time to work on the problem they'd been randomly assigned. The student was seated at a table and the task was placed placed in front of them. Video was collected using a GoPro camera facing the student. After completing the task, the student returned to class.

If the student had been assigned to the Gearbox-Assembly Task, they were told that the object in front of them had been disassembled, that it was their job to try and put it back together in five minutes, and that they should try their hardest and not be frustrated if they

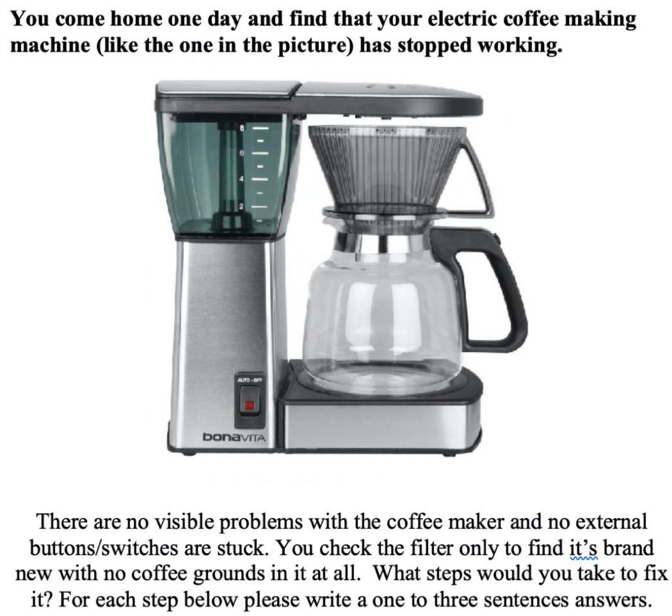


Figure 3.5: The Coffee-Maker Question was designed to assess strategies for troubleshooting a broken device

were unable to solve the puzzle. They were not given any further information about the object (i.e., no instructions on how to assemble the object). The participant was instructed not to touch the puzzle until the timer was started. Once the timer was started they had five minutes to try and solve the assembly puzzle. When the time expired, the puzzle was removed from the table.

If the student was in the Flashlight-Repair group, the two flashlights were placed in front of them. The participant was told that the red flashlight was not working, and that it was their job to repair it. They were shown how to turn on the green flashlight by twisting the head, which also demonstrated that the green flashlight was working. They were then given five minutes to repair the broken flashlight. When time expired, the puzzle was removed from the table.

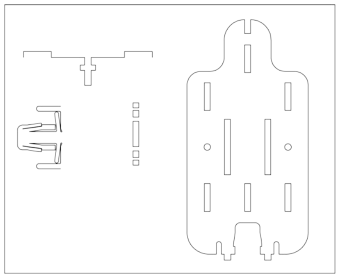
3.1.4.2 Course

This course took place in a makerspace on a university campus. The high-school students visited twice a week for roughly 1 hour per visit. The course was facilitated by the students' physics teacher and the lab manager.

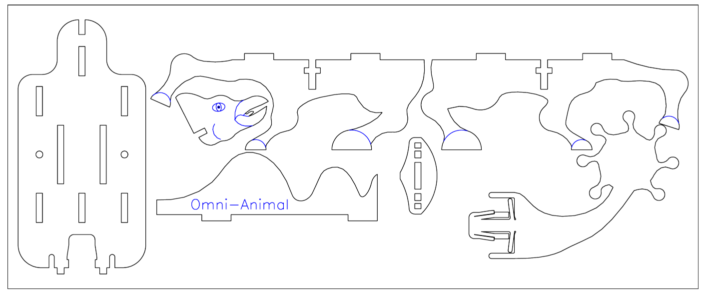
The high-school students worked on two multi-week design projects during the course. The projects were structured so the students had to meet certain, specific goals but were allowed creative freedom otherwise. The first project was the Omni-Animal, and the second project was the Rube Goldberg machine. The students worked on the Omni-Animal project in the first two months of the course. After finishing this project the students did not come back to the makerspace for two months due to administrative issues. When they returned, they worked on the Rube Goldberg project for an additional two months. The total amount of time that the students spent in the course was between 30 and 40 hours.

The Omni-Animal project required students to design a three-dimensional creature out of multiple two-dimensional pieces. This project was designed to give students experience with using two-dimensional vector-drawing software (CorelDRAW) to create a multi-part, three-dimensional construction. The students were given a starter template containing vector drawings of connectors (see Figure 3.6a) and an example Omni-Animal that had been cut out of wood and assembled. Additionally, they received direct instruction on using the software and had access to print-outs and other instructional resources on using the tools in CorelDRAW.

The Omni-Animal project was designed to engage students in multiple iterations of the design process. The students began with by sketching a number of possible designs, and then narrowing them down to a single idea. They then produced a sketch of their Omni-Animal idea, followed by a paper prototype of their design. Next, they converted their designs to vector drawings in CorelDRAW, then sent those designs to the laser cutter and cut the piece out of plywood. None of the students completed their project on the first try. In fact, the



(a) Omni-Animal template (provided to students)



(b) Example Omni-Animal (not provided to students)

Figure 3.6: Design files for the Omni-Animal project

students encountered problems at every stage of this process, with many problems being caused by errors in previous stages. For example, many students did not take the thickness of the material into account when designing their press-fit connections. When the female press-fit connector is wider or smaller than the material thickness, the connector does not work. This problem is not encountered until the pieces are cut from the plywood, and fixing the problem requires stepping backwards in the process and making changes to the vector drawing (Figure 3.7).

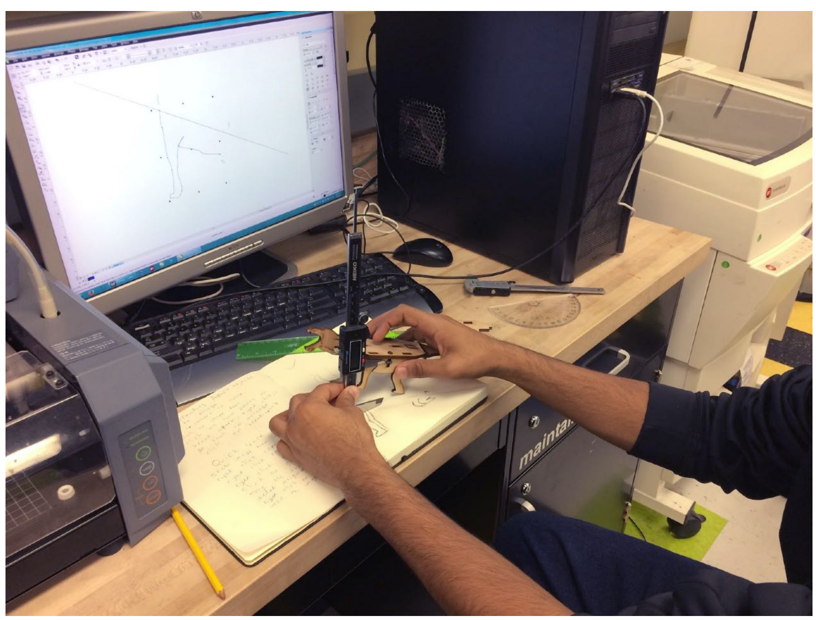


Figure 3.7: A student changing a vector drawing to fix problems with laser-cut Omni-Animal.

The project had a set of requirements that students had to meet: the completed Omni-Animal needed to utilize four different types of connectors: a snap-fit connector, a bolt-and-nut connector, a male-to-female press-fit connector, and a female-to-female notched connector. The snap-fit connector and bolt-and-nut connectors were both given in the design template, while the designs for the other two connectors were not provided. Outside of these requirements the students had full creative freedom.

The second project was the creation of a Rube Goldberg machine. A Rube-Goldberg machine is a mechanical contraption that uses a complicated series of interactions between a large number of components to perform a simple task. The students were broken into small groups, and each group was tasked with designing one of the components in a Rube Goldberg machine. The Rube Goldberg project was designed to give students experience with designing a multi-component, mechanistic system in a collaborative setting through an engineering design process. The students were broken into groups of two or three, and each group was responsible for constructing one stage of the machine. Each group was given a set of constraints having to do with the start and end actions of their component. For example, one group's constraints were that their stage should be activated when it detected heat, and it should trigger the following group's stage by creating a loud noise. The groups were required to collaborate with one another in order to ensure that the components they were designing would interact properly. Aside from satisfying these constraints, the groups were given full creative freedom. Like the Omni-Animal project, the Rube Goldberg project was intended to guide students through multiple iterations in the design cycle, and required students to make connections between the function, behavior, and structure of their stage in the machine.

3.1.4.3 Post-Course

The post-course phase occurred in the final week of the course. During this week, the high-school students were asked to leave class for a short period of time to work on a hands-on problem. If the student had been assigned to the Gearbox-Assembly Task in the pre-course phase, that student worked on the Flashlight-Repair Task in the post-course phase. Similarly, if the student had worked on the Flashlight-Repair Task in the pre-course phase, they worked on the Gearbox-Assembly Task in the post-course phase. Like the pre-course phase, each student was seated at a table with the task in front of them. During the task, video was collected using a GoPro camera facing the student. After time expired, the student returned to class. Additionally, on the final day of class each of the students spent 15 minutes working on the EFT assessment.

3.1.4.4 Experts

After the course had concluded, 18 graduate students in mechanical engineering (experts) were recruited to work on the hands-on tasks. Each expert was seated at a table and tasked with working on both the Gearbox-Assembly Task and the Flashlight-Assembly Task. The order of the tasks was randomized, and each expert had 5 minutes for each task. Video of their activity during both tasks was recorded using a GoPro camera.

3.1.5 Video Coding Schemes

We had two objectives which resulted in the development of two distinct video coding schemes. First, we were interested in evaluating how close each participant came to the correct solution on each task. The Correct-Combination Coding Scheme was developed for this purpose. Second, we were interested in understanding the types of actions that each participant took while working on the task. We designed the Actions-in-Time Coding Scheme for this purpose.

3.1.5.1 Correct Combinations

In order to meaningfully compare the participants on each task, it was necessary to develop a metric that accurately measured how close each participant came to the correct solution. This provided more information about each participant's progress than a binary complete/incomplete coding scheme. Because each task was different, we developed distinct coding schemes for each of them. Each of these coding schemes is detailed in the sections below.

3.1.5.1.1 Gearbox Correct Combinations Correctly assembling the Gearbox required 11 distinct combinations of parts. Each of these part combinations was assigned a distinct code. If the participant performed an action that matched one of the codes, they receive a point (1). If the participant's action matched a code partially, they received a half-point (0.5). No time information was recorded, so two participants who carried out the same actions in different orders would receive the same score. The scores for each combination were summed to create a single index of how close that participant came to solving the problem. A score of 0 meant the participant made no progress on the problem, while a score of 11 meant the participant completely solved the problem. The higher the score, the closer to finishing successfully.

3.1.5.1.2 Flashlight Correct Combinations We determined that correctly repairing the flashlight required four actions, and assigned each of these actions a distinct code. Participant actions that matched one of these codes were scored 1, and actions that partially matched a code were scored 0.5. A score of 0 meant the participant made no progress on the problem, while a score of 4 meant the participant completely solved the problem. The higher the score, the closer to the solution.

3.1.5.2 Actions in Time

In addition to comparing how close participants came to each solution, we were also interested in comparing differences in participants' actions during each task. We have developed a time-based coding scheme for the Gearbox-Assembly Task for this purpose.

3.1.5.2.1 Constructing Problem Traces The result of applying a time-based coding scheme is the construction of a problem trace for each participant. A problem trace is a sequence of codes that represent a participant's sequence of actions during a problem. There is no true or correct way to construct a problem trace for a given problem-solving episode; many different, representative problem traces exist for any given problem-solving episode. Once these problem traces were constructed, it was possible to identify groups of participants with similar approaches using unsupervised clustering methods.

3.1.5.2.2 Gearbox Actions in Time The Gearbox-Assembly Actions-in-Time coding scheme was developed to categorize and track the different types of actions participants carried out while attempting to solve the Gearbox problem. This coding scheme allowed us to analyze how sequences of actions differed between participants. The final coding scheme contained 14 codes that belonged to 4 categories: planning, evaluation, context, and action. The coding scheme complements the 11-point assessment scale and was designed so that it could be translated into other coding schemes used in similar analyses (Tschan, 2002; Worsley & Blikstein, 2013).

This coding scheme went through a number of iterations. Initially, the exact state of the gearbox as participants worked on the problem was recorded after each action. Each code consisted of a set of pieces with an optional sub-code indicating if any pieces were added or taken away in that moment. A finished sequence of codes provided a longitudinal description of how the state of the Gearbox changed over the course of five minutes. One major flaw of this coding scheme was the inability to distinguish between a correctly-assembled set of parts and an incorrectly-assembled set of parts. A second issue had to do with the number of unique codes. With just 10 parts, the number of unique Gearbox states able to be captured by this coding scheme was at least 2^{10} (1024). Ultimately, this coding scheme proved to be too noisy for any useful analysis.

The next and final iteration of our coding scheme treated the participants' actions in a more general way. Instead of hundreds of possible codes, we narrowed down the meaningful actions to 9 codes: exploring, looking, rotating, plastic connection, magnetic connection (correct), magnetic connection (incorrect), meshing gears, disassembling, and placing an axle into a hole or bracket.

We designed custom software to streamline the process of coding the videos. Each time the participant carried out an action, the appropriate code was entered and linked to the video using a timestamp (Figure 3.8). After coding a participant's video, we were left with a full sequence of the participant's actions during the problem. This process transformed video of the participants' actions into a time-stamped sequence of codes.

Figure 3.8: An example of a sequence of actions at different time points coded with the Time-Based Coding Scheme. From left to right: axle, rot(ate), plas(tic connection), mesh(ing gears), axle, mag(netic connection).

3.1.5.2.3 Flashlight-Repair Actions in Time The coding scheme for the Flashlight-Repair Task was designed to track each participant's actions as they worked through the problem. The coding scheme was also designed to keep track of which flashlight the participant was working on—functioning (green) or broken (red)—as well as which specific component each participant was focusing on.

The root of each code is the action. The codes for each action are as follows:

test testing

att attach

det detach

exam examine

comp compare

swap swap

org organize

frz no action

off-task off task

light-on light turned on

Each action code had 1 to 3 prefixes that provided additional detail. The first set of prefixes referred to the flashlight. Because the task involved one functioning (green) flashlight and one broken (red) flashlight, each action code was linked to one of these flashlights. This was signified by adding either a R (red) or G (green) before the action code.

{R, G} Red flashlight, Green flashlight

The second set of prefixes referred to the part that was being interacted with. The parts were as follows:

H Head of the flashlight

HR The reflector and glass cover that screws onto the head

L The actual bulb in the flashlight

RL The replacement bulb that is stored in the tail cap

B The batteries

T The tube of the flashlight

S The spring contact in the tail cap

C The tail cap

X Used when an action is incorrect (e.g., incorrectly attaching a part)

This coding scheme is best illustrated with an example. See Table 3.1 for an example of how this coding scheme was applied to a one participant's series of actions.

Table 3.1: Example of the Flashlight-Repair coding scheme. Note that the comp and swap codes do not have a flashlight prefix. This is because comparing and swapping always involve parts from both flashlights.

Qualitative Description of Participant's Problem-Solving Activity	Coding Scheme
Turn on the green flashlight. Since this is the working flashlight, the light turns on.	G,test G,light-on
Attempt to turn on the red flashlight. Since this is the broken flashlight, the light does not turn on.	R,test
Unscrew the tail cap from the red flashlight.	R,C,det
The contact spring falls out of the red flashlight.	R,S,det
Remove the batteries from the red flashlight.	R,B,det
Unscrew the tail cap from the green flashlight.	G,C,det
Remove the batteries from the green flashlight.	G,B,det
Compare the batteries from the red and green flashlights.	B,comp
Swap the batteries between the red and green flashlights.	B,swap
Put the batteries back into the red flashlight correctly.	R,B,att
Put the tail cap back onto the red flashlight.	R,C,att
Test the red flashlight.	R,test

3.2 Findings

3.2.1 Gearbox-Assembly Task

3.2.1.1 Closeness to the Gearbox Solution

The Gearbox-Assembly Task proved to be particularly challenging for the participants in this study. Few were able to completely solve the problem, regardless of their level of expertise: Only 2 out of 17 experts solved the problem, and none of the high-school students were able to solve it.

However, by counting the number of correct combinations that each participant carried out, we were able to measure how close each participant got to finding the correct solution. 11 unique part combinations were required to solve the Gearbox-Assembly Task. The maximum score a participant could receive was an 11, and the minimum score was zero.

The post-course students ($M = 3.69, SD = 1.85$) got significantly closer to the solution than pre-course students ($M = 1.94, SD = 1.15$), $t(11.69) = -2.27, p < 0.05$, Cohen's

$d = 1.14$. Additionally, the experts ($M = 6.82, SD = 2.44$) outperformed post-course students, $t(17.88) = -3.56, p < 0.01$, Cohen's $d = 1.38$ (Figure 3.9).

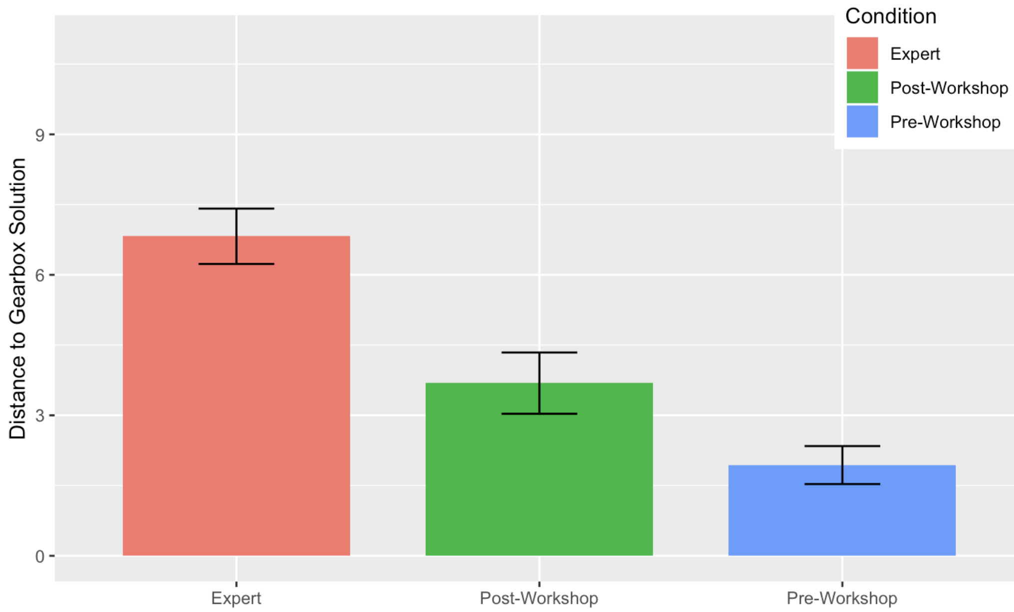


Figure 3.9: Correct combinations on the Gearbox-Assembly Task split between pre-course students, post-course students, and experts. The minimum possible score was 0 and maximum possible score was 11.

3.2.1.2 Cluster Analysis of Gearbox Problem-Solving Processes

One of the objectives of the study was to learn more about the differences in problem-solving approaches between experts and novices. Because we had no strong hypotheses about the precise nature of these differences, we began by using an unsupervised clustering method to identify groups of participants with similar approaches to the problems.

All of the participants' actions during the problem were coded, resulting in a string of codes representing each participant's sequence of actions taken during the problem. By computing the edit distance between all pairs of participants' sequences using TraMineR's optimal matching algorithm, we were able to construct a symmetric distance matrix that captured the similarity between all pairs of participants. After constructing this matrix, we used agglomerative hierarchical clustering (Maechler et al., 2016) to identify groups of participants who were most similar to each other using the R TraMineR (Gabadinho et al., 2011).

By clustering participants according to their entire set of actions on the Gearbox-Assembly Task, we identified two distinct groups of participants. The first group was composed of 16 experts and 3 post-course students. The second cluster was composed of all 7 pre-course students, 5 post-course students, and 1 expert (Figure 3.10). Subsequently, we will refer to the first cluster as the expert cluster (as it contains 94% of the experts) and the second cluster as the novice cluster (as it contains 100% of the pre-course high-school students).

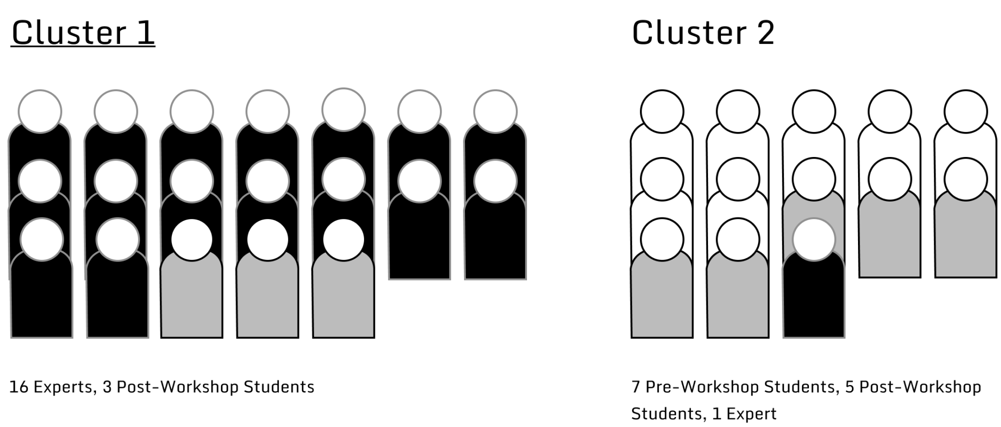


Figure 3.10: Makeup of the two groups found using hierarchical agglomerative clustering on the Gearbox problem.

The only sub-group split between the two clusters was the group of post-course high-school students. Our interpretation for this is that the course had a positive effect on the high-school students, making them more like experts on the task. After participating in the course, nearly half of the students were more similar in their problem-solving behavior to experts than to pre-course high-school students.

3.2.1.3 A Comparison of Problem-Solving Strategies on the Gearbox Problem

After performing the cluster analysis, we were interested in learning more about the differences between each cluster. We visualized the proportion of actions within each cluster and identified a number of differences (Figure 3.11). We identified four actions that the expert cluster performed at a higher frequency than the novice cluster: meshing gears (mesh), rotating pieces (rot), mounting axles (axle), and making correct magnetic connections (mag). In subsequent analysis we call these four actions "mechanical actions". We also identified two actions that the expert cluster performed at a lower frequency than the novices: incorrect plastic connections (plas) and incorrect magnetic connections (magx). We call these "structural actions" in subsequent analysis.

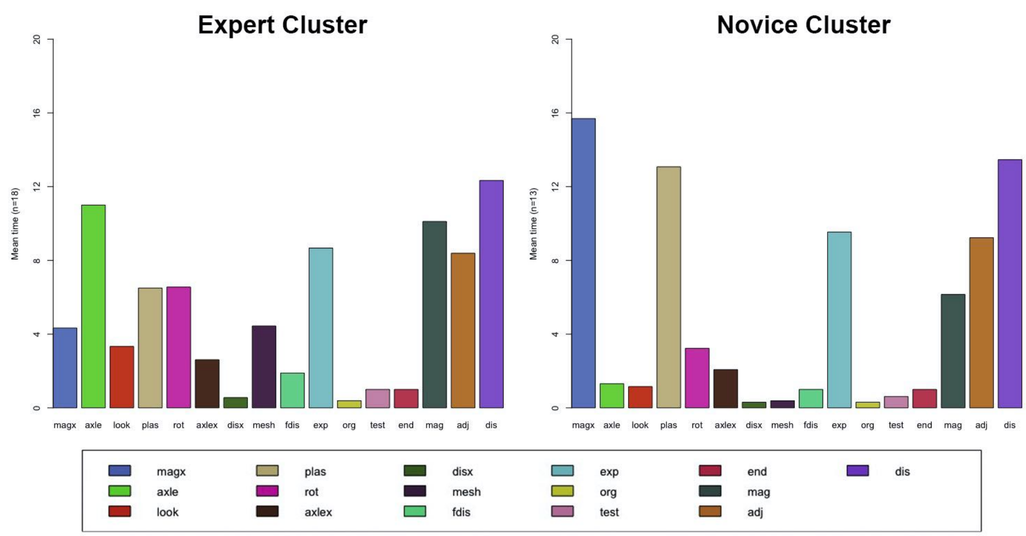


Figure 3.11: Proportion of actions for each cluster. Note the higher proportion of axle-related actions (green), meshing gears (dark purple), and rotation (fuchsia) in the expert cluster, and the higher proportion of incorrect magnetic connections (sky blue) and incorrect plastic connections (beige) in the novice cluster.

We created a mechanical-action index for each participant by dividing the sum of the four productive actions by the total number of actions for each participant. We performed 2 two-tailed t-tests to compare the differences in mechanical-action frequency between the pre-course students and the post-course students, as well as the differences between the post-course students and the experts. The post-course students ($M = 0.24, SD = 0.14$) performed significantly more mechanical actions than pre-course students ($M = 0.12, SD = 0.06$), $t(11.69) = -2.27, p < 0.05$. Cohen's $d = 1.14$. The experts ($M = 0.38, SD = 0.10$) performed significantly more mechanical actions than post-course students ($M = 0.24, SD = 0.14$), $t(17.87) = -3.56, p < 0.01$. Cohen's $d = 1.38$ (Figure 3.12).

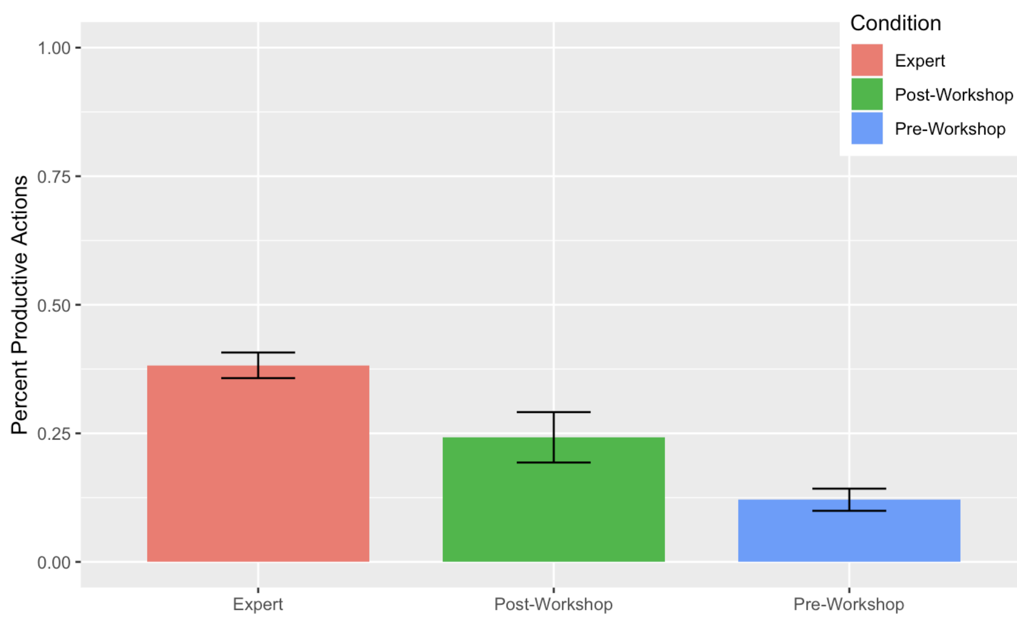


Figure 3.12: Proportion of mechanical actions taken by pre-course students, post-course students, and experts.

There is one action in particular that we want to highlight: meshing gears (mesh). Incredibly, none of the pre-course students meshed any of the gears during the five-minute task, despite the fact that five of the ten pieces included gears.

We compared the proportion of mesh actions using 2 two-tailed t-tests between pre-course students, post-course students, and experts. The post-course students ($M = 0.02, SD = 0.02$) performed significantly more mesh actions than pre-course students ($M = 0.0, SD = 0.0$), $t(7) = -3.5, p < 0.001$, Cohen's $d = 1.69$. The experts ($M = 0.08, SD = 0.05$)

performed significantly more productive actions than post-course students ($M = 0.02, SD = 0.02$), $t(21.34) = -4.88, p < 0.001$, Cohen's $d = 1.53$. It is worth highlighting that none of the pre-course students meshed any of the gears during the five-minute task despite the fact that five of the ten pieces included gears.

These results indicated that the expert and novice clusters had adopted qualitatively different approaches to the Gearbox problem. The experts and post-course students performed a significantly higher proportion of mechanical actions than the pre-course students. To determine if the proportion of mechanical actions a participant took was related to how close they came to the solution, a correlation analysis was done. These two measures were significantly correlated, $r(33) = 0.82, p < 0.001, r2 = 0.67$, indicating that the proportion of mechanical actions taken by a participant was a good predictor of how close they would come to solving the problem (Figure 3.13).

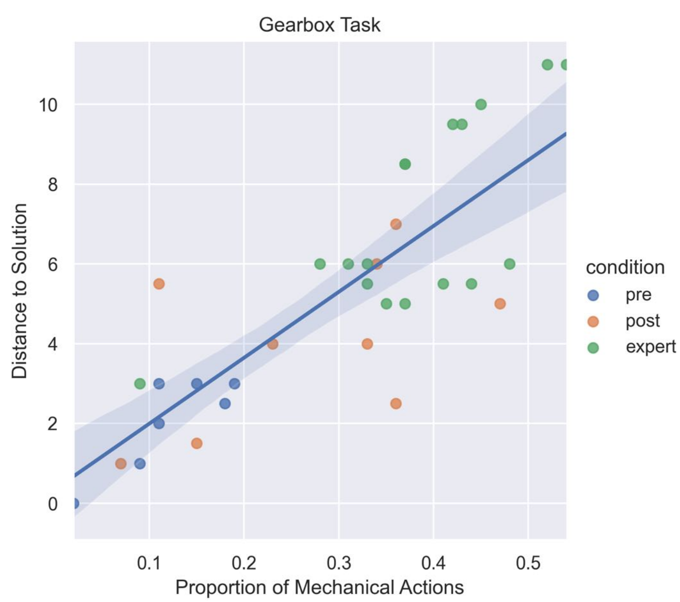


Figure 3.13: Scatterplot comparing scores on the Gearbox problem to the proportion of mechanical actions for each participant.

3.2.2 Flashlight-Repair Task

3.2.2.1 Closeness to the Flashlight-Repair Task Solution

The Flashlight-Repair Task was not as difficult as the Gearbox problem, but it still posed a significant challenge to the participants. 9 out of 17 experts solved the flashlight-repair problem, 1 out of 10 post-course students solved the problem, and 0 out of 9 pre-course students solved the problem.

In this problem, the broken flashlight had three sources of error: the batteries were inserted incorrectly, the spring contact in the cap was upside-down and failed to close the circuit, and the bulb was burned out. Successfully completing the task required repairing all three sources of error. By counting the number of errors corrected by each participant, it was possible to construct an index to measure distance to the solution. The minimum score a participant could receive was a zero (no sources of error corrected), and the maximum score was a three (all sources of error corrected).

The experts ($M = 2.47, SD = 0.62$) got significantly closer to the solution than the post-course students ($M = 1.9, SD = 0.57$), $t(20.52) = 2.43, p < 0.05$, Cohen's $d = 0.94$. Additionally, the post-course students made it marginally significantly closer to the solution than the pre-course students ($M = 1.22, SD = 0.97$), $t(12.61) = 1.83, p < 0.1$, Cohen's $d = 0.86$ (Figure 3.14).

3.2.2.2 Cluster Analysis of the Flashlight-Repair Problem-Solving Processes

To better understand the differences between expert and novice problem-solving approaches on the Flashlight-Repair problem that might have caused the difference in scores that we found, we used an unsupervised clustering method to identify groups of participants with similar approaches to the problems.

In order to extract meaningful clusters, it was first necessary to create a coding scheme that could allow differentiation between participants problem-solving approaches. Prior work

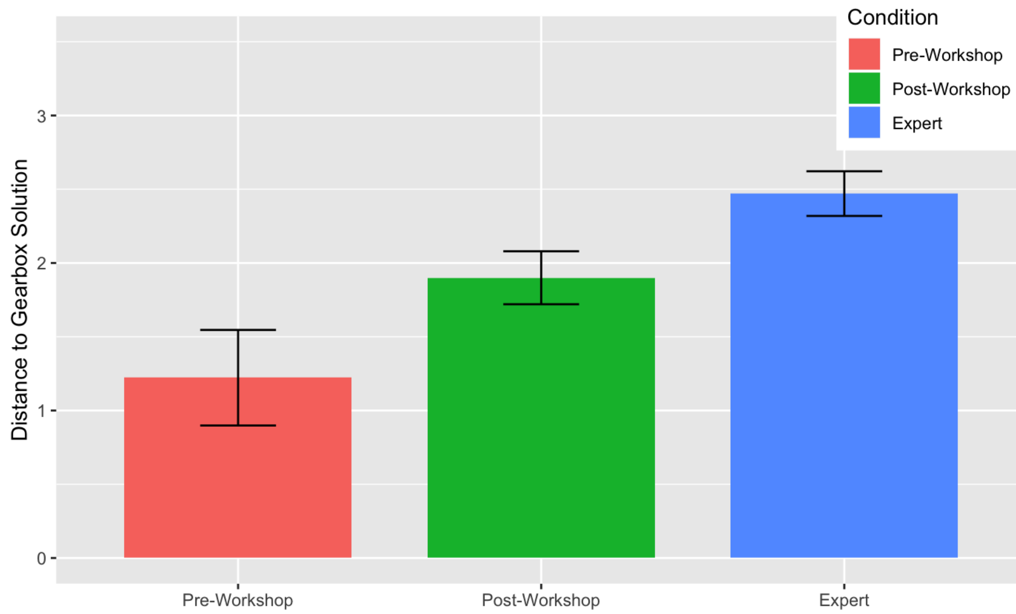


Figure 3.14: Closeness to flashlight solution

suggests that experts troubleshoot systems by engaging in a topographic search through a mental schematic (Rasmussen & Jensen, 1974). In the case of the Flashlight-Repair problem, the schematic would consist of an electrical circuit joining batteries, conductors, a switch, and a bulb. In the process of conducting the search, the expert-like participants would interact with all of these components, methodically checking them for errors by visually examining them or swapping them with parts from the working flashlight.

Based on this, we converted the existing sequence of codes containing action, part, and flashlight information into a simplified sequence that only contained part information. These simplified sequences of codes, which we refer to as interaction histories, allowed us to focus on which flashlight part each participant was attending to as they worked through the problem. Finally, we performed a cluster analysis on the participants' interaction histories using the same set of method described in Gearbox Cluster Analysis.

We identified two distinct clusters of participants. The first group contained 13 experts, 8 post-course students, and three pre-course students, and the second group contained four experts, one post-course student, and six pre-course students (Figure 3.15). We refer to the first group as the expert cluster since it contains 76% of the experts, and we refer to the second cluster as the novice cluster since it contains 67% of the pre-course students.

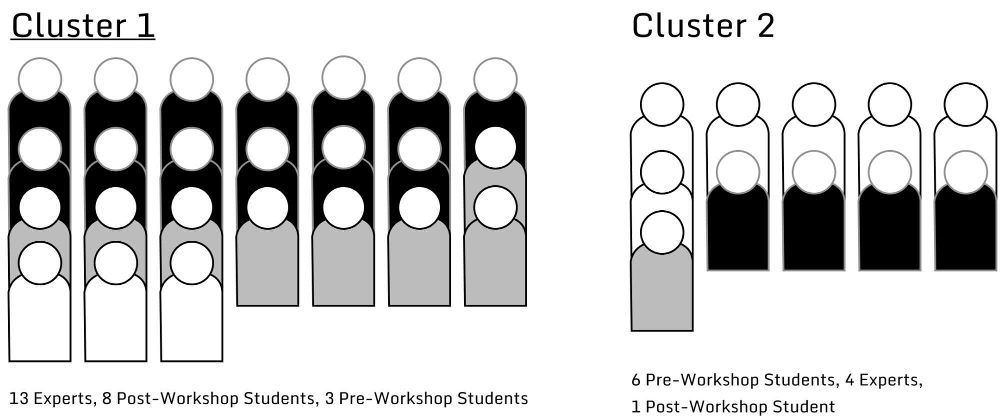


Figure 3.15: Makeup of the two groups found using hierarchical agglomerative clustering on the Flashlight-Repair task. The black figures represent experts, the grey figures represent post-course students, and the white figures represent pre-course students.

3.2.2.3 Exploring the Differences Between Problem-Solving Strategies on the Flashlight-Repair Problem

In order to understand the differences between clusters, we visualized the interaction histories within each cluster across the entire task (Figure 3.16). The expert histogram shows that the experts interacted with a more-uniform set of components, with more attention being paid to the batteries, cap, head, and bulb than the reflector, replacement bulb, and spring. This stands in contrast to the novice histogram, where the majority of interaction was weighted on a small number of components—the cap, batteries, and spring—with very little attention paid to the other components.

A second, longitudinal plot of interaction with components over the course of the task (Figure 3.17) provided more insight into this difference. Throughout the task, the expert cluster fluidly shifted their attention across components, presumably searching and testing for sources of error. In contrast, the novice cluster becomes increasingly fixated on a single source of error: the cap and spring. Additionally, the novice cluster paid little attention to the bulb, indicating that they had failed to consider it as a potential source of error.

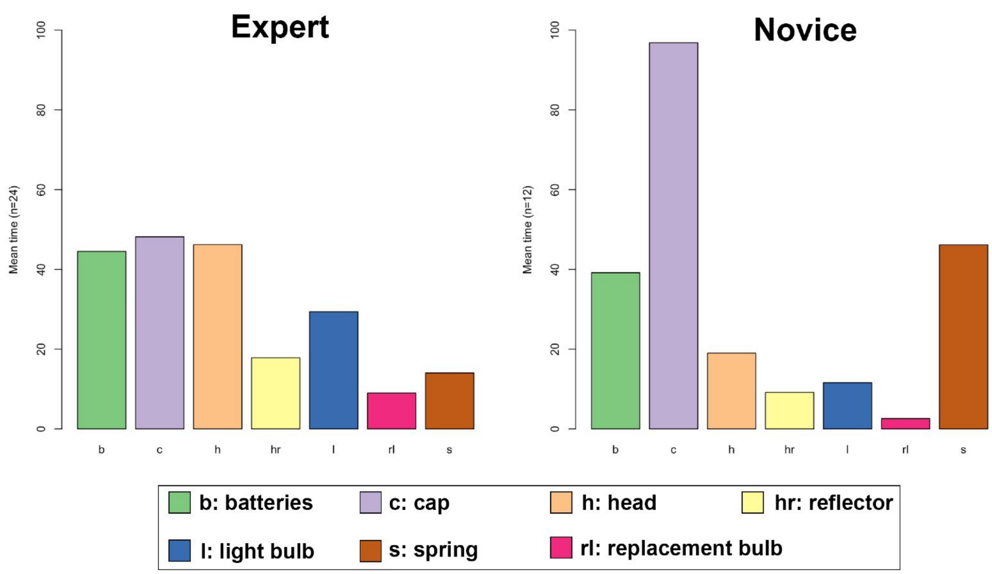


Figure 3.16: Cross-sectional plot of time spent attending to different components during the Flashlight-Repair Task.

This cluster analysis suggested that novices were more likely to interact with a smaller set of components (i.e., they became fixated on a specific source of error) while experts were likely to interact with a larger number of components (i.e., they considered all possible sources of error). To test this, we created an index to measure breadth of interaction by counting the number of components each participant interacted with during the problem.

Post-course students interacted with a significantly higher number of components ($M = 8.7, SD = 1.34$) than pre-course students ($M = 6.56, SD = 1.01$); $t(16.56) = 3.96, p < 0.01$. However, experts ($M = 9.24, SD = 1.30$) did not interact with a significantly larger number of components than post-course students, $t(18.56) = -1.01, p < 0.33$.

A between-group comparison of the number of unique configurations of components was

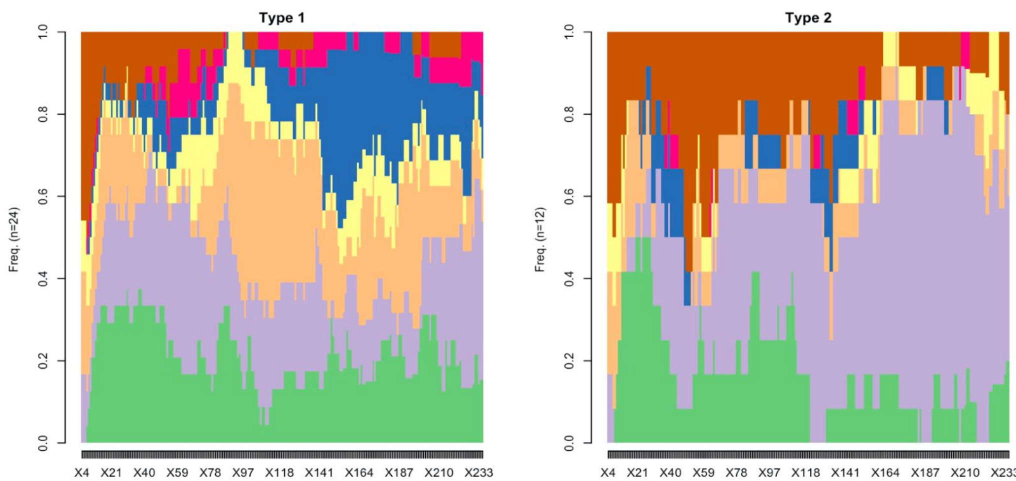
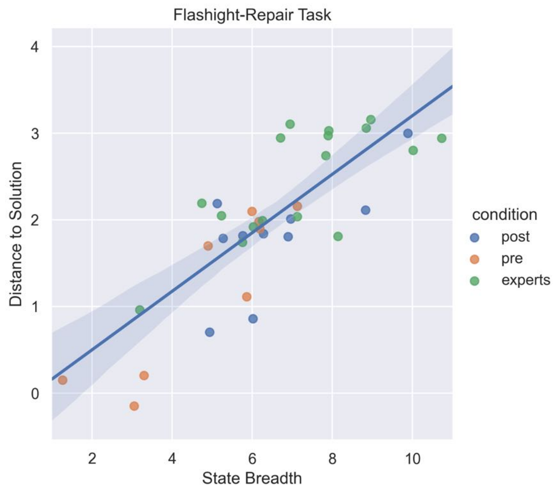


Figure 3.17: Longitudinal plot of attention paid to different components during the Flashlight-Repair Task.

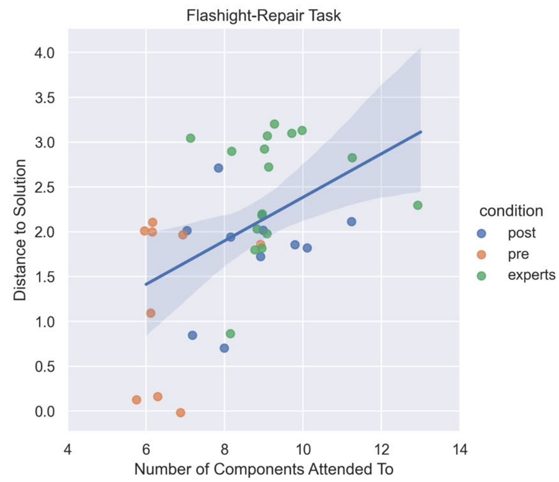
also performed. Each unique configuration of components in the flashlight represented one state in the full problem space. Post-course students visited a significantly-higher number of states ($M = 6.60, SD = 1.71$) than pre-course students ($M = 4.78, SD = 1.98$); $t(15.94) = 2.13, p < 0.05$. However, experts ($M = 7.24, SD = 1.99$) did not visit significantly more states than post-course students, $t(21.35) = -0.88, p = 0.39$.

To determine if there was a relationship between the number of components a participant interacted with and the number of problem-states they visited, a correlation analysis was performed. The correlation between number of states visited and number of components interacted with was not significant, $r(36) = 0.27, p = 0.11, r2 = 0.07$.

Two correlation analyses were performed to determine the relationships between the problem-solving approaches and progress towards the solution. First, the participants' scores on the Flashlight-Repair problem were compared to the number of unique states visited. These were significantly correlated, $r(36) = 0.83, p < 0.001, r2 = 0.69$. Second, scores on the task were compared to the number of parts interacted with. These were also significantly correlated, $r(36) = 0.46, p < 0.01, r2 = 0.46$. See Figure 3.18 for scatterplots of each of these



(a) Comparison of scores on the Flashlight-Repair task to number of states visited for each participant. These were significantly correlated, $r(36) = 0.83, p < 0.001, r2 = 0.69$



(b) Comparison of scores on the Flashlight-Repair task to number of componenents interacted with. These were significantly correlated, $r(36) = 0.46, p < 0.01, r2 = 0.46$

Figure 3.18: Comparing problem-solving strategies to scores on the Flashlight-Repair Task.

relationships.

Finally, to determine the contribution of both number of states visited and number of components interacted with on progress towards the solution, a multiple-regression was performed. The coefficient on number of parts interacted with was 0.13, and the coefficient on number of states visited was 0.31. Both of the coefficients on each of these independent variables were significant at $\alpha = 0.05$, and the r2 value for the complete model was 0.75.

3.2.3 The EFT Assessment

3.2.3.1 Knowledge of the Components in a Device

Students were presented with a photo of a blender and a list of ten components, and were asked to determine which parts were present in the blender. They were also asked to rate their confidence from 1 (not sure at all) to 10 (completely certain) for each component (Figure 3.4).

In the list of ten components, five of the components were present in the pictured blender and the other five components were absent. The components that were present were motors, gears, LEDs, screws, and glue; and the components that were not present were sensors, batteries, programs, light bulbs, and cameras.

The hit rate (true positive) and false-alarm rate (false positive) was computed for each student before and after the course. The hit rate was calculated by determining the proportion of present components that each student selected as present, and the false-alarm rate was calculated by determining the proportion of absent components that the student selected as present. The hit rate and false-alarm rates were used to compute d-prime values for each student using the Hautus adjustment for extreme values (Hautus, 1995). D-prime is a measure of accuracy that is calculated by taking the difference between the z-transforms of the hit rate and the false-positive rate. The higher a student's d-prime value, the better the student was able to identify which components are present and which components are absent.

After computing the d-prime values for each student in the pre-course and post-course conditions, an independent-samples t-test was performed to test for changes in d-prime. There were no significant differences between the pre-course students ($M = 0.53, SD = 0.58$) and the post-course students ($M = 0.66, SD = 0.55$), $t(26.60) = -0.61, p = 0.54$.

3.2.3.2 Confidence About One's Knowledge of Components in a Device

In addition to identifying the components in the blender, the students were also asked to rate how confident they were that their selections were correct. We averaged the confidence scores across all components for each student, creating a single index that represented each student's overall confidence in their knowledge about the components in a device. We conducted an independent-samples t-test to test for changes in confidence during the course. There was a marginally significant increase in confidence from the start of the course ($M = 6.82, SD = 1.85$) to the end of the course ($M = 8.11, SD = 1.76$), $t(26.756) = -1.98, p <$

0.06.

3.2.3.3 Strategy for Troubleshooting a Broken Device

Students were presented with an imaginary situation where their electric coffee maker suddenly stopped working, and that there were no visible problems with the coffee maker. The students were asked to detail the steps they would take to fix the coffee maker by providing one-to-three sentence explanations (Figure 3.5).

The students' responses were categorized according to a coding scheme that contained five main categories: open the coffee maker, make sure the coffee maker is broken, get help, check power, and give up. If the students' imagined approach to troubleshooting a broken device had changed during the course, there were two things we would have seen in their responses. First, we would have seen a change in the frequencies of actions taken. For example, we might have found that fewer students chose to give up or seek help after the course. Second, we would have expected to see a change in the sequence of actions taken.

To test for a change in the overall frequencies of actions, we conducted five Poisson regressions with the action as the dependent variable and the pre/post condition as the independent variable. None of these were significant. Thus, we did not find evidence that participation in the course led to a change in imagined troubleshooting actions.

To test for a change in the action sequence, we conducted three chi-squared tests of independence—one for each step in the full troubleshooting process—to examine the relationship between pre and post conditions and the imagined actions. We did not find significant differences between conditions in the distribution of imagined actions for the first or second steps; step 1, $\chi^2(4, N = 34) = 7.04, p < 0.14$; step 2, $\chi^2(4, N = 31) = 2.43, p < 0.66$. That is, there was no shift in the students' imagined troubleshooting strategy for the first two steps of repairing the coffee maker. However, there was a significant difference between conditions in the third step, $\chi^2(4, N = 32) = 16.42, p < 0.003$. The table of responses for this step can be found in Table 3.2. The largest difference between the two groups appears to be that the pre-course students report giving up on the third step, while the post-course students continued imagining ways to try and fix the problem.

Table 3.2: Actions that students reported they would take in the third step of troubleshooting the coffee maker

	Open and look inside	Check if broken	Get help	Check power	Give up	Total
Pre-course	2	0	4	0	10	16
Post-course	5	2	1	6	2	16

3.3 Discussion

This study was designed to learn more about how taking part in a long-term digital-fabrication course could affect high-school seniors' problem-solving skills. The research questions in this study were as follows:

1. Do students improve at solving mechanistic problems after taking part in a year-long course in digital fabrication?
2. If so, are the problem-solving skills that students develop during the course similar to those of expert mechanical engineers?

To answer these questions, we designed a set of hands-on problems involving simple mechanisms and asked the pre-course students, post-course students, and expert mechanical engineers to try and solve them. The participants' performance on the problems was coded in two ways. The first method was to score each participant on how close they got to solving the problem. The second method was to code each action or state visited as the participant worked through the problem, which produced a temporal sequence of codes that represented the problem-solving process.

The answer to the first question was that after the course, students were significantly better at solving a set of hands-on, mechanistic problems. Across both problems, the post-course students made significantly more progress towards the solutions than the pre-course students, which was related to a difference in problem-solving approaches. The higher the proportion of mechanical actions a participant took during the Gearbox task, the closer to the solution they came; and on the Flashlight-Repair problem, the more components that a participant interacted with, the closer they came to the solution.

Despite the fact that the high-school students did not learn about or work with geared mechanisms or electrical devices in the course, they still performed significantly better on both the Gearbox-Assembly Task and Flashlight-Repair task after taking part in the course. This suggested that the students had developed a set of problem-solving skills or approaches during the course that affected their ability to solve a class of problems involving engineered mechanisms. To determine the nature of these changes, the high-school students' process data were analyzed in comparison to the expert mechanical engineers. This provided the answer to the second research question: The approaches taken by the post-course students were significantly more like those of expert mechanical engineers across both problems.

Because we had no *a priori* hypotheses about the precise differences in approaches between experts and novices on these types of hands-on mechanistic problems, we used unsupervised clustering methods to find groups of participants who had adopted similar approaches. By examining the composition of these clusters, it was possible to identify distinct approaches to the problems and gain more insight into the differences between expert and novice mechanistic problem-solving.

3.3.1 Different Approaches to the Flashlight-Repair Problem

On the Flashlight-Repair problem, the experts were more likely to interact with all of the components in the flashlight that could have been potential sources of error, fluidly shifting their attention across components. In contrast, the pre-course students became increasingly fixated on a single source of error—the cap and spring—while failing to attend to or interact with the burned-out bulb. After taking part in the course, the high-school students were significantly more like expert engineers in their approach to solving the Flashlight-Repair problem. In fact, they were statistically indistinguishable from experts on this measure.

The same pattern was found when analyzing the number of unique problem states visited during the task. Each problem-state corresponded to a unique configuration of components in the flashlight. For example, the fully-assembled flashlight with the batteries correctly inserted and the fully-assembled flashlight with the batteries incorrectly inserted were each a state in the problem space. The experts visited significantly more problem states than the pre-course students, but no significant differences were found between experts and post-course students. Additionally, the post-course students visited significantly more states than the pre-course students. In sum, after participating in the course the high-school students interacted with more components, visited more states, and made more progress towards the solution than the pre-course students; and all of these changes made them more like expert engineers.

One explanation for these differences is that the participants who interacted with more components and visited more problem states were using a more systematic troubleshooting approach. In the literature on troubleshooting, there are two approaches that have been described that fit this pattern: topographic search and "swaptronics". A topographic search is theorized to involve the use of an organizational mental model, and involves tracing potential sources of error through the mental model and testing them systematically (Rasmussen & Jensen, 1974). The swaptronics approach, described by Lajoie and Lesgold (1992), involves systematically swapping out components one at a time until the source of error is discovered. The patterns of behavior identified in the process data suggest that the experts may have been using either of these approaches. Both of these approaches are theorized to involve the use of some sort of mental model (Halasz & Moran, 1983; Norman, 2014). Topographic search is theorized to rely on mental models of the causal and structural organization of a system, while a swaptronics approach can be conducted with a partial mental model containing the parts in the system alone.

However, it is likely that the pre-course and post-course students possessed mental models of the same quality. This is because the students did not learn about or work with electrical circuits or flashlights at any point during the course. This implies that the difference in approaches reflected a difference in strategy, not of prior knowledge, and that during the course the students gained experience with troubleshooting that they were able to utilize on the Flashlight-Repair problem. How this might have occurred is described in more detail in the section The Digital-Fabrication Course and Mechanistic Problem Solving below.

3.3.2 Different Approaches to the Gearbox Problem

On the Gearbox problem, experts performed a higher proportion of mechanical actions, while novices (i.e., pre-course students) performed a higher proportion of structural actions. The experts were more likely to mesh gears, mount axles in bushings, and rotate the pieces, while the pre-course students almost exclusively stacked pieces on top of each other and connected them with magnets. The post-course students were significantly more likely to perform mechanical actions than the pre-course students, indicating that participation in the course made the high-school students more like experts in this regard. A striking finding was that none of the pre-course students meshed two gears during the five-minute task, despite the fact that 5 of the 10 components were gears. A strong correlation was found between the proportion of mechanical actions taken and progress towards the solution, indicating the difference in approach (mechanical vs. structural actions) was responsible for the difference in performance on the problem between groups.

The difference in approaches on the Gearbox problem can be explained in terms of different ways that participants might have categorized the problem. The relationship between the way someone categorizes a problem and their problem-solving approach has been studied in the literature, where they are known as problem schemas (Chi et al., 1981; Hardiman et al., 1989). The problem schema that is activated during a problem is theorized to play a role in determining the problem-solving approach. In the Gearbox problem, one group seemed to have categorized the problem as a structural puzzle that could be solved by searching for a way to connect all of the magnets and stack all of the pieces. The other group seemed to have categorized the problem as reassembling a mechanical object, and tried to figure out how the parts worked together (e.g., trying to determine which gears interacted, and which wheels connected to which axles) in addition to looking for structural connections.

Categorizing the problem as involving a mechanism had the practical effect of making it easier to identify combinations of parts that brought the problem closer to the solution. For example, the gears contained useful information (tooth pitch, tooth depth, bevel angle) that could be used to determine the other gears that it meshed with. Each combination of parts found to work together also had the effect of restricting the problem space by reducing the number of free parts and ways of combining them, while simultaneously producing a more coherent understanding of the object-to-be-constructed. In contrast, the magnets contained little useful information (polarity). Connecting two pieces magnetically did little to limit the search space, as there was no clear way to test the correctness of the connection (as opposed to potential mechanical connections).

Similar to the previous problem, the change in approach cannot be explained in terms of a difference in content knowledge or experience working with gears. The students did not work with gears or geared mechanisms during the course. An alternative explanation is that during the course the high-school students gained a more general skill of identifying and working on mechanistic problems that affected the way they approached the Gearbox problem. A description of how this might have occurred is provided in the following section.

3.3.3 The Digital-Fabrication Course and Mechanistic Problem Solving

Students who took part in the digital-fabrication course experienced an increase in their mechanistic problem solving skills. They did not work with circuits or gears, and they were not explicitly taught mechanistic problem-solving strategies. However, they did iterate through multiple cycles of the engineering design process while working on their projects, which provided them with the opportunity to practice solving problems. The workshop was structured so these students had a number of forms of support to aid them in approaching these problems. They were encouraged to investigate their designs and talk to their peers, teachers, and facilitators in order to figure out the possible causes of their problems; to make changes to their designs that addressed these possible sources of error; and to create and test prototypes to verify whether they had solved the problem. This provided the students with practice in working on problems through a process that involved improving their understanding of the situation by testing and interacting with their prototypes. Because the students were using digital-fabrication tools to work on their projects, they were able to quickly test and refine their designs, sometimes making multiple iterations within a single class period. Working on these projects provided the students with guided practice, which has been shown to support the development of problem-solving skills (Lajoie & Lesgold, 1992).

The engineering-design process provides a way of improving one's understanding through the iterative process of prototyping, testing, reflecting, and redesigning. When a problem is encountered that is due to incorrect or incomplete understanding, a common way forward is to test and interrogate the prototype, to use this information to identify potential solutions, and to put these new ideas to the test through the creation of an updated prototype. As the process repeats, two things evolve: a more detailed and complete artifact, and a more detailed and complete understanding of the artifact. Of course, this is an idealized description of what actually occurs during the process of engineering design. In actuality, the process is messier, the types of problems encountered are of a much wider variety, and the process of solving problems is not guaranteed to result in learning. However, over a long-enough time period, students working on these sorts of projects are likely to encounter problems of this sort and gain some experience working on them. The high-school students who took part in this study were no exception. Over the months of working on their projects, many of the students were faced with problems related to their own projects, and in some cases they gained experience helping their peers work through these sorts of problems as well.

This study showed that high-school students who worked on projects through multiple iterations of the engineering design cycle built a stronger set of mechanistic problem-solving skills. After taking part in the course, the high-school students were better able to categorize problems involving mechanisms and adopted more productive approaches to solving them. These skills not only improved their performance on the problems, but also made them more like expert mechanical engineers.

3.3.4 Limitations

The analyses of the process data were valuable in identifying and characterizing the distinct approaches taken to work on the hands-on problems. However, because we did not assess the students' knowledge of the mechanisms used in these problems, it was not possible to conclude with full certainty whether the differences in approaches were due to differences in knowledge, skill, or some combination of both. One way of determining would be to use a think-aloud protocol during the problems. This method would not only make it possible to determine each participant's familiarity and prior knowledge about the problem, but it would also provide more insight into the problem-solving strategies that each participant was using. In a future study, using a think-aloud protocol to compare experts and novices on the same set of hands-on tasks could provide deeper insight into how the differences in actions reflected differences in strategy.

A second limitation was that this study used a between-subjects design to assess changes in problem-solving skills. This choice that was made to avoid test-retest effects; however, this had the effect of reducing the power of the study, and left open the possibility that the effects were not due to changes in problem-solving skill, but simply due to an uneven distribution of students within each group (e.g., the students with prior electrical knowledge ended up in one group, and the students with prior knowledge of gears ended up in another). A within-subjects design would have avoided these problems, and would have also made it possible to investigate the effects of the course on individual students, as opposed to only being able to examine group effects.

A within-subjects design would not be able to account for the possibility that observed changes were due to something that occurred outside the course. This would only be possible with the use of a control group who did not participate in the course, which the current study did not use. Because of this, there is a possibility that the changes we observed were due to other experiences that the high-school students had over the course of the school year. For example, there is the possibility that the students learned about circuits or geared mechanisms in another course, which could explain the differences found in this study.

3.3.5 Conclusion and Next Steps

This is one of the first studies to provide empirical evidence that working on long-term design projects in a makerspace has positive effects on students' problem-solving skills. In order to capture this change, we designed a set of hands-on problems involving mechanistic systems and developed two distinct coding schemes for each problem. The first coding scheme allowed us to quantify how close each participant got to solving the problem. By analyzing this score, we were able to show that post-course students were significantly better at solving the problems than pre-course students. The second coding scheme was designed to capture process data. Analyzing this data showed how the high-school students' approaches to working on the problems changed after the course, and by comparing the high-school students to a group of mechanical engineers it was possible to show that the post-course students were more like experts than the pre-course students.

This study did not just provide evidence that students who worked in a makerspace were better at solving problems than students who did not (though this would have been a useful contribution). It showed how working on a specific type of activity (engineering design) using a specific set of tools (digital fabrication) affected a specific type of problem-solving (mechanistic), provided details about the nature of these changes (changes in problem-solving approaches), and showed how participating in the course made the students significantly more like expert mechanical engineers. These findings suggest that it is worth looking deeper into the theorized link between engineering design and mechanistic understanding discussed in the Background section, which is the focus of the next chapter.

Chapter 4

Study 2: Engineering Design and Mechanistic Understanding

> What I cannot create, I do not understand.
>
> —Richard Feynman

The previous study provided evidence that engaging in the engineering design process supported the development of mechanistic problem-solving skills. The goal in the current study was to investigate the ways in which engaging in the engineering design process might also support the construction of mechanistic knowledge.

In the Background, I argued that engineering design should be considered a form of reflective inquiry, and like all forms of reflective inquiry, engaging in it could support learning about the causal relations in a mechanism. In reflective inquiry, the primary mechanism through which learning is theorized to occur is encountering and solving problems. The problem-solving process occurs in multiple stages: Identifying the problem, generating ideas and hypotheses, working out implications, testing and evaluation, and reflection.

One of the objectives in this study was to determine whether the problem-centered aspect of engineering design was necessary for learning about a mechanism, or whether any form of making was sufficient. Some types of making do not involve working through this process, such as assembling a kit with instructions, 3D printing a model that was downloaded from the Internet, or producing trinkets on a laser cutter. All of these are common activities that occur in Makerspaces in and out of schools (e.g., see Blikstein (2013)). By comparing problem-centered making to assembling with step-by-step instructions, it was possible to learn more about how these different forms of making support learning about mechanisms.

Another objective in this study was to investigate the role of explanation in learning through engineering design. This was based on informal observations in the previous study regarding students who were stuck on a problem with their design. Sometimes, during the process of explaining the problem to a peer or instructor, the student would suddenly see the solution. Explaining is a well-known strategy for solving problems in computer engineering, where programmers will sometimes explain their code, line-by-line, to a rubber duck to help them identify the bug in their code. Explanation can be thought of as a form of reflection, which is also theorized to play a role in learning through reflective inquiry.

There is a substantial body of research demonstrating the effect of self-explaining. Self-explaining has been shown to be an effective method across a number of domains, including physics (Chi & Bassok, 1989; Conati & Vanlehn, 2000), biology (Chi et al., 1994), and computer programming (Bielaczyc et al., 1995; Vihavainen et al., 2015). The results from these studies is in near-total agreement that self-explanation is one of the most effective methods for building explicit, declarative knowledge about mechanistic systems.

Self-explaining is theorized to support the construction of "systematic" understanding. A person possessing this kind of understanding is able to decompose a system into the structure, function, and behavior of its components (Chi et al., 1994). Recall that mechanistic understanding of a phenomenon is theorized to involve four elements: understanding of the phenomenon, understanding of the parts or entities, understanding of how the parts are organized, and understanding of the activities (i.e., the direct, causal interactions between the parts) (Craver, 2006). This is practically identical to the kind of knowledge that self-explanation has been shown to support. The SBF framework can be brought into rough alignment with the framework used in this dissertation. “Structure” corresponds to “parts and organization”, “function” corresponds to “the phenomenon”, and “behavior” corresponds to “activities” (Table 4.1).

Table 4.1: Comparing the dimensions of mechanistic understanding from Craver to those of the structure-behavior-function (SBF) framework.

Mechanistic Understanding Craver (2006)	SBF (Vattam et al., 2011)
Parts	Structure
Phenomenon	Function
Organization	Structure
Activity	Behavior

Self-explaining is billed as a constructive activity that helps build a coherent, systematic mental model while reading a text. In the process of learning about something, one’s understanding evolves from fragmented and incorrect to unified and accurate. During the process of reading a text, some of the information encountered contradicts pieces of fragmented understanding. Self-explaining is theorized to bring these contradictions to the awareness of the learner, triggering a process through which understanding is made more systematic and complete.

This theory helps explain the phenomenon observed in the previous study of students spontaneously solving problems during the process of explaining. Self-explaining forced the students to step back and talk through the problem, bringing contradictions between their ideas and their designs to the fore. Once made aware of this mismatch, the student was able to repair their conception and solve the problem.

Research on self-explanation has mostly focused on its effects while reading a text or working through worked examples. There is little work on how self-explanation and other constructive activities such as engineering design might interact. Chi et al. (1994) noticed a possible relationship between self-explanation and another constructive activity, drawing.

When comparing high self-explainers to low self-explainers, they found that the only students who produced a large number of drawings while reading a text were high-self explainers. (Chi et al., 1994) wrote:

> This suggests that drawing diagrams may be an alternative constructive activity for enhancing learning, so that the benefit of talking science may be its constructive nature, rather than the learning of the discourse and argument structure of science. However, it is not clear whether drawing diagrams alone would have been adequate at promoting greater learning, or whether it merely accompanied the activity of self-explaining... It needs to be studied more systematically. p. 460

Since then, there has been little work exploring this relationship.

Problem-solving and reflection are both aspects of the engineering design process theorized to support the construction of knowledge about mechanisms. This study was designed to investigate the role of problem-solving by comparing different modes of making (building with instructions vs. building with information) and different modes of self-explaining (self-explaining vs. reading) with respect to learning about mechanisms. The research questions in this study were as follows:

1. Is the problem-centered nature of engineering design important to supporting learning about mechanisms, or will other forms of making such as assembling with instructions also build mechanistic understanding?
2. How does self-explaining support learning about mechanisms? Does this process support learning about some dimensions of mechanistic knowledge more than other dimensions?
3. How does the combination of problem-solving and self-explaining support learning about mechanisms?
    (a) Do both activities support learning about the same dimensions of mechanisms, or does each activity foster a distinct form of understanding?

A 2x2 factorial design was used to test these hypotheses (Figure 4.1). The first factor compared building with information to assembling with instructions and aimed to isolate the effects of problem-solving during making. The second factor compared self-explaining to reading the text twice. The main activity in the study was learning about how a pendulum clock worked. Students in all four groups were presented with a partially-deconstructed LEGO pendulum clock. Additionally, all of the students had access to a webpage containing text, videos, diagrams, and animations about pendulum clocks. The webpage was broken up into sections, with each section corresponding to a specific component in the clock. The ways this webpage was contextualized within the learning activity differed between groups.

For the two problem-centered making groups, the primary learning activity was using the information in the webpage to help them figure out how to reconstruct the clock. The webpage was contextualized as a source of information that students could use to help them solve the problem. In contrast, the primary learning activity for students in the other two groups was to read through the webpage to learn about the pendulum clock. They were asked to read through the webpage, section-by-section, and at the end of each section they were shown a photograph of where to place the component. The activity of reconstructing the pendulum clock with access to information about how clocks work, but without step-by-step instructions, was intentionally designed to create a situation where the specific effects of problem-solving in engineering design could be investigated. As soon as the activity began, students found themselves in a problematic situation that involved figuring out how to reconstruct an unknown mechanism. Most engineering design projects will bring students into similarly problematic situations, where the current state of the project is distinct from the desired outcome, and closing the gap requires figuring out the root of the problem[1]. In the pendulum-clock task used in this study, the root of the problem was lack of knowledge about the pendulum clock mechanism, which made it difficult for students to figure out where

[1]These types of problems include troubleshooting and debugging, reverse-engineering, analysis, and design.

to reattach the pieces to get the clock working correctly. So, while this problem did not arise organically during the course of a larger engineering design project, it was representative of the types of problem that do arise during these types of projects, and as such provided a way to investigate the pedagogical impact of working on these types of problems when making. Analyzing the differences between the problem-centered making groups (those who had to figure out how to reconstruct the clock) and the groups who assembled with instructions on a set of posttest questions provided data that was used to answer the first research question in this study.

To answer the second research question, two groups were asked to explain, out-loud, each section of the webpage after they finished reading it, while the other two groups did not explain the text on the webpage. An analysis of the differences between the conditions who self-explained and those who did not provided data that was used to answer the second research question.

Because the study employed a factorial design, one of the groups received both treatments: first, they rebuilt the clock using the webpage as a source of information; then, they were asked to read through the webpage and self-explain at the end of each section. A comparison of the differences between this group and the other three groups was done to answer the third research question.

4.1 Methods

4.1.1 Participants

69 community college students taking part in social science courses at a local community college were recruited through a participant pool and received course credit for participating in the study. Students were prescreened by the number of engineering courses they had taken in high school or college. Any students who reported having taken more than two engineering courses were excluded from participation.

4.1.2 Design

Students were randomly assigned to one of four groups: Build, Explain, Build+Explain, or Read-Twice (Figure 4.1). All groups began the study with a short pretest, moved onto a learning phase, and concluded with a posttest (Figure 4.4). During the learning phase, all groups worked with the same basic set of materials: a webpage providing information about pendulum clocks, and a partially-deconstructed LEGO model of a pendulum clock. The only difference between groups was the treatment they received during the learning phase.

The study employed a 2x2 factorial design, with the two independent variables being building [rebuilding the clock with step-by-step instructions vs. figuring out how to rebuild the clock using information from the webpage] and explaining [reading the webpage twice vs. self-explaining at the end of each section]. The dependent variables were depth and quality of mechanistic knowledge about the pendulum clock and ability to transfer that knowledge to explain a related mechanism (a mechanical watch). More details about how these dependent variables (DVs) were operationalized and measured are provided in the Materials section.

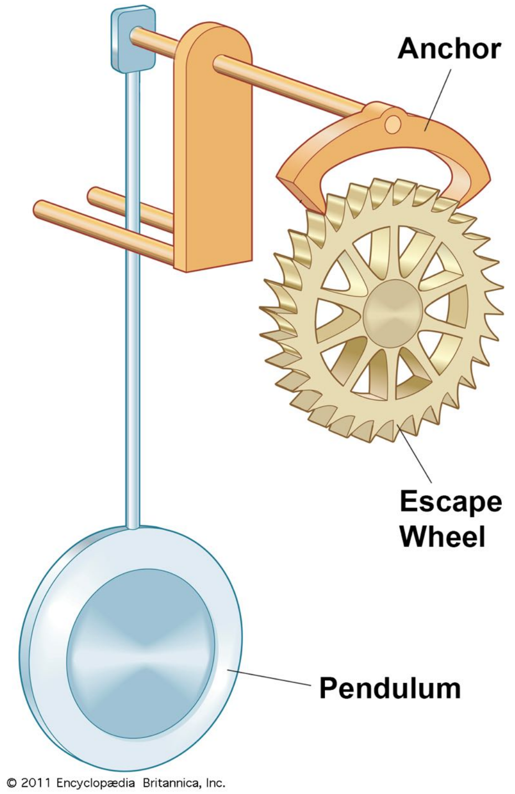
		Build	
		With Instructions	With Information
Self-Explain	No	Read Twice	Build
	Yes	Explain	Build+Explain

Figure 4.1: 2x2 factorial design for the mechanical clock study.

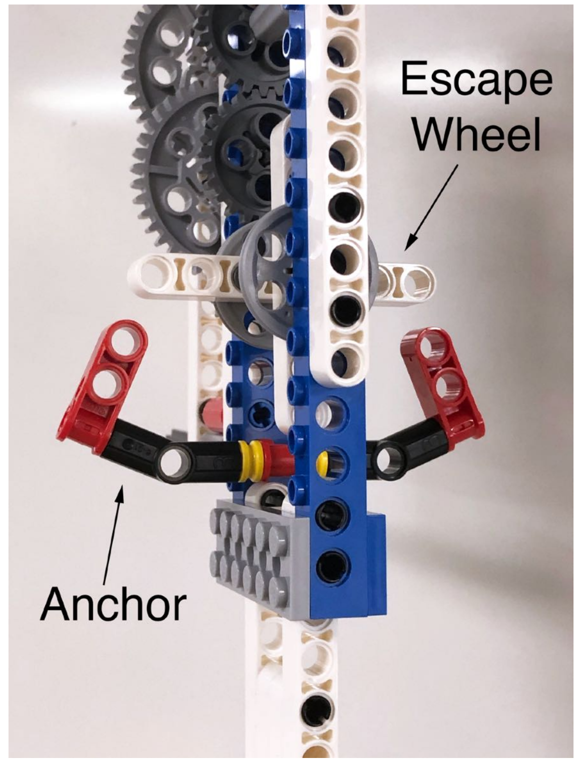
At the start of the learning phase students in the Explain group (N=17) were instructed to go through the webpage one section at a time, reading the text, looking at the diagrams and animations, and watching the videos. They were told to inform the researcher when they reached the end of a section. Each time a student reached the end of a section, they were asked to re-explain the material from that section in their own words while referencing the materials on the webpage. After explaining, the participants were instructed to find the corresponding LEGO component on the table and correctly attach it to the LEGO clock. To help students find the LEGO part, cut-out pieces of paper showing labeled photographs of each component in the LEGO clock were placed on the table as well (Figure 4.2). Additionally, participants were shown a photograph of where to place the part. Only after explaining every section on the webpage and assembling the clock did students in the Explain group move onto the posttest.

Figure 4.2: Five labeled images presented to students during the learning phase

Students in the Build group (N=18) and Build+Explain group (N=17) were both tasked with reconstructing a LEGO pendulum clock by figuring out where to attach four pieces: the weight, the pendulum, the anchor, and the hand. They were instructed to use the webpage as a resource to learn about the clock and its parts and to use that information to help them solve the problem. Additionally, they were informed that if they got stuck or had questions, the researcher was there to help them. All of the diagrams and videos on the webpage showed the pendulum clock with its parts oriented in ways that were substantially different from their orientation in the LEGO model. For example, on the webpage pendulum clocks were always shown with the anchor above the escape wheel, while the actual placement of the anchor in the LEGO model was below the escape wheel (see Figure 4.3). This was done to ensure that students in both the Build and the Build+Explain groups could not simply match the diagrams to the LEGO model to determine where to place each component. Instead, students would need to learn about the purpose of each component as well as its relations



(a) A diagram used on the webpage with the anchor above the escape wheel.



(b) The actual placement of the anchor on the LEGO model was below the escape wheel.

Figure 4.3: An example of the differences between diagrams on the webpage and the actual placement of parts in the LEGO model.

to other components in order to figure out where it belonged. After reassembling the clock, the Build group moved onto the posttest, while the Build+Explain group was asked to go through the webpage and explain each of the sections out loud. Like the Explain group, the Build+Explain group was instructed to first read the text of each section and then explain it to the researcher in their own words. This additional explanation phase was the sole difference between the Build+Explain group and the Build group.

Students in the Read Twice group (N=17) were asked to read through the materials on each section twice, and to inform the researcher before moving onto the next section. Each time they reached the end of a section, they were asked to place the corresponding part back on the LEGO clock. Like the Explain group, they were provided with images of the parts and their correct location on the LEGO clock.

4.1.3 Procedure

The study took place in a small room at a local community college. The room contained a single table and two chairs, one for the participant and one for the researcher, and the student and researcher sat side-by-side during the study. A laptop computer was provided so that students could access the webpage about pendulum clocks. A partially-assembled LEGO clock was present on the table along with the four parts that had been removed: the anchor, the pendulum, the weight, and the hand (Figure 4.5a). Five labeled photographs of the parts in the LEGO clock were placed on the table underneath each of the parts (Figure 4.2). The full study was video recorded using a GoPro Hero 5, and the screen of the laptop computer was recorded as well.

The study took place over the course of one hour, and participants were run one-at-a-time. In the first 5-10 minutes of the study, students worked on a short pretest. After this, they moved onto the learning phase, which lasted 15-30 minutes. Finally, the students worked through a two-part posttest. The first part of the posttest involved answering questions about the pendulum clock; the second part of the posttest involved answering questions about an analogous mechanism, a mechanical watch; and the third part of the posttest involved working on a physical problem, the reconstruction of a verge & foliot clock. The learning phase differed between groups, but the pretest and two-part posttest were identical for all groups (Figure 4.4).

The written pretest and posttest were administered on an Apple iPad. During the written pretest and posttest, the researcher read all text out loud to the student. This helped ensure that students read all of the questions, that they considered all of the options for multiple-choice answers, and helped keep the amount of time spent on the written pretest and posttest as close as possible between students. Students were given 10 minutes for the

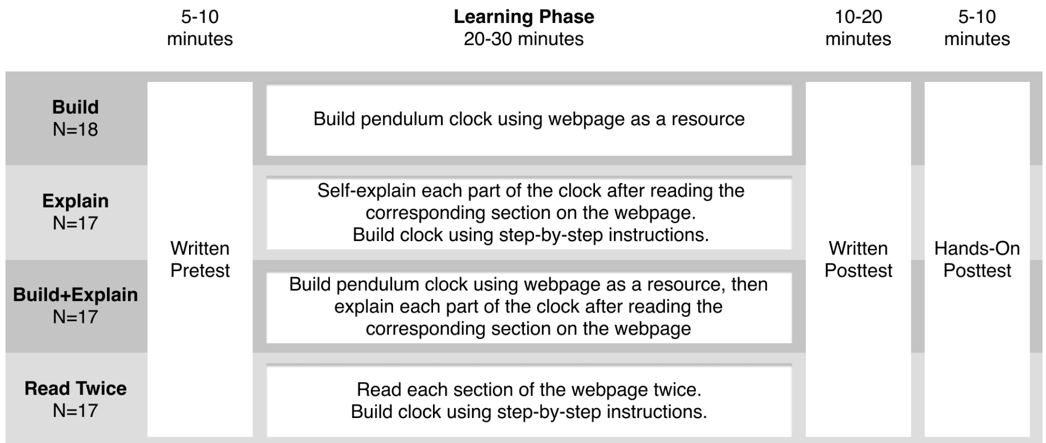


Figure 4.4: Study Procedure.

hands-on posttest, reconstructing the verge & foliot clock. The study concluded once a student successfully reconstructed the clock or reached the 10 minute mark.

4.1.3.1 Learning Phase Procedure: Explain Group

The learning phase for the Explain group started with the researcher opening the webpage about the pendulum clock and moving the laptop in front of the student. On the table next to the laptop was a partially-deconstructed LEGO pendulum clock. The parts that had been removed were placed on labeled images (Figure 4.5a). The researcher began by informing the students that their goal was to learn about how pendulum clocks work. The researcher read the instructions at the top of the webpage to the student, which instructed students to go through the webpage one section at a time, reading the text, looking at any diagrams or animations, and watching any videos. The researcher then showed how each section of the webpage corresponded to a part in the pendulum clock, and that each of these parts could be found in the functional LEGO model.

Students were asked to work through the webpage one section at a time, and to stop once they reached the end of a section and felt that they understood the material. At this point, the researcher asked them to explain the section in their own words. They were allowed to consult the materials on the webpage, but were asked to explain in their own words. The instructions on how to self-explain each section, provided at the top of the webpage, were as follows:

> After you read each section, explain out loud what it means to you. Your explanation should address these three questions:
>
> 1. What is the name of this part?
> 2. What is the purpose of this part? What role does it play in the clock?
> 3. How does this part work? That is, what other parts does this part touch or directly interact with, and how does this part work together with those parts to do its job?

These instructions were read out loud to the students at the beginning of the learning phase by the researcher. During the learning phase, students were prompted to self-explain using a similar set of questions each time they reached the end of each section. The questions asked to the students at the end of each section were as follows:

1. What is the name of the part? Where is it in the LEGO clock?
2. What does it do? What is its job?
3. How does it work?

In cases where the explanation was vague, confused, or incorrect, the researcher prompted the students for clarification. If the student was unable to provide a coherent explanation, the student was asked to return to the webpage section and then try again.

The webpage began with a section called the Big Picture, which provided an overview of mechanical clocks as well as a short video about how they work. Students in the Explain condition did not explain this section, since explaining the pendulum clock was one of the posttest questions. However, they did explain all of the other sections corresponding to individual parts of the clock.

The choices to prompt students about each part's purpose (prompt 2) and its mechanical relations to other parts (prompt 3) was made to align the self-explanation prompts to those of Chi et al. (1994) as closely as possible. In that study, 22 specific function prompts were used throughout their 101-sentence passage. These prompts were provided whenever a component was discussed (e.g., the atrium), and asked the student to state explicitly the component's function. As all of the sections in the webpage corresponded to components in the pendulum clock, it was logical to use functional prompts exclusively. The choice to prompt students for clarifications when their explanations were vague or confused was also based on Chi et al., who asked students questions like "What do you mean?" when the explanations were vague.

4.1.3.2 Learning Phase Procedure: Build Group

The Build group's learning phase began with the researcher opening the webpage about the pendulum clock and moving the laptop in front of the student. The researcher stated that the goal was to learn about how pendulum clocks work by reading and learning from the webpage in combination with figuring out how to reconstruct the mechanical clock. In this condition, the mechanical LEGO clock had been partially disassembled. The four parts that had been removed from the LEGO clock were placed on the table—the anchor, the pendulum, the weight, and the hand—along with labeled images of each part (Figure 4.5a). The researcher informed the student that they would not need to remove or reconfigure parts that were already on the clock. Their goal was to use the information on the webpage to help them figure out where to reattach the four components so that the LEGO clock would properly function.

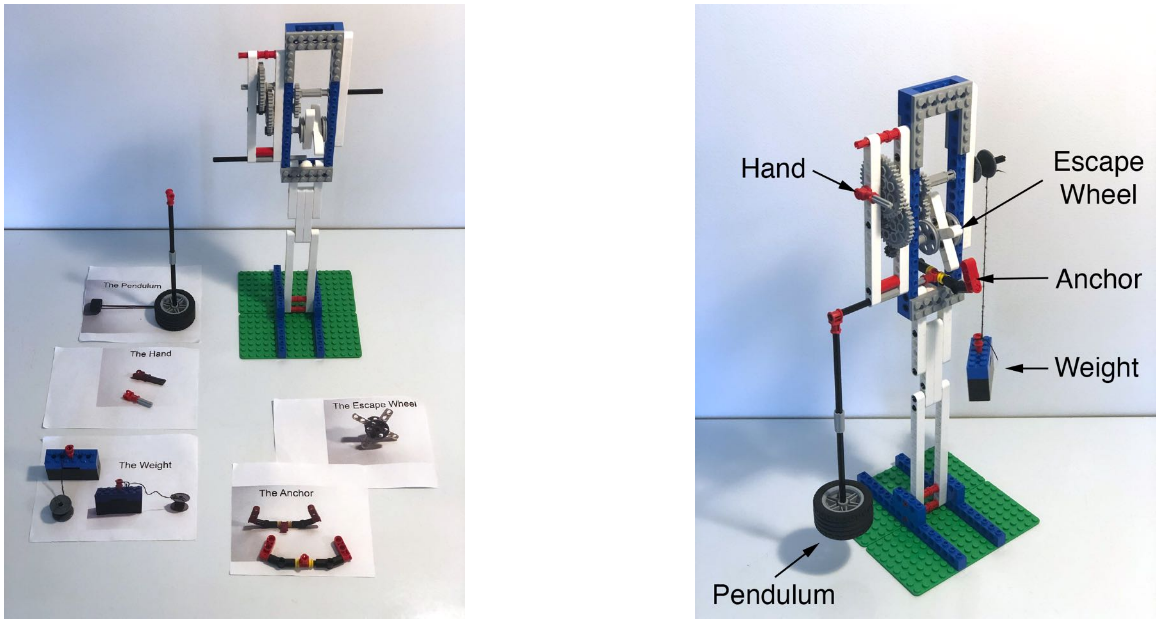
As discussed in the Materials section, there were 208 ways that components could be attached to the clock. This meant that it was inevitable that some students would become stuck while working on the task. Students were told that if they got stuck, they could ask the researcher for help. When asked for help, the researcher would point the student to the section of the webpage that covered the part they were struggling with, asking them to re-read the text and re-watch the videos. Additionally, the researcher would encourage the student to pay attention to how the part interacted with the other parts instead of focusing on the orientation of the part in the diagrams and videos. In most cases, this form of help was successful in getting the student unstuck. In the cases where the student remained stuck even after receiving help, the researcher would step in after 15 minutes and explicitly tell the student which parts were incorrectly attached. If the student failed to make progress after 20 minutes, the researcher would correctly attach a part, then give the student roughly 60 seconds to attach another part correctly. Each minute, the researcher would attach a part and give the student the opportunity to attach another part. This continued until the clock was correctly assembled. Once the clock was correctly assembled, the student moved onto the posttest.

4.1.3.3 Learning Phase Procedure: Build+Explain Group

The learning phase for the Build+Explain group was a combination of the Build and Explain learning phases. First, students were asked to reconstruct the LEGO clock. The procedure used was identical to that of the Build group. However, after successfully reconstructing the clock the students did not move onto the posttest. Instead, students in this condition were then asked to explain each section of the webpage to the researcher. The prompts and procedure used during this explanation phase were identical to those used with the Explain group. After explaining each section of the webpage, students in the Build+Explain group moved onto the posttest.

4.1.3.4 Learning Phase Procedure: Read-Twice Group

The procedure used during the learning phase for the Read-Twice group started in the same way as the procedure used with the Explain group. However, students in this condition were simply asked to read each section of the webpage twice, and were not allowed to move on from a section until they confirmed with the researcher that they had done so. After reaching the bottom of the webpage, the students moved onto the posttest.



(a) Partially-constructed LEGO clock and parts with labeled images.

(b) The fully-assembled LEGO clock with all parts labeled.

Figure 4.5: The start-state and goal-state of the LEGO pendulum clock.

4.1.4 Materials

4.1.4.1 The Pendulum Clock

The LEGO pendulum clock used in the study was fully-functional and mechanically-analogous to other pendulum clocks, though it only had a single hand and did not keep accurate time. This clock was from the LEGO Simple & Powered Machines educational kit[2]. Like all pendulum clocks, the LEGO pendulum clock worked through the interaction of a weight, a series of gears, an escape wheel, an anchor, a pendulum, and a hand (Figure 4.2). After winding up the weight and releasing it, the LEGO pendulum clock would run for about one minute before the weight reached the ground and the clock stopped. Before the participant entered the room, four of the parts were removed from the LEGO clock and placed on the table on top of labeled images (Figure 4.5a).

[2]A full set of instructions for building to clock can be found at https://education.lego.com/en-us/lessons/spm/click-clock.

4.1.4.2 The Webpage about How Pendulum Clocks Work

A webpage with information about pendulum clocks was provided as resource during the learning phase in all four conditions. The webpage content consisted of a mix of original content created for the study and media collected from third-party websites such as YouTube and Wikipedia. The text on the webpage describing the clock and its components was written to model a high-quality textbook chapter, and the videos, animations, and images were included to complement the text by providing dynamic visualizations of the various mechanisms in the clock. Although in the original work on self-explanation figures from the textbook were deleted, leaving only text (Chi et al., 1994), more recent work has shown that the effect of self-explanation is undiminished when learning with multiple representations (Berthold et al., 2009; van der Meij & de Jong, 2011). The decision to include multiple forms of media was made in order to mimic the types of resources a student might find when searching for information about a problem on the Internet. In fact, all of the third-party material used on the webpage were resources that a student would actually find if they were to look for information about pendulum clocks on the Internet.

There were three webpage variants, one for each condition. These webpages can be accessed at the following URLs:

- https://foothill2019.weebly.com/clocka.html (Webpage for the Build and Build+Explain conditions)
- https://foothill2019.weebly.com/clockb.html (Webpage for the Explain condition)
- https://foothill2019.weebly.com/clockc.html (Webpage for Read-Twice condition)

There were two differences between the webpage variants. First, the instructions at the top of the webpage were different across conditions. Details about these differences can be found in the Procedure section. Second, images showing the location of each component in the LEGO clock were provided for the Explain and Read-Twice conditions, but not for the Build and Build+Explain conditions. These images were included as a form of step-by-step instructions to help students in the Explain and Read-Twice conditions locate the correct location to place each part on the LEGO clock. Providing these images for the Build and Build+Explain groups would have ruined the treatment by making the reconstruction task trivial, as students would be able to simply look at the image and immediately put the part on in the correct location, as opposed to needing to figure this out from the information on the webpage.

The webpage consisted of six sections on a single page. Students could scroll from section to section, or use the navigation bar at the top of the webpage to jump directly to sections. The six sections were the Intro, the Big Picture, the Weight, the Anchor and Escape Wheel, the Pendulum, and the Hand. The Intro section contained the instructions for the task, which differed between groups (as described above). The Big Picture section provided a 14-sentence written overview of the parts in a mechanical clock and how they work together. This section also contained a labeled diagram of a pendulum clock and a 71-second video that explained how a pendulum clock worked[3]. The section on the Weight contained a nine-sentence passage that explained the role of the weight and how it interacts with other parts in the clock. This section also contained an animation showing that as the weight falls, it causes gears inside the clock to turn. The Anchor and Escape Wheel section contained a 13-sentence passage that explained how the two parts work together to regulate the speed of the clock. This section also contained an animation, a static diagram, and two videos. The first 50-second video showed an animated 3D model of the escape wheel, anchor, and pendulum, and explained how these parts worked together. The second 84-second video showed a large working pendulum clock and explained how the entire clock worked, with specific focus on the anchor, escape wheel, and pendulum. During pilot testing it was found that students struggled to understand this part of the clock more than any other, which motivated the choice to include extra information, including multiple videos. The Pendulum section contained an 8-sentence passage about the pendulum, in addition to an

[3]As mentioned in the Procedure section, students in the Explain conditions did not explain this section.

animation and a diagram. Finally, the section on the Hand contained a six-sentence passage as well as a static image.

4.1.4.3 The Written Pretest and Posttest

Each student started the study by taking a written pretest, and ended by taking a written posttest and working on hands-on posttest problem [4]. The pretest contained questions in the following categories: previous experience (4 questions); cognitive reflection (3 questions); confidence related to mechanical systems (5 questions); interest in mechanical systems (2 questions); and knowledge (2 questions). The posttest contained questions in the following categories: mechanical reasoning ability (1 question); confidence related to mechanical systems (7 questions); interest in mechanical systems (3 questions); mental effort and perceived difficulty of the learning phase (4 questions); and knowledge about mechanical clocks (12 questions). See Table 4.2 for a complete mapping from the question IDs on the pretest and posttest to their respective categories.

Table 4.2: The IDs for questions on the pretest and posttest by category.

Category	Question ID(s) on Pretest	Question ID(s) on Posttest
Knowledge	13, 16	2, 3, 4, 5, 6, 7, 17, 18, 19, 22, 26, 27
Previous Experience	1, 2, 3, 12	NA
Mech. Reasoning	NA	1
Cognitive Reflection	4, 5, 6	NA
Confidence with Mech. Systems	7, 8, 9, 14, 15	8, 9, 10, 20, 21, 23, 24
Interest in Mech. Systems	10, 11	11, 12
Difficulty of the Task	NA	13, 14, 15
Enjoyment of the Task	NA	16

4.1.4.3.1 Knowledge Questions 12 knowledge questions were included in the posttest, and 2 were included in the pretest. The choice to use an unmatched pretest and posttest

[4] The full pretest and posttest can be found in the Appendix.

was made based on the observation during pilot testing that none of the students possessed any prior knowledge about mechanical clocks or other related systems. Matched pretests and posttests are typically used in situations where students have varying levels of prior knowledge which needs to be accounted for when measuring learning. This was not a concern in the present study, since none of the 24 students who took part in the pilot study were found to have any prior knowledge about mechanical clocks or watches.

There were seven types of knowledge questions in total. Two of the question types were based on prior work by Chi et al. (1994):

Verbatim Recall information explicitly stated in the text

Inference Draw new inferences that were not explicitly stated in the text (e.g., inferring the function of a component from its structure, inferring how components relate to one another)

The remaining fives question types were as follows:

Manipulation check A basic question that every student should have been able to answer

Drawing from Memory Draw a diagram of the pendulum clock from memory with no prompts

Verbatim Explanation Explain out loud how a pendulum clock works

Near Transfer Answer questions about a similar mechanism, a mechanical watch

Near Transfer Explanation Explain out loud how a similar mechanism works, a mechanical watch

Each of these question types, with the exception of the manipulation check, were chosen to allow the assessment of multiple dimensions of students' understanding about the clock mechanism. Using a suite of question types provided a more complete picture of the nature of students' understanding, which made it possible to determine whether different dimensions of knowledge about the clock were affected by the different interventions. The types of knowledge assessed on the posttest were declarative knowledge, mental models, mechanistic understanding, transfer of mechanistic understanding, and problem-solving transfer. Table 4.3 shows which question types were intended to assess these forms of knowledge.

In the context of this study, declarative knowledge was assessed by (1) asking students to answer questions about the content in the text and videos and (2) by asking them to reason about the behavior of the clock in hypothetical situations. The first type of question assessed verbatim declarative knowledge (i.e., how well a student was able to recall information encoded during the learning task), and the second question type assessed inferential knowledge (i.e., a student's ability to combine multiple pieces of information from the text with prior knowledge to construct new knowledge). This is a common form of assessment in research on self-explanation, and is most similar to the types of questions found on typical exams.

Declarative questions are well-suited to assessing some types of knowledge, but are not ideal for others. For example, it was difficult to fully-assess the students' mental models using questions that assessed declarative knowledge alone. A mental model of a working pendulum clock encodes the various components, their structural organization in the system, and their mechanistic interactions. In other words, a mental model of a pendulum clock is three-dimensional, dynamic, and contains many distinct elements. The Draw-a-Pendulum-Clock question was designed to capture information about students' mental models of the mechanical clock that would have otherwise been difficult to elicit. The use of drawings to investigate students' mental models and understanding has been used in a number of domains, such as information science (Gray, 1990), science education (Reiner et al., 1995; Smith et al., 2019), and ecology (Eberbach et al., 2012; Yu et al., 2016).

However, there are limitations to the use of drawings as assessments. The drawings students' made were static and two-dimensional, while the mechanical clock itself was a dynamic, three-dimensional system. Because of this, drawings fail to capture the dynamic, mechanistic interactions between components, which is crucial to understand when assessing one's knowledge of a mechanistic system. This made it difficult to understand the full nature of the students' mental models from the drawings alone. For example, just because two parts in a drawing were represented as interacting, it could not be definitively concluded that the student who made the drawing understood how or why they interacted. The only thing that could be definitively concluded from this was that the student was able to recall those parts, their structure, and their organization in the clock (Russ et al., 2008).

In order to compensate for this limitation, two explanation questions were included in the posttest. One of the explanation questions asked students to explain how the entire pendulum clock worked in their own words, while the other asked them to explain the workings of the escapement and pendulum, which is the most complex and important subsystem in the clock. By coding and analyzing the students' explanations, it was possible to gain deeper insight into their mechanistic understanding.

The whole suite of questions provided more coverage of the students' knowledge than a single question type would have. This strategy has been used in other research investigating knowledge of mechanistic systems, such as Hmelo-Silver et al. (2007), who used both drawings and explanations to assess expert and novice knowledge about an aquarium. A careful analysis of performance across the different posttest questions made it possible to determine the distinct effects of the two interventions of interest on learning. More details about the questions themselves are provided in the sections below.

Table 4.3: Types of knowledge questions included in the posttest and what dimensions of knowledge they were designed to measure.

Questions	Dimension of Mechanistic Knowledge	Posttest ID
Verbatim Questions	Declarative knowledge	5, 6, 7
Inference Questions	Declarative knowledge	2, 3, 4,
Draw-a-Pendulum-Clock	Mental models	17
Explain-a-Pendulum-Clock	Mechanistic understanding	18, 22
Explain-a-Mechanical-Watch	Transfer of mechanistic understanding	26, 27
Reconstruct the Verge & Foliot Clock	Problem-solving transfer	

4.1.4.3.1.1 Verbatim Questions The three verbatim questions were all multiple choice, and asked students to recall details about the pendulum clock that they had learned about from the webpage. The three verbatim questions were as follows:

- How does the escape wheel move in the clock? (Correct answer: "It rotates in a single direction")
- What keeps the escape wheel moving? (Correct answer: "The weight falling")
- What is the purpose of the pendulum in the clock? (Correct answers "To keep the parts in the clock moving at a constant rate" and "To regulate the gears from spinning out of control")

The answers to all three verbatim questions were explicitly stated in the text or in the narration of the videos.

4.1.4.3.1.2 Inference Questions Two of the three inference questions were multiple choice, and one was open response. Answering these questions correctly required making inferences about the working of the clock based on what had been learned during the learning phase. The three inference questions were as follows:

- Clocks with pendulums tend to run faster when cold. Why? (Correct answer: "The pendulum becoming shorter when cold")
- What would happen if you were to double the weight in the mechanical clock? (Correct answer: "Nothing would change")
- Which of these parts would you remove to stop the motion of the clock and prevent the gears from turning and the weight from falling? (Correct answer: "The pendulum")

All three of these questions involved making inferences about the clock mechanism, and none of the answers were explicitly stated on the webpage. Answering these questions required integrating multiple pieces of information. For example, answering the first inference question ("Why do clocks with pendulums run faster when cold?") required integrating three pieces of information. First, that the pendulum is the part that controls the frequency of the clock. Second, that the length of the pendulum is the sole factor that determines the speed at which it oscillates. Third, that things (including the pendulum) contract when they get cold. So, the clock runs faster when cold because the pendulum contracts, which increases the frequency at which it oscillates, which leads the clock to run faster.

The intuitive, incorrect answer to the second question ("What would happen if the weight were doubled?") was that it would cause the clock to go faster. This is related to the common misconception that heavier objects fall faster than light objects, and also to the idea that more weight means more energy, and more energy means a faster rate of operation. Correctly answering this question required students to not immediately accept this intuitive response (the same ability measured by the CRT-2 questions in the pretest), but it also required them to think about the fact that the pendulum alone determines the rate of the clock. The third question ("Which part would you remove to stop the clock?") required students to understand the mechanical role each part played in the clock, and to correctly infer that the pendulum was the only choice that could be correct[5].

4.1.4.3.1.3 Draw-a-Pendulum-Clock Question The Draw-a-Pendulum-Clock question asked students to draw a diagram of the mechanical clock. The purpose of this question was to gain insight into the completeness and coherence of the students' mental models of the mechanical clock. Of particular interest were the parts that were present in the drawing, and how the parts were connected or interacted with one another.

4.1.4.3.1.4 Explain-a-Pendulum-Clock Questions In the Explain-a-Pendulum-Clock questions, students were provided with a diagram of a pendulum clock that they had

[5]If either the anchor or escape wheel were removed, there would be nothing to regulate the weight from falling immediately to the ground, which would cause the clock to run faster. And the hand is just an indicator, so removing it would have no effect.

not previously seen in the study (Figure 4.6) and were asked to "provide an explanation of what this system does and how it works". During the learning phase, students watched two videos that each contained a complete explanation of how a pendulum clock worked. Additionally, the text in the first section of the webpage (The Big Picture) provided a full description of how a mechanical clock worked. However, at no point during the learning phase were any students asked to provide a full explanation of the pendulum clock. Even students in the Explain and Build+Explain groups never generated a complete explanation of the clock. This question was the first time that any of the students generated a full explanation during the study.

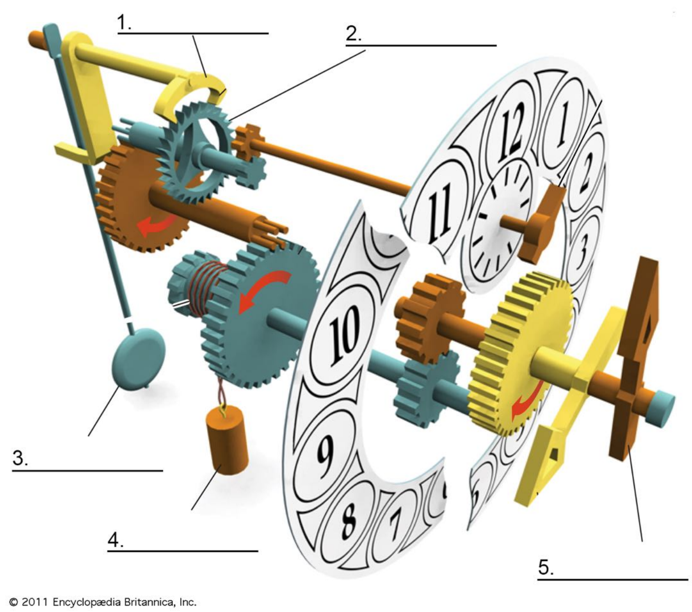


Figure 4.6: Pendulum-clock diagram

4.1.4.3.1.5 The Explain-a-Mechanical-Watch Transfer Question The Explain-a-Mechanical-Watch transfer question asked students about a mechanical watch. Mechanical watches work analogously to pendulum clocks, and each part in a mechanical watch has a corresponding part in a pendulum clock. For example, the source of power in a pendulum clock is the weight, which turns the gears in the clock as it falls to the ground. In the mechanical watch, the source of power is the mainspring, which turns the gears in the watch as it slowly uncoils. The weight and the mainspring have the same role—the source of power—but they have no obvious visual similarities. Many of these parts in the mechanical watch are like this—they have different names, look different, and do their jobs via different mechanical behaviors.

In this question, students were first asked to match the parts of the mechanical watch with the parts of a pendulum clock that served the same purpose. The parts of the mechanical watch were listed in one column, and the parts of the pendulum clock were listed in a second column. The instructions read "All of the parts in the mechanical clock have analogous parts in the mechanical watch. Draw lines connecting the parts from the mechanical clock to the analogous parts of the mechanical watch." Students were also provided with a labeled diagram of a mechanical watch (Figure 4.7), but not with a diagram of a pendulum clock.

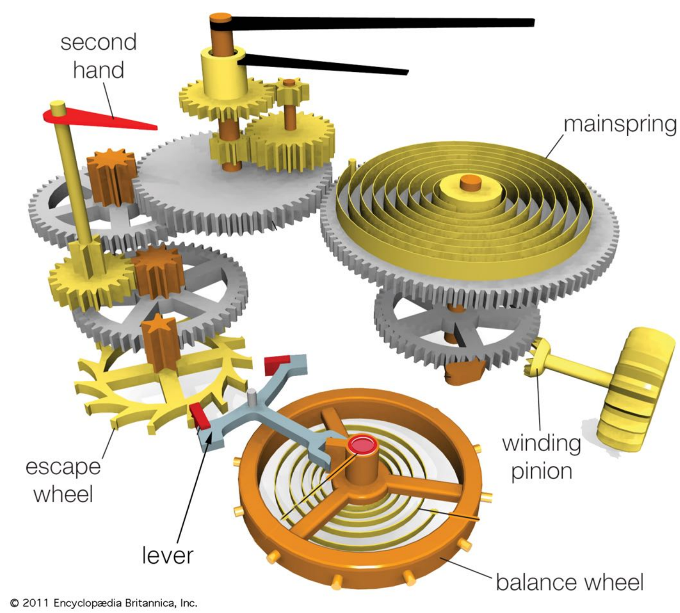


Figure 4.7: The diagram of the mechanical watch provided to students on the posttest.

After answering this question, students were asked to "Explain how you think this mechanical watch might work."

4.1.4.3.2 Questions about Previous Experience Four questions on the pretest asked about students' prior experience with programming, engineering, design, and using LEGO to build mechanisms and robots.

4.1.4.3.3 Mechanical Reasoning One question that was designed to measure mechanical reasoning ability was included at the beginning of the posttest. This question showed five meshed gears with an arrow indicating the direction of rotation for a single gear. The task was to determine which of the other gears would also turn in the same direction. This question was included to measure the distribution of mechanical reasoning ability between groups, so that it could be controlled for if the distribution was unequal.

4.1.4.3.4 Cognitive Reflection Questions (The CRT and CRT-2) Regardless of condition, learning about the pendulum clock required students to process and integrate a large amount of information. Additionally, correctly answering the posttest questions required students to understand what each question was asking, to consider all of the options, and to avoid simply providing the first answer that came to mind. These can be framed in terms of two types of cognitive processes called System 1 and System 2 (Kahneman, 2011). System 1 processes are executed with little conscious effort, and are considered to be automatic or spontaneous, while System 2 processes are slower and require more conscious effort (Stanovich & West, 2000). A tendency to rely on System 1 is known as miserly processing. As there was a possibility that students with a tendency towards miserly processing would receive lower scores on the posttest, it was important to measure this to ensure that there was not an unequal distribution of miserly processors across the conditions in the study. In this study, the Cognitive Reflection Test 2 (CRT-2) was used to measure miserly processing (Thomson & Oppenheimer, 2016).

The CRT-2 is based on the CRT (Frederick, 2005), a widely used and repeatedly validated measure of miserly processing. Low scores on the CRT are associated with a preference for miserly processing, and high scores are associated with a tendency to avoid miserly processing. The CRT consists of three questions that are designed to elicit an immediate, intuitive, and incorrect response. Reaching the correct answer requires suppressing System 1 responses in favor of System 2. The three questions on the CRT are as follows:

1. A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost? (intuitive answer: 10 cents; correct answer: 5 cents)
2. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? (intuitive answer: 100 minutes; correct answer: 5 minutes)
3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? (intuitive answer: 24 days; correct answer: 47 days)

The CRT has been validated against a number of other measures, including SAT scores (Frederick, 2005). However, scores on the CRT have been shown to be highly-correlated with numeracy (Thomson & Oppenheimer, 2016), which means that a person with high cognitive reflection ability but low numeracy skills would be likely to receive an inaccurate (too-low) score.

In pilot testing, it was found that a large number of participants were receiving very low scores on the CRT, but also exhibiting behaviors consistent with high cognitive reflection during the study. Many of the students who attempted to solve the problems using the intuitive, but incorrect method failed to perform correct calculations (e.g., failing to divide 48 by 2 correctly on the third question). This indicated that many students were struggling with the numerical nature of the problems on the CRT. For this reason, the CRT was replaced with the CRT-2 (Thomson & Oppenheimer, 2016), which was developed in part to address this problem. Like the original CRT, the CRT-2 consists of four questions that are designed to elicit an immediate, but incorrect response. In contrast to the CRT, these questions have little to no numerical content. The four questions on the CRT-2 are as follows:

1. If you're running a race and you pass the person in second place, what place are you in? (intuitive answer: first; correct answer: second)
2. A farmer had 15 sheep and all but 8 died. How many are left? (intuitive answer: 7; correct answer: 8)
3. Emily's father has three daughters. The first two are named April and May. What is the third daughter's name? (intuitive answer: June; correct answer: Emily)
4. How many cubic feet of dirt are there in a hole that is 3' deep x 3' wide x 3' long? (intuitive answer: 27; correct answer: none)

The fourth question was dropped based on results from pilot testing. Students responded to this fourth question in the same way they had responded to the three questions in the CRT, by attempting (and struggling) to perform numerical calculations. Students did not have similar issues with the first three questions, so they were included in the pretest.

4.1.4.3.5 Questions about Confidence with Mechanical Systems On the pretest, three questions were asked about students' general confidence in working with mechanical systems, and two questions were asked about their confidence in their ability to explain and in their understanding of how a pendulum clock worked. They were asked the same five questions on the posttest, but were also asked two questions about their confidence in their ability to explain how the mechanical watch worked and in their ability to figure out how it worked.

4.1.4.3.6 Interest in Mechanical Systems Two questions about interest in mechanical systems were included in the pretest, and the same two questions were included again in the posttest. There were two goals for these questions. The first goal was to measure the distribution of interest between groups before the study, so that it could be controlled for if the distribution was unequal. The second goal was to measure any change in interest due to taking part in the study.

4.1.4.3.7 Perceived Difficulty and Enjoyment of the Learning Phase Three questions about the perceived difficulty of the learning phase were included in the posttest. These were included to learn more about how the different learning phases were perceived, and to determine if perception of difficulty had any relationship to learning. One question about how much the participant enjoyed the learning-phase task was also included in the posttest.

4.1.4.4 The Hands-On Posttest: Repairing a Verge & Foliot Clock

The final question in the posttest was a hands-on problem. Students were presented with a partially-disassembled verge & foliot clock (Figure 4.8a) and asked to reconstruct it so that it worked properly. They were told that when the clock was correctly reassembled, the clock would tick and the hands move just like the pendulum clock.

Like the mechanical watch, the verge & foliot clock works analogously to the pendulum clock, though many of the parts look and act different. There were two major differences that students needed to account for in this task. The first was the difference in the oscillators, and the second was the difference in escapements.

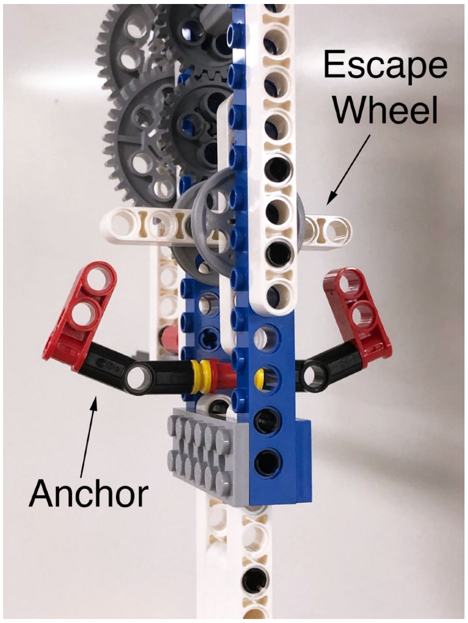
The oscillator is the part of the clock that keeps everything moving at a steady rate. In the pendulum clock the oscillator is the pendulum, while in the verge & foliot clock the oscillator is the foliot. The pendulum hangs down from the clock and swings side to side at a constant rate, while the foliot sits on top of the clock and wobbles back and forth. The pendulum has a single bob and the foliot has two bobs. Though these two parts have different shapes and behaviors, they both have the same purpose, which is to regulate the entire clock mechanism through their periodic oscillation.

The oscillator in the clock is connected to the escapement. The escapement is made up of two parts that mechanically interact with each other in order to regulate the rate of the

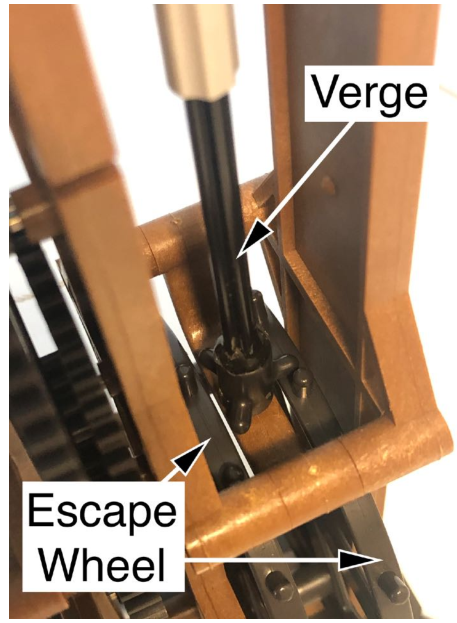
(a) The partially-disassembled verge & foliot clock that students were asked to repair

(b) The correctly-assembled verge & foliot clock

Figure 4.8: The verge & foliot clock students were asked to rebuild as part of the posttest.



(a) The escapement in the pendulum clock



(b) The escapement in the verge & foliot clock

Figure 4.9: Differences between the escapements of the pendulum clock and the verge & foliot clock

clock. In the pendulum clock the escapement is made up of the anchor and escape wheel, while in the verge & foliot clock the escapement is made up of two parts called the verge and the escape wheel. The anchor and the verge have the same purpose in both clocks, which is to catch and release successive teeth in the escape wheel, causing it to lock and unlock. However, the shapes of the anchor and the verge are different, and the ways that they interact with the escape wheel are also different. The anchor is large—nearly as wide as the escape wheel itself—and sits above or below the escape wheel (Figure 4.9a). The verge is small—not much bigger than one of the teeth on the escape wheel—and sits nestled in between the teeth of the escape wheel (Figure 4.9b).

Correctly reconstructing the verge & foliot clock required the students to attach the weight to a string and hang it from a pulley, to correctly insert the verge, to correctly build the foliot from the LEGO axles and wheels, and to correctly attach the foliot to the verge.

4.2 Findings

4.2.1 Pre-Analysis: Measuring Possible Confounds Between Conditions

Though the study employed a randomized experimental design, it was important to ensure that randomization resulted in an even distribution of possible confounding factors across conditions. For example, if students in one condition had a significantly higher GPA than students in the other conditions, this would be a potential confounding factor as students with higher GPAs might be expected to learn more, regardless of the intervention. To safeguard against this, data on potential confounding factors was collected from four sources. The first was a prescreen survey administered by the research-experience program to all the students in the pool at the start of the semester; the second was the video recorded during the study; the third was the pretest administered to students at the start of the study; and the fourth was the posttest administered at the end of the study.

The prescreen survey contained the bulk of data on potential confounds. The following variables in the prescreen data were chosen for analysis:

- GPA
- Age
- Gender
- Reading and writing ability
- English language competency
- Mother's level of education
- Socio-economic status (SES)
- Math background
- Engineering background
- Programming background
- Prior experience in makerspaces

Other variables of interest were learning phase duration (i.e., time-on-task), mechanical reasoning ability, and cognitive reflection ability (i.e., miserly processing). The video data was used to determine the amount of time each student spent in the learning phase, and the pretest contained students' answers to three cognitive-reflection questions from the CRT-2 designed to measure their miserly processing (Thomson & Oppenheimer, 2016). The pretest also contained questions about the students' prior experience in makerspaces and prior engineering experience, which were redundant with similar questions on the prescreen survey.

Finally, the posttest contained a single question designed to measure mechanical reasoning ability. The choice to include this question on the posttest instead of the pretest was made to keep the length of the pretest as short as possible, and was based on the assumption that students' mechanical reasoning ability was unlikely to change over the course of the study. The full list of variables as well as their source of data can be found in Table 4.4.

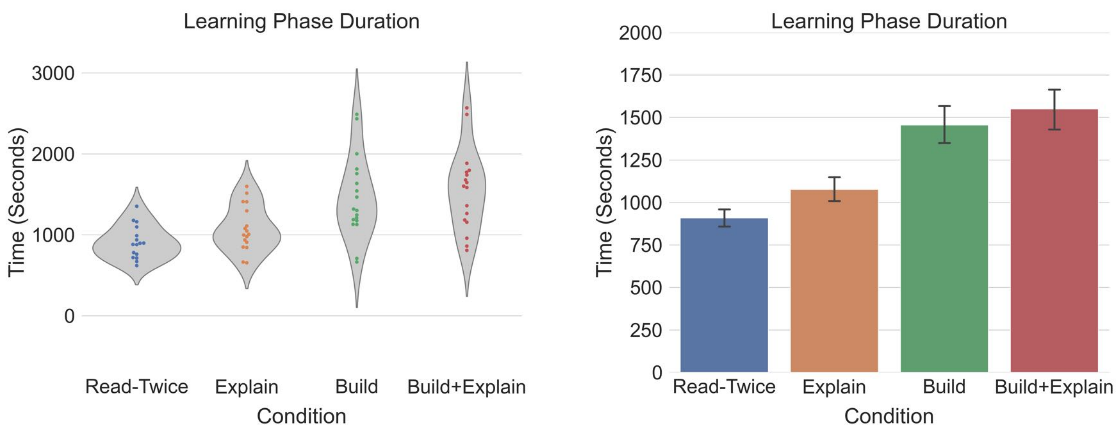
Table 4.4: Potential confounds in the study

Possible Confound	Source of Data	Differences Between Groups	Statistical Results
Learning phase duration	Video	**Yes**	$F(3, 65) = 10.80, p < 0.001$
Cognitive Reflection (CRT-2)	Pretest	No	$F(3, 65) = 0.20, p = 0.89$
Engineering background	Pretest	No	$F(3, 65) = 0.13, p = 0.93$
Prior experience in makerspaces	Pretest	No	$F(3, 65) = 0.51, p = 0.73$
Mechanical reasoning	Posttest	No	$\chi^2(3, N = 69) = 1.64, p = .65$
GPA	Prescreen	No	$F(3, 55) = 0.50, p = 0.68$
Age	Prescreen	No	$F(3, 65) = 1.87, p = 0.14$
Gender	Prescreen	No	$F(3, 65) = 0.26, p = 0.85$
Reading and writing ability	Prescreen	No	$F(3, 65) = 0.26, p = 0.85$
English level	Prescreen	No	$F(3, 65) = 0.16, p = 0.92$
Mother's level of education	Prescreen	No	$F(3, 65) = 0.82, p = 0.48$
SES	Prescreen	No	$F(3, 65) = 0.87, p = 0.46$
Math background	Prescreen	No	$F(3, 65) = 0.68, p = 0.57$
Engineering background	Prescreen	No	$F(3, 65) = 0.18, p = 0.91$
Prior experience in makerspaces	Prescreen	No	$F(3, 65) = 0.99, p = 0.40$
Programming background	Prescreen	No	$F(3, 65) = 1.26, p = 0.29$

To determine whether there were significant differences between groups, a one-way ANOVA was performed for each of the continuous variables and a chi-square test of independence was performed for each of the binary variables. Out of the 16 variables, only learning-phase duration was found to vary significantly between groups, $F(3, 65) = 10.80, p < 0.001$. The full statistical results for all tests are reported in Table 4.4.

Post-hoc comparisons using the Holm-Bonferroni correction revealed that the differences in time-on-task were primarily between (a) the Build and Build+Explain conditions and (b) the Read-Twice and Explain conditions, with the two Build conditions having significantly longer learning phases than the other two conditions (Figure 4.10a). There was no difference in learning-phase duration between the Build and Build+Explain groups ($M = 1456.316, SD = 501.690$ and $M = 1550.559, SD = 499.323$ respectively, $p = 0.58$), and there was a trending difference between durations of the Read-Twice and Explain groups ($M = 862.026, SD = 277.736$ and $M = 1077.111, SD = 280.625$ respectively, $p < 0.07$). However, there were significant differences between the Build group and the Explain group, $t(26.98) = -2.78, p < 0.03$, Cohen's $d = -0.93$; as well as between the Build group and the Read-Twice group, $t(26.82) = 4.37, p < 0.001$, Cohen's $d = 1.45$. Similarly, there were significant differences between the Build+Explain group and the Explain group, $t(32) = -3.41, p < 0.01$, Cohen's $d = -1.17$; and between the Build+Explain group and the Read-Twice group, $t(32) = 4.97, p < 0.001$, Cohen's $d = 1.70$. See Figure 4.10b for a graphical comparison of all four group means.

The data clearly shows that the building task—reconstructing the pendulum clock—required significantly more time than explaining the text or reading the text on the webpage twice. On average, students in the two Build conditions spent 25.03 minutes in the learning phase, while students in the other two conditions spent 16.16 minutes in the learning phase. Students in the Build conditions spent nearly ten more minutes in the learning phase than students in the other two conditions. Because of this, all subsequent analyses are conducted twice; once with no covariates, and once with learning-phase duration included as a covariate



(a) Distribution of time-on-task for students in each condition. Each point represents one student's time.

(b) Mean time-on-task for each group. Error bars show the standard error of the mean.

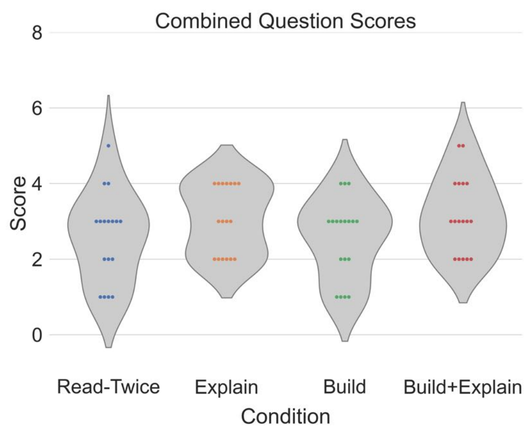
Figure 4.10: Amount of time spent in the learning phase for each condition.

in order to control for these differences.

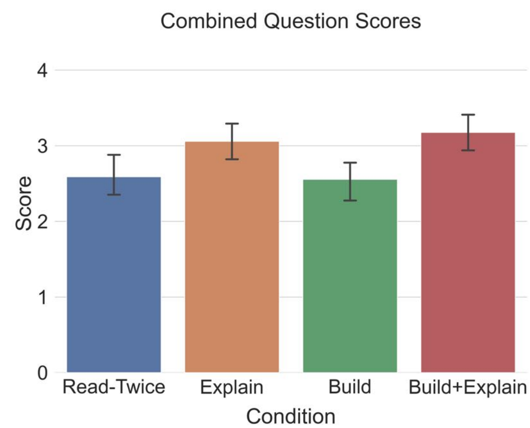
Although there were not significant differences between groups on the other variables, a select few were included in the covariate analyses as well. The variables that were selected were mechanical reasoning ability, cognitive-reflection ability, and GPA. Each of these variables have theoretical significance, in that there are plausible mechanisms linking scores on each to learning outcomes in the study. Including them in the subsequent analyses made it possible to investigate how these factors impacted learning about the mechanical clock, and to explore interactions between these factors and the different interventions. Additional information about each of these variables is included in the Additional Covariates section of the Appendix.

4.2.2 Verbatim and Inference Questions

Three multiple-choice questions on the posttest were designed to assess verbatim knowledge, defined as questions with answers that were explicitly written in the text or stated in the videos. There were also items on the posttest designed to assess students' ability to answer



(a) Distribution of Combined question scores for students between conditions. Each point represents one student's score.



(b) Mean Combined question scores for each group. Error bars show the standard error of the mean.

Figure 4.11: Combined question scores for each condition.

inference questions about the mechanical clock. Answering these questions required students to draw inferences that went beyond what was written explicitly in the text by integrating knowledge about different aspects of the mechanical clock. These six questions were either scored 1 (correct) or 0 (incorrect), and the scores across all six questions were summed to give a single Combined-Question score.

A one-way ANOVA was conducted to compare each condition's Combined-Question scores. No significant differences were found, $F(3, 65) = 1.63, p = 0.19$. The distribution of Combined scores can be found in Figure 4.11a, and the mean and standard deviation of Combined question scores can be found in Table 4.5 and Figure 4.11b.

Table 4.5: Combined question scores in the pendulum-clock study.

Condition	N	Mean	SD
Read-Twice	17	2.59	1.18
Explain	17	3.06	0.90
Build	17	2.56	1.04
Build+Explain	18	3.18	1.01

A two-way ANOVA was conducted to examine the main effects of (a) building and (b) explaining on the Combined question score. The main effect of explaining was significant, $F(1,65) = 4.79, p < 0.05$, indicating that students who explained during the learning phase (the Explain and Build+Explain conditions) scored significantly higher on the Combined questions than students who did not explain (the Read-Twice and Build conditions). The main effect of building during the learning was not significant (Figure 4.12).

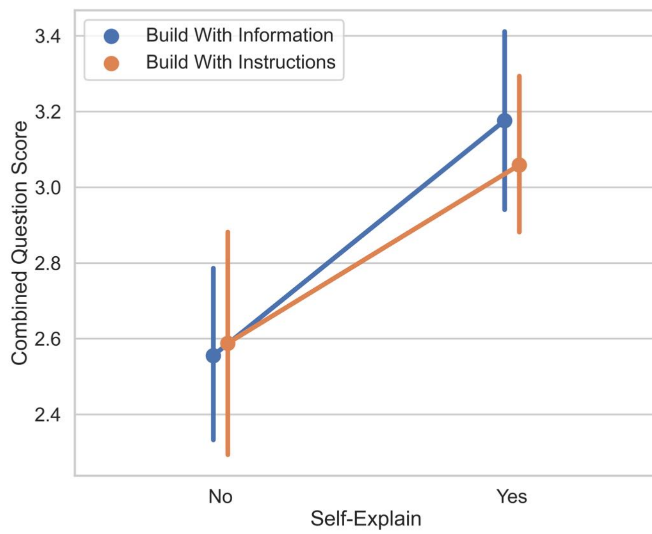


Figure 4.12: Two-way interaction plot showing the effects of building and explaining on the Combined question score.

A covariate analysis was conducted to investigate the effect of controlling for time-on-task, mechanical-reasoning scores, cognitive-reflection scores, and GPA on the Combined scores. First, a regression with time-on-task included as the sole covariate was conducted, as this was the only potential confound that differed significantly between groups. Next, a complete regression was conducted with all covariates of interest included. Finally, a stepwise regression was run in order to determine the best-fitting model with the smallest subset of independent variables that best predicted the Combined score. The stepAIC function in R using the forward-backward algorithm was used to perform each of these stepwise regressions. The results of the stepwise regression suggested that the best model was the one that contained time-on-task as the sole covariate.

When controlling for time-on-task, the coefficient on the Build+Explain group became significant at α=0.05, indicating that the Combined question scores for the Build+Explain group were significantly higher than those of the Read-Twice group (column 2 in Table 4.6). When all the covariates were included, none of the coefficients were significant (column 3 in Table 4.6).

Taken together, these findings are consistent with the conclusion that explaining during the learning phase had a positive impact on students' ability to answer a combination of verbatim and inference questions in the posttest.

4.2.2.1 Analyzing Verbatim and Inference Questions Separately

The Combined question score contained information from two distinct types of questions: Verbatim and Inference. The Verbatim questions and Inference questions were analyzed separately to determine whether the intervention effected these question types differentially.

The three Inference questions were summed to create an Inference question score. First, a one-way ANOVA was performed to test the treatments' effects on inference-question scores. No significant differences were found, $F(3, 65) = 0.60, p = 0.61$. Second, a two-way ANOVA was conducted to examine the main effects of (a) building and (b) explaining on the Inference question score. Neither of the main effects were significant. The distribution of inference-question scores can be found in Figure 4.13a, and the means and standard errors of inference-question scores can be found in Table 4.7 and Figure 4.13b.

A covariate analysis was conducted to investigate the effect of controlling for time-on-task, mechanical-reasoning scores, cognitive-reflection scores, and GPA on Inference question scores. First, a regression with time-on-task as the sole covariate was conducted, as this was the only potential confound that differed significantly between groups. Next, a complete regression was conducted with all covariates of interest included. Finally, a stepwise regression was run in order to determine the best-fitting model with the smallest subset of independent

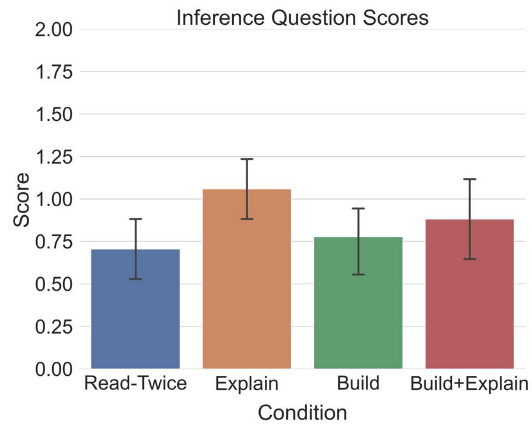
Table 4.6: Results of regression analyses of the Combined question scores. The Read-Only condition served as the baseline, and Time-on-Task, CRT-2, and GPA were mean-centered and standardized.

	Dependent variable:		
	Combined Question Score		
	(1)	(2)	(3)
Build	-0.03	0.27	-0.05
	(0.35)	(0.40)	(0.44)
Build+Explain	0.59	0.94**	0.60
	(0.36)	(0.42)	(0.47)
Explain	0.47	0.58	0.38
	(0.36)	(0.36)	(0.40)
Time-on-Task		-0.24	-0.15
		(0.15)	(0.17)
Mech. Reasoning			0.21
			(0.31)
CRT-2			-0.10
			(0.14)
GPA			-0.003
			(0.14)
Constant	2.59***	2.39***	2.37***
	(0.25)	(0.28)	(0.40)
Observations	69	68	59
R^2	0.07	0.11	0.08
Adjusted R^2	0.03	0.05	-0.04
Note:	*p<0.1; **p<0.05; ***p<0.01		

Table 4.7: Inference-question scores in the pendulum-clock study.

Condition	N	Mean	SD
Read-Twice	17	0.71	0.69
Explain	17	1.06	0.83
Build	17	0.78	0.81
Build+Explain	18	0.89	0.93

(a) Distribution of inference-question scores for students between conditions. Each point represents one student's score.



(b) Mean inference-question scores for each group. Error bars show the standard error of the mean.

Figure 4.13: Inference-question scores for each condition.

variables that best predicted Inference question scores. The results of the stepwise regression suggested that the best model was one with GPA as the sole independent variable.

Overall, there was no difference between conditions on ability to answer Inference questions, even when controlling for Time-on-Task and other covariates. There was a marginally-significant negative relationship between GPA and inference-question scores, though including this in the model did not affect the relationship between the condition and the outcome (Figure 4.14).

Next, the three Verbatim questions were summed to produce a Verbatim question score. First, a one-way ANOVA was conducted, and no significant differences were found, $F(3, 65) = 1.11, p = 0.35$. Second, a two-way ANOVA was conducted to examine the main effects of (a) building and (b) explaining on the Verbatim question score, and neither of the main effects were significant. The distribution of verbatim-question scores can be found in Figure 4.15a, and the means and standard errors of verbatim-question scores can be found in Table 4.9 and Figure 4.15b.

The same covariate-analysis procedure described in the previous section was used again

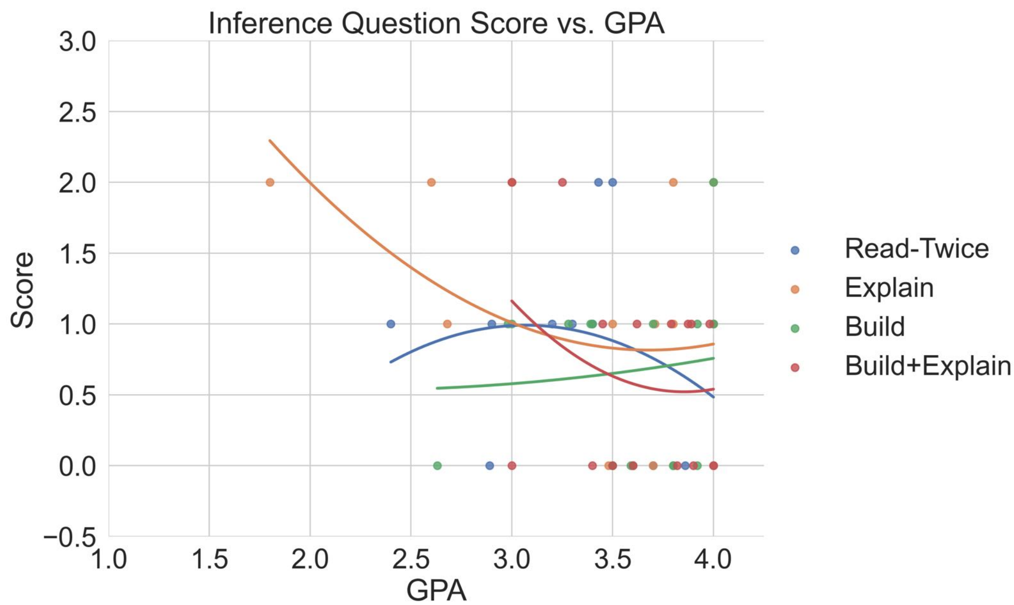
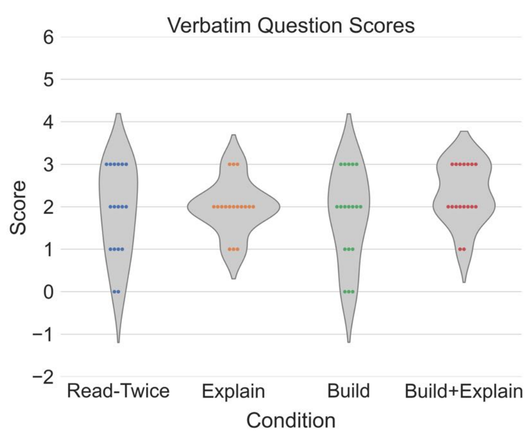
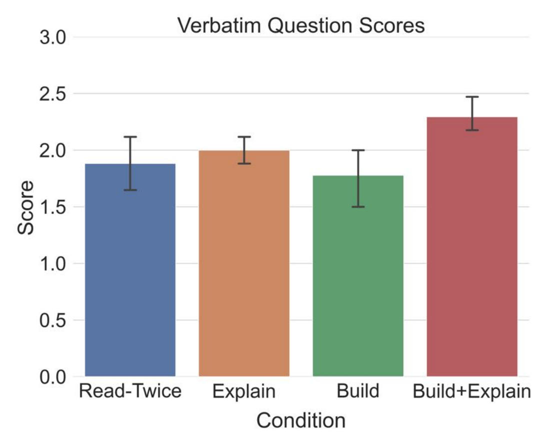


Figure 4.14: Relationship between GPA, inference-question scores, and experimental condition. Lines for each condition were fit using a second-order polynomial. Each point represents one student.



(a) Distribution of verbatim-question scores for students between conditions. Each point represents one student's score.



(b) Mean verbatim-question scores for each group. Error bars show the standard error of the mean.

Figure 4.15: Verbatim-question scores for each condition.

Table 4.8: Results of regression analyses of Inference question scores. The Read-Only condition served as the baseline. Time-on-Task, CRT-2, and GPA were mean-centered and standardized.

	Dependent variable:			
	Inference Question Score			
	(1)	(2)	(3)	(4)
Build	0.07	0.29	0.004	
	(0.28)	(0.31)	(0.31)	
Build+Explain	0.18	0.43	0.04	
	(0.28)	(0.33)	(0.33)	
Explain	0.35	0.43	0.23	
	(0.28)	(0.29)	(0.28)	
Time-on-Task		-0.17	-0.07	
		(0.12)	(0.12)	
Mech. Reasoning			0.04	
			(0.22)	
CRT-2			-0.08	
			(0.09)	
GPA			-0.17*	-0.18*
			(0.10)	(0.09)
Constant	0.71***	0.56**	0.69**	0.78***
	(0.20)	(0.22)	(0.28)	(0.09)
Observations	69	68	59	59
R^2	0.03	0.06	0.10	0.06
Adjusted R^2	-0.02	-0.002	-0.02	0.05
Note:	*p<0.1; **p<0.05; ***p<0.01			

Table 4.9: Verbatim-question scores in the pendulum-clock study.

Condition	N	Mean	SD
Read-Twice	17	1.88	1.05
Explain	17	2.00	0.61
Build	17	1.78	1.06
Build+Explain	18	2.29	0.69

here. Like the Inference question analysis, the results of the stepwise regression suggested that the best model was one with GPA as the sole independent variable. However, no significant relationships were found between any of the variables and the Verbatim question score in any of the models at $\alpha = 0.10$ (Table 4.10).

In summary, there were no significant relationships between the experimental conditions and the Verbatim or Inference question scores. This meant that the finding from the previous section, that there was a main effect of explaining on Combined question scores, was not due to the explaining intervention having a stronger effect on one type of question. Instead, these results indicate that explaining had a mild, statistically insignificant positive effect on each type of question, but that when both question types were combined the effect became significant.

4.2.3 The Draw-a-Pendulum-Clock Task

In this task students were asked to draw a diagram of a pendulum clock from memory. During the task, students could not view the webpage, nor could they see the LEGO model of the clock. Because of these restrictions, the quality and content of a drawing was a reflection of each student's mental representation of the clock as well as their ability to recall and render details of the clock from memory unprompted.

Students' drawings were coded twice, once for the presence of components and a second time for the presence of interactions between components. This resulted in two scales for each drawing, one for "Parts" and one for "Mechanism". A drawing would receive a perfect Parts score if all six components were clearly-rendered, and would receive a perfect Mechanism score if all five mechanical relations were present. For both scales, each item was scored out of two points. An unambiguous rendering was coded 2 out of 2, an ambiguous rendering was coded 1 out of 2, and no rendering was coded 0 out of 2. The maximum score on the Parts scale was 12, as there were six possible components that could be rendered; and the maximum score on the Mechanism scale was 10, as there were five possible mechanical

Table 4.10: Results of regression analyses of Verbatim question scores. The Read-Only condition served as the baseline. Time-on-Task, CRT-2, and GPA were mean-centered and standardized.

	Dependent variable:			
	Verbatim Question Score			
	(1)	(2)	(3)	(4)
Build	-0.10	-0.02	-0.05	
	(0.30)	(0.34)	(0.38)	
Build+Explain	0.41	0.51	0.56	
	(0.30)	(0.36)	(0.40)	
Explaining	0.12	0.15	0.15	
	(0.30)	(0.31)	(0.34)	
Time-on-Task		-0.07	-0.08	
		(0.13)	(0.14)	
Mech. Reasoning			0.18	
			(0.27)	
CRT-2			-0.03	
			(0.12)	
GPA			0.17	0.19
			(0.12)	(0.12)
Constant	1.88***	1.83***	1.68***	1.97***
	(0.21)	(0.24)	(0.34)	(0.11)
Observations	69	68	59	59
R^2	0.05	0.05	0.11	0.05
Adjusted R^2	0.005	-0.01	-0.01	0.03
Note:	*p<0.1; **p<0.05; ***p<0.01			

relations that could be rendered. After coding each drawing, the two scores were summed to create a third, "Combined" scale.

The following example helps illustrate the difference between a clearly-rendered and an ambiguously-rendered component. A clearly-rendered escape wheel would be visually distinct from the other gears in the clock, while an ambiguously-rendered escape wheel would have the same size and shape as the other gears. Two examples of clearly-rendered escape wheels can be seen in Figure 4.16a and Figure 4.16b. In both of these drawings, the escape wheels are shaped like crosses, sit in the center of the clock, and are visually distinct from the other gears. Both of these students were given two points for their rendering of the escape wheel. An example of an ambiguously-rendered escape wheel can be seen in Figure 4.16d. In this drawing, the escape wheel and other gears are all rendered as circles, and are visually indistinguishable from one another. The student who made this drawing was given one point out of two for their rendering of the escape wheel. Complete details of the coding scheme used to analyze the drawings can be found in the section Coding Schemes for the Draw-a-Pendulum-Clock Task in the Appendix.

The drawings that students produced were highly variable, receiving a wide range of scores on both the Parts and Mechanism scales. Figure 4.16a shows an example of a drawing that received a perfect score on both scales. In this drawing the student clearly rendered all six components—the weight, the gears, the escape wheel, the anchor, the pendulum, and the hand—and also clearly rendered all five mechanical relations between parts—the connection between the weight and gears, the connection between the gears and the escape wheel, the interaction between the escape wheel and the anchor, the connection between the anchor and the pendulum, and the connection between the gears and the hand. The drawing shown in Figure 4.16b received a perfect score on the Parts scale, but received a medium score on the Mechanism scale. The reason for the lower Mechanism score is that many mechanical connections, such as the connection between the pendulum and the anchor, are missing or ambiguous. Finally, the drawing shown in Figure 4.16c received a medium score on the

Parts scale and a zero score on the Mechanism scale, as none of the mechanical connections rendered in the drawing were correct.

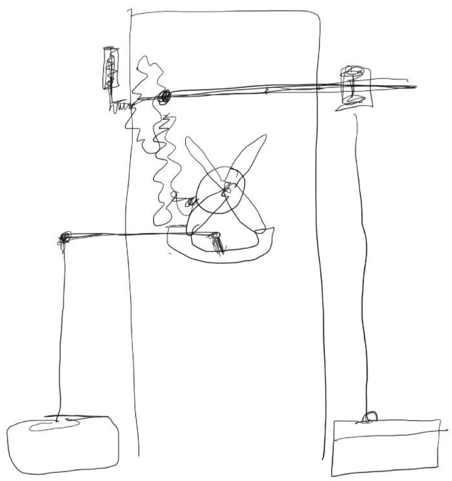
A one-way ANOVA was performed to compare students' scores on the Combined drawing scale (Mechanism + Parts score) between experimental conditions. There was a significant effect of condition on the Combined score, $F(3, 64) = 5.45, p < 0.01$. A follow-up two-way ANOVA was conducted to examine the main effects of (a) building and (b) explaining on the Combined drawing scores. The main effect of building was significant, $F(1, 64) = 11.68, p < 0.01$, indicating that students who rebuilt the clock during the learning phase (the Build and Build+Explain conditions) scored significantly higher on the Combined drawing scale than students who did not build (the Read-Twice and Explain conditions) (Figure 4.17).

Post-hoc analyses corrected for multiple comparisons using the Holm-Bonferroni correction indicated that the Build+Explain group scored significantly higher on the Combined drawing scale than the Explain or Read-Twice groups, $t(32) = -3.09, p < 0.05$ and $t(32) = 3.79, p < 0.01$, respectively. There was not a significant difference between the Build+Explain group and the Build group, $t(31.63) = -2.10, p < 0.17$. See Table 4.11 for means and standard deviations for each condition. The distribution of Combined drawing scores can be seen in Figure 4.18a and the means and standard errors can be seen in Figure 4.18b.

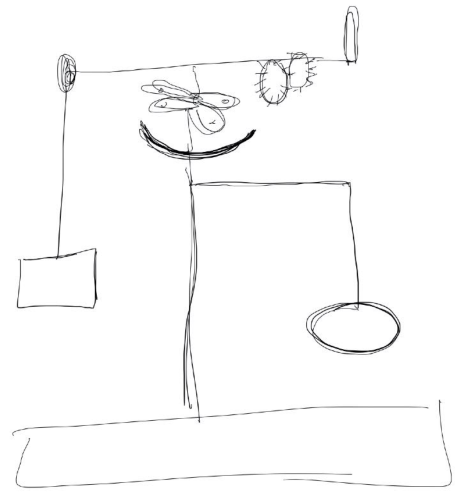
Table 4.11: Mean and standard deviations of Combined scores in the Draw-a-Pendulum-Task.

Condition	N	Mean	SD
Read-Twice	17	12.59	3.26
Explain	17	13.25	4.02
Build	17	14.67	3.32
Build+Explain	18	17.23	3.87

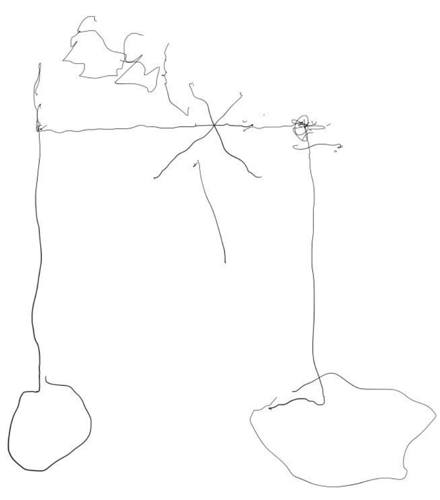
A covariate analysis was conducted to investigate the effect of controlling for time-on-task, mechanical-reasoning scores, cognitive-reflection scores, and GPA on the Combined drawing scores. First, a regression with time-on-task as the sole covariate was conducted,



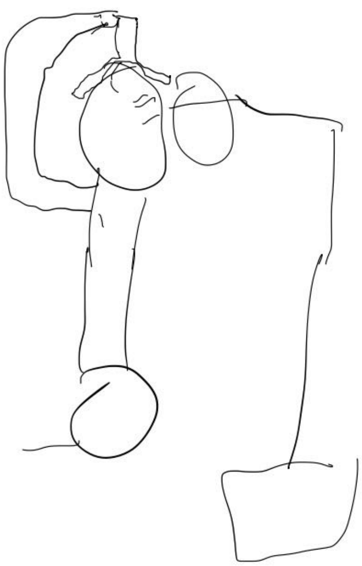
(a) This drawing received a perfect score for Parts (12/12) and Organization (10/10).



(b) This drawing received a perfect score for Parts (12/12) and low score for Organization (5/10).



(c) This drawing received a low score for Parts (7/12) and a zero for Organization (0/10).



(d) This drawing received a low score for Parts (7/12) and a low score for Organization (4/10).

Figure 4.16: Example drawings created by four different students during the posttest.

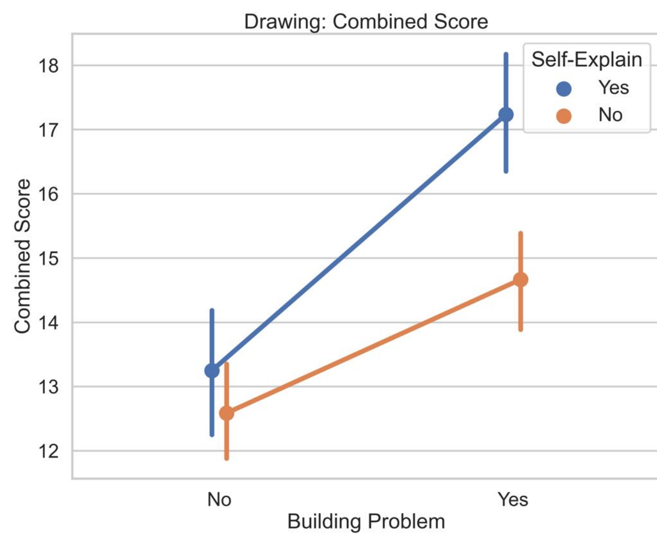
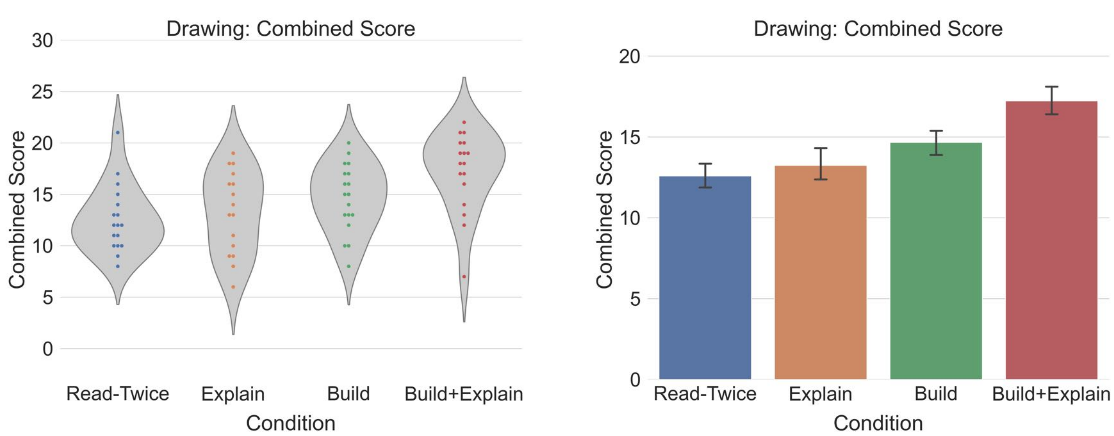


Figure 4.17: Two-way interaction plot showing the effects of building and explaining on the Combined drawing score.



(a) Distribution of Combined scores for the Draw-a-Pendulum-Clock task.

(b) Mean Combined scores for the Draw-a-Pendulum-Clock task. Error bars show the standard error of the mean.

Figure 4.18: Combined scores for the Draw-a-Pendulum-Clock task.

as this was the only potential confound that differed significantly between groups. Next, a complete regression was conducted with all covariates of interest included. Finally, a stepwise regression was run in order to determine the best-fitting model with the smallest subset of independent variables that best predicted Combined scores. The results of the stepwise regression suggested that the best model contained two covariates: Time-on-Task and Mechanical Reasoning Ability.

In the Time-on-Task model, the coefficients on the Build and Build+Explain conditions were both significant at α=0.05 (column 2 of Table 4.11). This was a somewhat surprising finding, as both of these groups spent significantly more time in the learning phase than the Explain or Read-Twice groups. The value of the coefficient on Time-on-Task was -0.92, indicating that for every standard-deviation increase in the duration of the learning phase (472 seconds), there was a corresponding 0.92 point decrease in the Combined drawing score. Figure 4.19 provides a visualization of this relationship. The students who spent the most time in the learning phase received some of the lowest scores, and these students were all in the Build and Build+Explain groups. However, when comparing students between conditions who spent similar amounts of time in the learning phase, the students in the Build and Build+Explain conditions outperformed students in the other groups.

The model identified as providing the best fit by the stepwise regression contained Time-on-Task and Mechanical Reasoning Ability as covariates. In this model, the coefficients on Build, Build+Explain, and Mechanical Reasoning Ability were all significant at α<0.05, and the coefficient on Time-on-Task was significant at α=0.10 or better (column 4 of Table 4.11). The coefficient on Mechanical Reasoning was 2.65 and significant at α=0.01, indicating that students with high mechanical reasoning ability scored 2.65 points higher on the Combined drawing scale than students with low mechanical reasoning ability (Figure 4.20). In this mode, the coefficients on the Build and Build+Explain groups both remained significant, and the coefficient on the Explain group was not significant.

Taken together, these analyses were all consistent with the conclusion that there was a

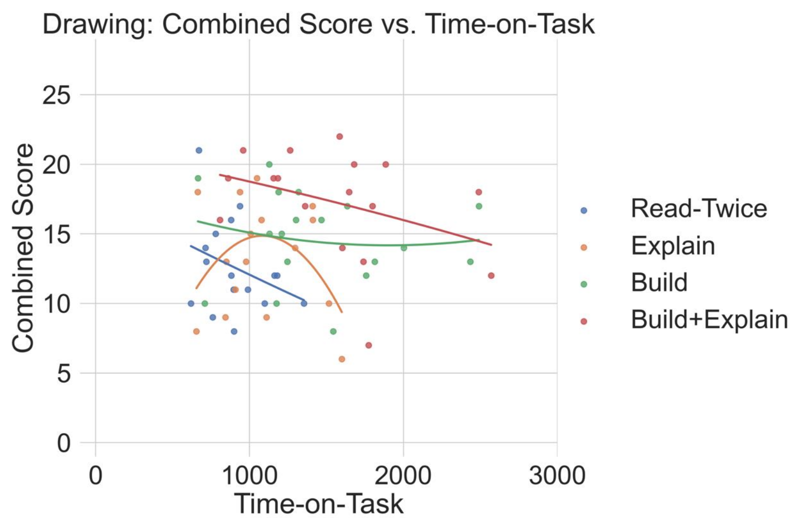


Figure 4.19: Relationship between Time-on-Task, Combined drawing scores, and experimental condition. Lines for each condition were fit using a second-order polynomial. Each point represents one student.

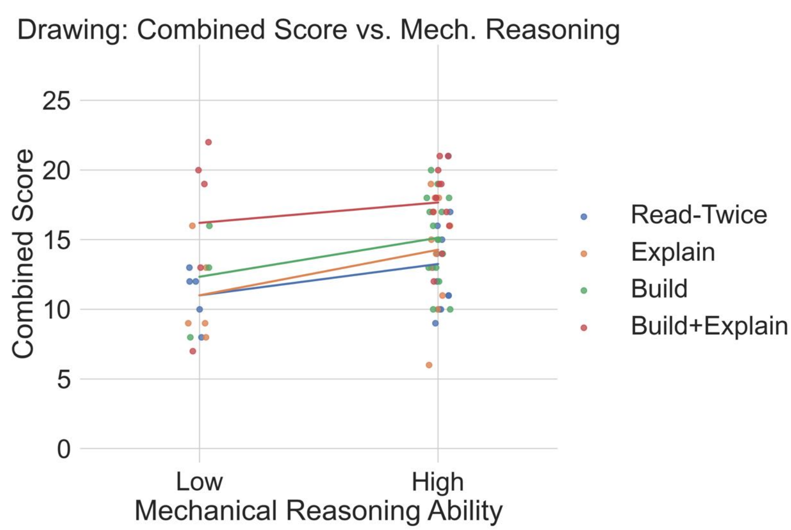


Figure 4.20: Relationship between Mechanical Reasoning Ability, Combined drawing scores, and experimental condition. Each point represents one student.

significant main effect of building on the students' free recall of the pendulum clock mechanism. When controlling for Time-on-Task, this effect became even stronger, with both the Build and Build+Explain groups scoring significantly higher on the Combined drawing score than the Read-Twice group. Of particular note is the Build+Explain group, which achieved significantly higher Combined drawing scores than all other groups, regardless of whether the analysis controlled for possible confounds. This indicated that the combination of building and explaining had the strongest effect overall on students' ability to remember accurate and complete representations of the pendulum clock's parts and mechanism.

4.2.3.1 Individual Analyses of the Mechanism and Parts Scales

A set of follow-up analyses were run to determine whether the effects of building on the Combined drawing scores were due to improved ability to recall and draw (a) the parts of the clock, (b) the mechanical relations between them, or (c) both. The Combined drawing scale was pulled apart into two scales, one for Parts and a second for Mechanism.

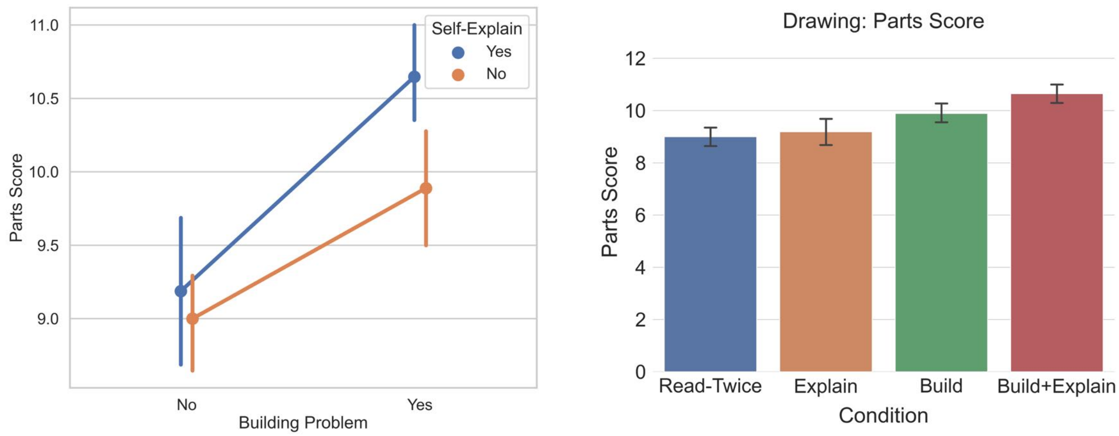
First, a one-way ANOVA was performed to compare students' scores on the Parts drawing scale between experimental conditions. There was a significant effect of condition on the Combined score, $F(3, 64) = 3.40, p < 0.05$. A follow-up two-way ANOVA was conducted to examine the main effects of (a) building and (b) explaining on the Parts drawing scores, and the main effect of building was significant, $F(1, 64) = 8.29, p < 0.01$. This indicated that students who rebuilt the clock during the learning phase (the Build and Build+Explain conditions) scored significantly higher on the Parts drawing scale than students who did not build (the Read-Twice and Explain conditions) (Figure 4.21a).

In post-hoc tests using the Holm-Bonferroni correction, the only significant difference was between the Build+Explain group and the Read-Twice group, $t(32) = 3.34, p < 0.05$ (Figure 4.21b).

Next, a one-way ANOVA was conducted to compare scores on the Mechanism drawing scale between conditions. There was a significant difference between conditions, $F3, 64) =$

Table 4.12: Results of regression analyses of Combined scores on the Draw-a-Pendulum-Clock task. The Read-Only condition served as the baseline. Time-on-Task, CRT-2, and GPA were mean-centered and standardized.

	Dependent variable:			
	Combined Drawing Score			
	(1)	(2)	(3)	(4)
Build	2.08*	3.17**	3.14**	3.05**
	(1.23)	(1.38)	(1.49)	(1.31)
Build+Explain	4.65***	5.93***	6.50***	6.15***
	(1.24)	(1.45)	(1.58)	(1.38)
Explain	0.66	1.03	0.83	1.22
	(1.26)	(1.29)	(1.36)	(1.23)
Time-on-Task		-0.92*	-1.10*	-1.01*
		(0.53)	(0.57)	(0.51)
Mech. Reasoning			2.61**	2.65***
			(1.07)	(0.97)
CRT-2			0.43	
			(0.46)	
GPA			-0.02	
			(0.48)	
Constant	12.59***	11.88***	9.81***	9.83***
	(0.88)	(0.98)	(1.36)	(1.20)
Observations	68	67	58	67
R^2	0.20	0.24	0.35	0.32
Adjusted R^2	0.17	0.19	0.26	0.27
Note:		*p<0.1; **p<0.05; ***p<0.01		



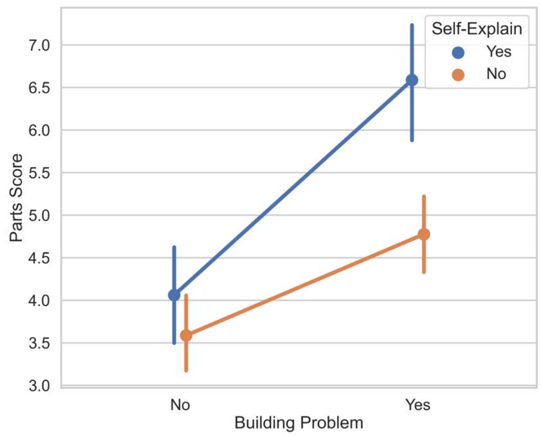
(a) Two-way interaction plot showing the effects of building and explaining on the Parts drawing score.

(b) Mean Parts scores for the Draw-a-Pendulum-Clock task. Error bars show the standard error of the mean.

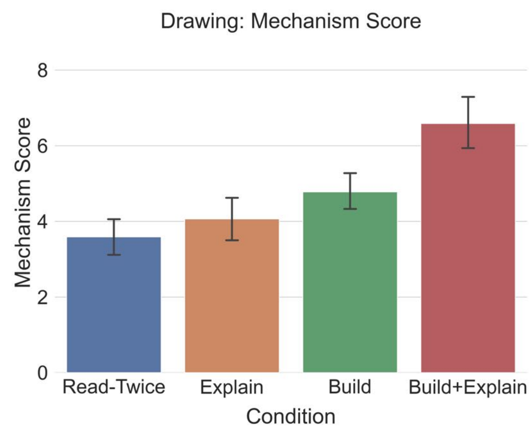
Figure 4.21: Parts scores for the Draw-a-Pendulum-Clock task.

$5.30, p < 0.01$. To test for main effects of building or explaining a two-way ANOVA was performed. The main effect of building on the Mechanism score was significant, $F(1, 64) = 10.38, p < 0.01$, as was the main effect of explaining, $F(1, 64) = 4.14, p < 0.05$. A post-hoc test indicated that both of these main effects were being driven by the Build+Explain condition. The Build+Explain condition scored significantly higher on the Mechanism drawing score than the Read-Twice group, $t(32) = 3.48, p < 0.01$, and scored marginally-significantly higher than the Explain group, $t(30.35) = -2.72, p < 0.06$ (Figure 4.22).

Across both the Parts and Mechanism dimensions of the Drawing scale, the story was the same: the groups who reconstructed the clock during the learning phase (the Build and Build+Explain groups) made more complete, accurate drawings. Their drawings contained more parts that were clearly rendered, and the mechanical connections and interactions between those parts were also more accurately depicted. Thus, reconstructing the clock during the learning phase facilitated the construction of more accurate, complete, and robust mental models.



(a) Two-way interaction plot showing the effects of building and explaining on the Mechanism drawing score.



(b) Mean Mechanism scores for the Draw-a-Pendulum-Clock task. Error bars show the standard error of the mean.

Figure 4.22: Mechanism scores for the Draw-a-Pendulum-Clock task.

4.2.4 Assessing Mechanistic Understanding with the Explain-a-Pendulum-Clock Task

In the Pendulum-Clock Explanation task, students were presented with a novel, unlabeled diagram of a pendulum clock (Figure 4.6) and asked to explain how it worked in their own words. These explanations were video recorded, and the recordings of these explanations were coded along seven dimensions. Four of these dimensions captured types of correct statements, and three captured types of misconceptions. The four types of correct statements were correctly identifying parts, correctly stating the purpose of a part in the clock, correctly stating a direct causal relation between parts, and correctly stating an indirect causal relation between parts. The three types of incorrect statements were misconceptions about a part's purpose, misconceptions about the direct causal relations between parts, and misconceptions about the indirect causal relations between parts.

This coding scheme was based on a checklist provided by Craver (2006) for assessing the quality of mechanistic explanations. According to Craver, a complete explanation must describe four things: the phenomenon, the parts, the activities, and the organization. Table 4.13 shows how the current coding scheme can be mapped onto this checklist as well as the similar structure-behavior-function (SBF) framework.

Table 4.13: Mapping the coding scheme used to code students' explanations of the pendulum clock onto the checklist provided by Craver as well as the widely-used structure-behavior-function (SBF) framework.

Explanation Coding Scheme	Craver (2006)	SBF
Parts	Parts	Structure
Purpose	Phenomenon	Function
Direct Cause	Activity/Organization	Behavior
Indirect Cause	Activity	Behavior

In the process of coding the explanations the need for a refinement to the activity code emerged. A simple definition of "activity" provided by Craver is "the things that entities do... Activities are the causal components of mechanisms" (2006, p. 371). However, statements having to do with the causal relations between parts could not be coded into a single activity category. Students were found to make two distinct types of causal statements when explaining how the clock worked: direct causal statements and indirect causal statements. The distinction between direct and indirect causal statements had to do with the type of interaction described. Direct causal statements described how two parts directly or physically interacted with one another, while indirect causal statements described how two parts were related without mentioning the direct interactions between them. For example, the statement "as the weight falls, it rotates the axle which turns the gears" would be coded as a direct causal statement, while the statement "the weight falling causes the pendulum to swing" would be coded as an indirect causal statement. Both of these statements are correct, but direct causal statements indicate a different type of understanding—one that is more precise and involves an immediate, interactive relationship between two entities. For this reason, direct causal statements and indirect causal statements were coded along separate dimensions. The full coding schemes are provided in the Appendix.

The Parts, Purpose, Direct Cause, and Indirect Cause scores were summed to create a single Combined explanation score. A one-way ANOVA was performed to investigate differences between groups on this score. There was a significant effect of condition on the Combined score, $F(3,65) = 2.89, p < 0.05$. A follow-up two-way ANOVA revealed a significant main effect of explaining on the Combined score, $F(1,65) = 5.21, p < 0.05$. The interaction plot showed that this main effect appeared to be primarily driven by the Build+Explain group (Figure 4.23).

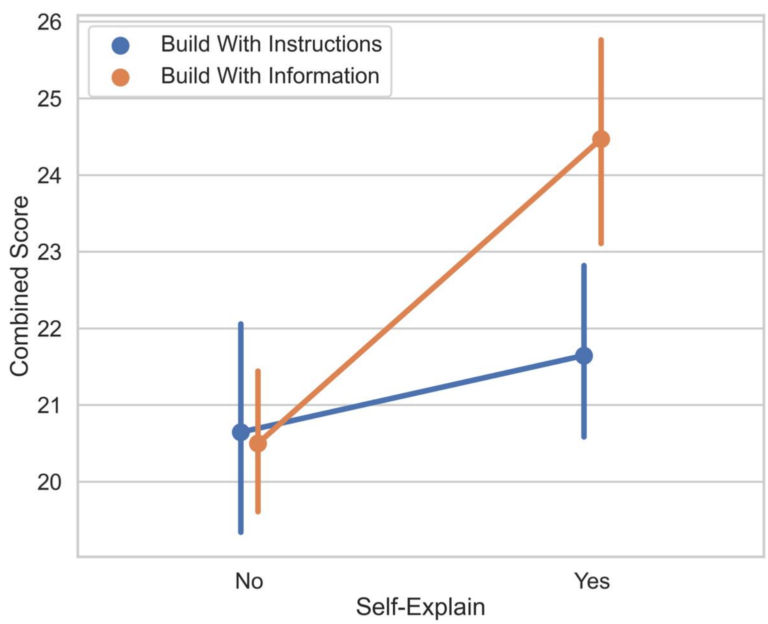
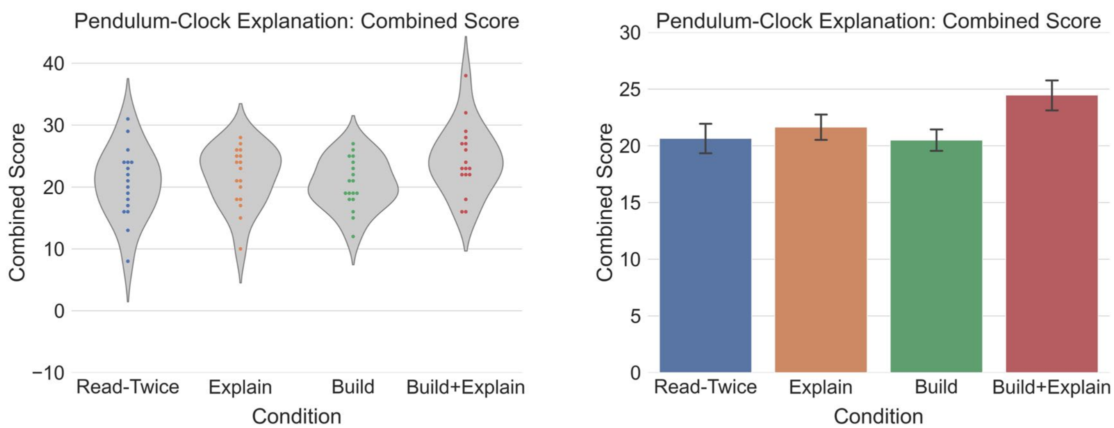


Figure 4.23: Two-way interaction plot showing the effects of building and explaining on the Combined score for the Explain-a-Pendulum-Clock task.

To investigate this further post-hoc tests using the Holm-Bonferroni correction were conducted to compare pairwise means between groups. The only significant difference was between the Build and Build+Explain groups, $t(31.35) = 2.89, p < 0.05$, though there were trending differences between the Build+Explain group and the other two groups as well. Means and standard deviations for each group can be found in Table 4.14. The distribution of Combined scores can be seen in Figure 4.24a and the means and standard errors can be



(a) Distribution of Combined scores on the Explain-a-Pendulum-Clock task.

(b) Mean Combined scores for the Explain-a-Pendulum-Clock task. Error bars show the standard error of the mean.

Figure 4.24: Combined scores for the Explain-a-Pendulum-Clock task.

seen in Figure 4.24b.

Table 4.14: Mean and standard deviations of Combined scores on the Explain-a-Pendulum-Clock Task.

Condition	N	Mean	SD
Read-Twice	17	21.76	5.88
Explain	17	22.82	4.59
Build	17	22.67	3.91
Build+Explain	18	25.88	4.65

A covariate analysis was conducted to explore these effects while controlling for potential confounds. This analysis was conducted via a series of linear regressions. The first regression model contained one covariate—Time-on-Task—as this was the only potential confound that differed significantly between groups. Next, a regression model containing all covariates was conducted. Finally, a stepwise regression was run in order to determine the best-fitting model with the smallest subset of independent variables that best predicted the Combined score. Similar to the previous covariate analyses, the results of the stepwise regression suggested that the best model contained two covariates: Time-on-Task and Mechanical Reasoning

Ability.

Across all models, the coefficient on the Build+Explain group remained positive and significant at $\alpha = 0.01$, and the coefficient on Time-on-Task remained negative and significant at α=0.01. Figure 4.25 shows the relationship between Time-on-Task, condition, and Combined scores. The coefficient on the Build+Explain group in the best-fitting model (column 4 of Table 4.15) was 6.65, indicating that when controlling for Time-on-Task and Mechanical Reasoning Ability the Build+Explain group scored 6.65 points higher on the Combined score than the baseline group (Read-Twice). When setting the Explain group as the baseline, the coefficient dropped to 2.870; and when setting the Build group as the baseline the coefficient dropped to 3.12. However, in both of these cases the coefficient on Build+Explain remained significant at α=0.01. Taken together, this meant that when controlling for possible confounds, the Build+Explain group scored significantly higher on the Explain-a-Pendulum-Clock task than any of the other groups.

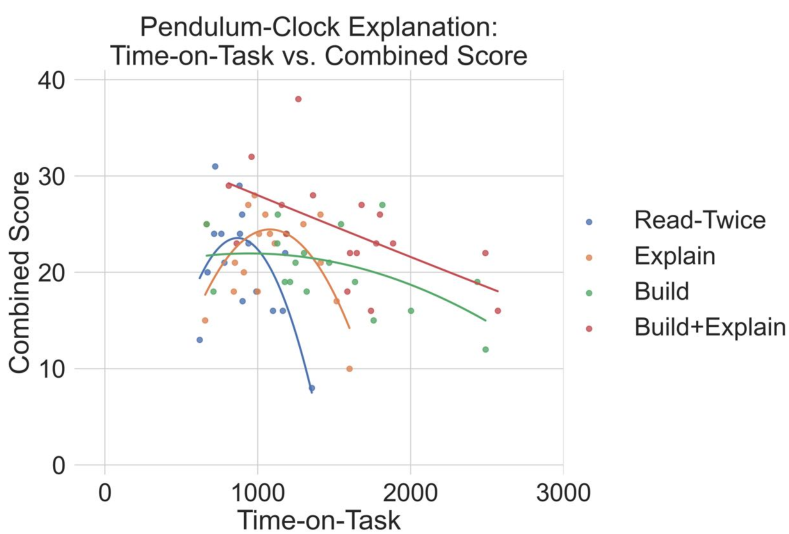


Figure 4.25: Relationship between Time-on-Task, Combined scores on the Explain-a-Pendulum-Clock task, and experimental condition. Lines for each condition were fit using a second-order polynomial. Each point represents one student.

All of the analyses of the Explain-a-Pendulum-Clock task were consistent. There was a positive main effect of explaining during the learning phase on the quality of the students' explanations of the pendulum clock in the posttest. This may seem tautological (students who self-explained during the learning phase performed better on the explanation question on the posttest), but it is important to note that none of the students in either of the two explain conditions (Explain and Build+Explain) generated a complete explanation of entire clock during the learning phase. Even so, it did make sense that students who explained individual components during the learning phase would score higher when asked to explain the entire clock as a single system. Further analyses revealed that this main effect was largely driven by the Build+Explain group. Students who explained after they reconstructed the clock (the Build+Explain group) generated explanations of the pendulum clock that demonstrated deeper mechanistic understanding than the other groups. This indicated that the combination of building and explaining had the strongest effect overall on students' understanding of the pendulum clock mechanism.

4.2.4.1 Individual Analyses of the Parts, Purpose, and Direct Cause Scales

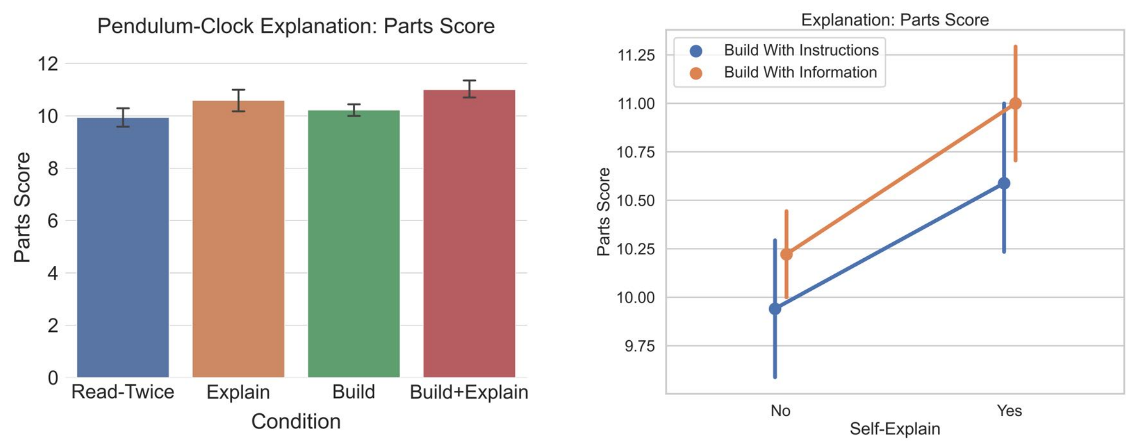
The previous analysis combined all four dimensions of mechanistic knowledge together into a single, combined score. However, one of the questions this study was designed to answer was how different treatments might impact specific dimensions of mechanistic knowledge. For this reason, individual analyses were conducted for each of the four scales: Parts, Purpose, Direct Cause, and Indirect Cause.

The only significant effect that was found was a main effect of self-explaining on knowledge about Parts in the Pendulum Clock. A one-way ANOVA was performed to investigate differences between groups on the Parts score. There was not a significant effect of condition on the this score, $F(3, 65) = 1.82, p = 0.15$. However, a follow-up two-way ANOVA revealed a significant main effect of explaining on the Parts score, $F(1, 65) = 4.44, p < 0.05$. Additional details are provided in Figure 4.26.

There were two trending effects as well. First, there was a trending main effect of

Table 4.15: Results of regression analyses of Combined scores on the Explain-a-Pendulum-Clock Task. The Read-Only condition served as the baseline. Time-on-Task, CRT-2, and GPA were mean-centered and standardized.

	Dependent variable:			
	Combined Clock Explanation core			
	(1)	(2)	(3)	(4)
Build	-0.15	2.72	3.26	2.61
	(1.72)	(1.80)	(2.04)	(1.76)
Build+Explain	3.82**	7.20***	8.20***	7.42***
	(1.75)	(1.88)	(2.16)	(1.85)
Explain	1.00	1.81	2.30	2.09
	(1.75)	(1.65)	(1.83)	(1.62)
Time-on-Task		-2.55***	-2.67***	-2.64***
		(0.69)	(0.77)	(0.68)
Mech. Reasoning			3.00**	2.46*
			(1.43)	(1.28)
CRT-2			0.47	
			(0.67)	
GPA			-0.71	
			(0.65)	
Constant	20.65***	18.86***	15.41***	16.96***
	(1.23)	(1.28)	(2.10)	(1.59)
Observations	69	68	59	68
R^2	0.09	0.25	0.32	0.30
Adjusted R^2	0.05	0.21	0.23	0.24
Note:		*p<0.1; **p<0.05; ***p<0.01		



(a) Mean Parts scores for the Explain-a-Pendulum-Clock task. Error bars show the standard error of the mean.

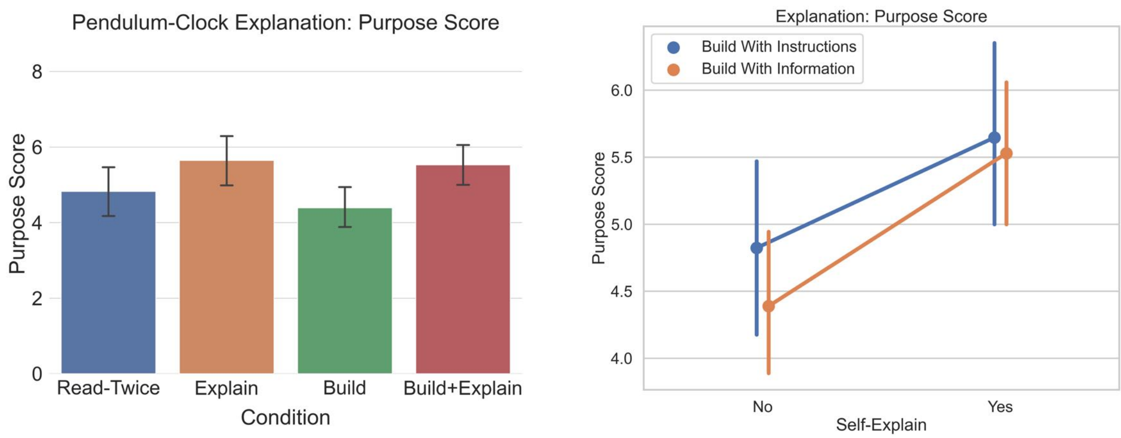
(b) Two-way interaction plot showing the effects of building and explaining on the Parts score for the Explain-a-Pendulum-Clock task.

Figure 4.26: Parts scores for the Explain-a-Pendulum-Clock task.

explaining on knowledge about the Purpose of components. Second, there was a combined effect (Build+Explain) on the Direct Cause score when compared to all other conditions. The reason for reporting these trends is that they became significant on the second explanation question, where students were asked to transfer their knowledge to explain how a mechanical watch worked.

The first trend was that students who self-explained scored higher on the Purpose score. The results of the one-way ANOVA were not significant $F(3, 65) = 1.82, p = 0.41$, and there was a trending main effect of Explaining $F(1, 65) = 2.57, p = 0.11$. Although these differences were not statistically significant, the trend between self-explaining and understanding of Purpose was clearly visible in the data visualizations (Figure 4.27).

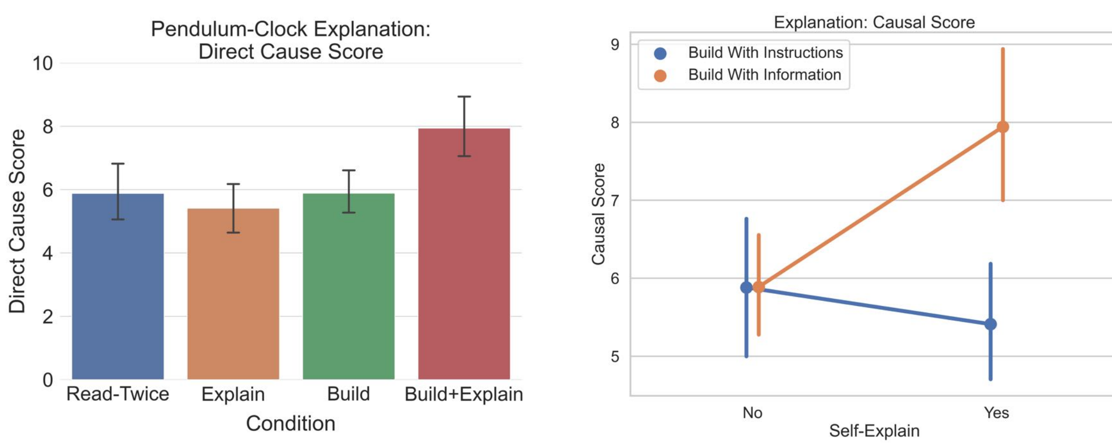
The second trend was that students in the Build+Explain group scored higher on the Direct Cause scale. The results of the one-way ANOVA were not significant $F(3, 65) = 1.37, p = 0.25$. However, the trend between building then self-explaining on the Direct Cause scale was apparent in the plots (Figure 4.28).



(a) Mean Purpose scores for the Explain-a-Pendulum-Clock task. Error bars show the standard error of the mean.

(b) Two-way interaction plot showing the effects of building and explaining on the Purpose score for the Explain-a-Pendulum-Clock task.

Figure 4.27: Purpose scores for the Explain-a-Pendulum-Clock task.



(a) Mean Direct Cause scores for the Explain-a-Pendulum-Clock task. Error bars show the standard error of the mean.

(b) Two-way interaction plot showing the effects of building and explaining on the Direct Cause score for the Explain-a-Pendulum-Clock task.

Figure 4.28: Direct Cause scores for the Explain-a-Pendulum-Clock task.

4.2.5 Assessing Transfer of Mechanistic Understanding with the Explain-a-Mechanical-Watch Task

In the mechanical-watch explanation task, students were presented with a labeled diagram of a mechanical watch (Figure 4.7) and asked to explain how it worked in their own words. At this point in the study the students had only learned about the pendulum clock, but had not seen any diagrams of a mechanical watch, nor had they learned about the mechanical watch (or any other type of mechanical clock for that matter). In order to explain how the mechanical watch worked, students had to transfer their knowledge of the pendulum clock to this similar, but novel system.

Students' explanations were video recorded, and these recording were coded along six dimensions. These were the same six dimensions used to code the pendulum-clock explanations, minus the Parts dimension. The Parts dimension was not coded because the diagram was clearly labeled with the names of each part. Three of the dimensions captured the types of correct statements students made during their explanations, and the remaining three dimensions captured the types of incorrect statements (i.e., misconceptions) students made. The three correct dimensions were Direct Cause, Indirect Cause, and Purpose; and the three incorrect dimensions were Direct Cause Misconceptions, Indirect Cause Misconceptions, and Purpose Misconceptions. The same distinction between direct causal and indirect causal statements used in the Pendulum-Clock Explanation Task was also used here.

The three correct dimensions were summed to create a single Combined score for each student, and a one-way ANOVA was performed to compare groups on this outcome. There was a significant effect of condition on the Combined score, $F(3, 65) = 4.91, p < 0.01$. A follow-up two-way ANOVA was then conducted, which found a significant main effect for explaining $F(1, 65) = 7.78, p < 0.01$, a marginally-significant main effect for building, $F(1, 65) = 3.79, p < 0.06$, and a marginally-significant interaction $F(1, 65) = 3.29, p < 0.08$. The interaction plot suggested that all three of these effects were driven by the Build+Explain group

(Figure 4.29).

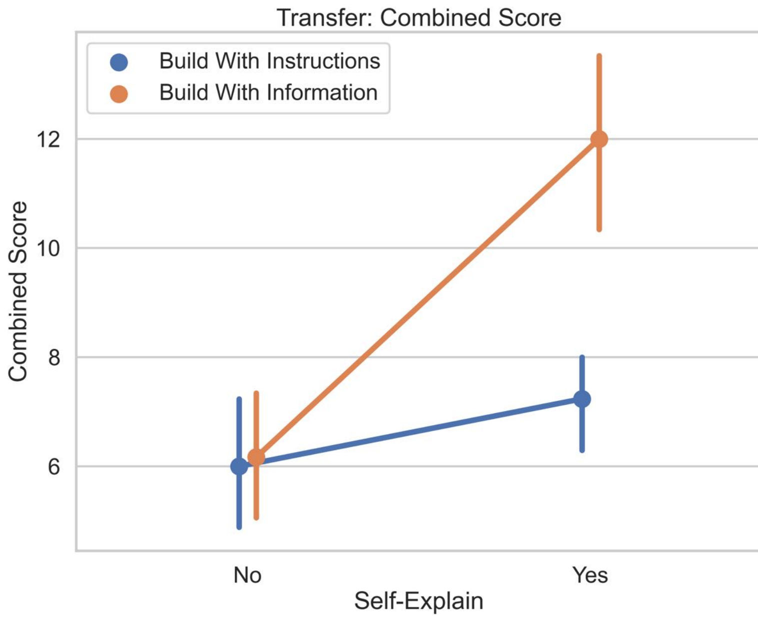


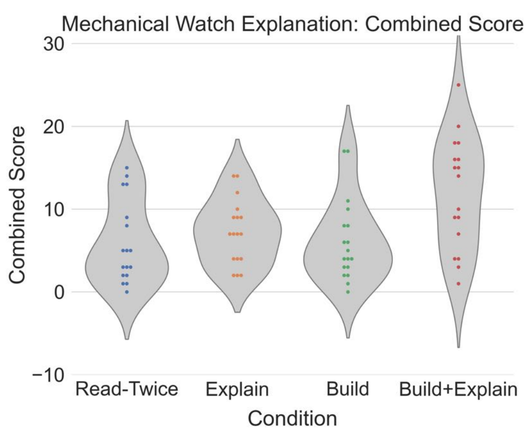
Figure 4.29: Two-way interaction plot showing the effects of building and explaining on the Combined score for the Explain-a-Mechanical-Watch task.

To investigate this further post-hoc tests using the Holm-Bonferroni correction were conducted to compare pairwise means between groups. There were significant or marginally-significant differences between the Build+Explain group and all three of the other groups, with the Build+Explain group scoring higher on the task than the other three groups. The Build+Explain group scored significantly higher on the task than the Build group, $t(29.15) = -2.89, p < 0.05$, the Read-Twice group, $t(32) = 2.93, p < 0.05$, and marginally-significantly higher than the Explain group, $t(32) = -2.5, p < 0.07$. The means and standard deviations for all four groups can be found in Table 4.16, the distributions of combined scores can be found in Figure 4.30a, and a comparison of means can be found in Figure 4.30b.

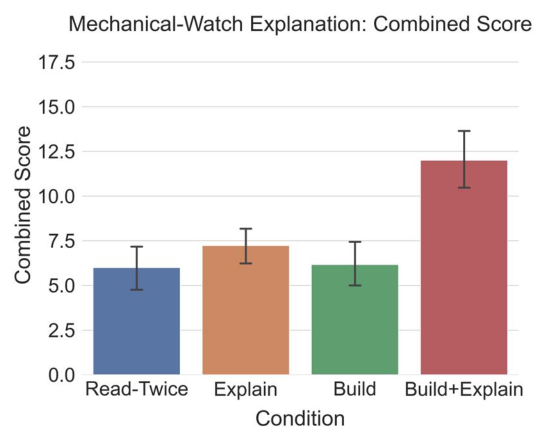
Next, a covariate analysis was conducted to explore the effect of controlling for potential confounds. This analysis was conducted via a series of linear regressions. The first regression model controlled for Time-on-Task, the second model controlled for all covariates, and the

Table 4.16: Mean and standard deviations of Combined scores on the Explain-a-Mechanical-Watch Task.

Condition	N	Mean	SD
Read-Twice	17	6.00	5.02
Explain	17	7.23	3.91
Build	17	6.17	4.94
Build+Explain	18	12.00	6.78



(a) Distribution of Combined scores on the Explain-a-Mechanical-Watch task.



(b) Mean Combined scores for the Explain-a-Mechanical-Watch task. Error bars show the standard error of the mean.

Figure 4.30: Combined scores for the Explain-a-Mechanical-Watch task.

third regression model contained the covariates identified as providing the best fit by a stepwise regression. In this case, the only covariate identified for inclusion in the third regression model was Time-on-Task.

Across all the regression analyses, the coefficient on the Build+Explain group remained significant at α=0.01, and the coefficient on Time-on-Task remained significant at $\alpha = 0.05$. When including Time-on-Task as a covariate, the coefficient on Build+Explain increased from 6.00 to above 8. This meant that when controlling for Time-on-Task, students in the Build+Explain group scored over 8 points higher on the task than the baseline (Read-Only) group. See Figure 4.31 for a visualization of the relationship between Time-on-Task, Combined score, and the experimental condition. When setting each of the other groups as the baseline, the coefficient on Build+Explain remained significant at $\alpha = 0.01$, though the coefficient dropped from around 8 to around 6.

In sum, the Build+Explain group dramatically outperformed all of the other groups on the Explain-a-Mechanical-Watch transfer task, and this effect became even strong when controlling for Time-on-Task. These results implied that knowledge transfer between analogous mechanisms was best supported by building, then explaining. While there was some knowledge transfer in the other three conditions, the combined effect of building then explaining was dramatic. The mean Combined score of Build+Explain group was nearly double that of the other three groups, indicating that students in this condition were able to transfer twice as much information to the novel, analogous mechanism.

4.2.5.1 Individual Analyses of the Purpose, Direct Cause, and Indirect Cause Scores

In order to learn more about the dramatic difference between the Build+Explain group and the other three groups, analyses of the differences between groups on the three individual scales were performed. The three scales that made up the Combined scale were Purpose, Direct Cause, and Indirect Cause.

Table 4.17: Results of regression analyses of Combined scores on the Explain-a-Mechanical-Watch Task

	Dependent variable:			
	Combined Watch Explanation Score			
	(1)	(2)	(3)	(4)
Build	0.17	2.07	2.15	2.07
	(1.78)	(1.95)	(1.98)	(1.95)
Build+Explain	6.00***	8.27***	8.45***	8.27***
	(1.80)	(2.04)	(2.07)	(2.04)
Explain	1.24	1.65	1.78	1.65
	(1.80)	(1.79)	(1.82)	(1.79)
Time-on-Task		-1.85**	-1.92**	-1.85**
		(0.75)	(0.76)	(0.75)
Mech. Reasoning			1.00	
			(1.44)	
CRT-2			-0.41	
			(0.68)	
Constant	6.00***	4.89***	4.80**	4.89***
	(1.28)	(1.39)	(2.06)	(1.39)
Observations	69	68	68	68
R^2	0.18	0.25	0.26	0.25
Adjusted R^2	0.15	0.20	0.19	0.20
Note:		*p<0.1; **p<0.05; ***p<0.01		

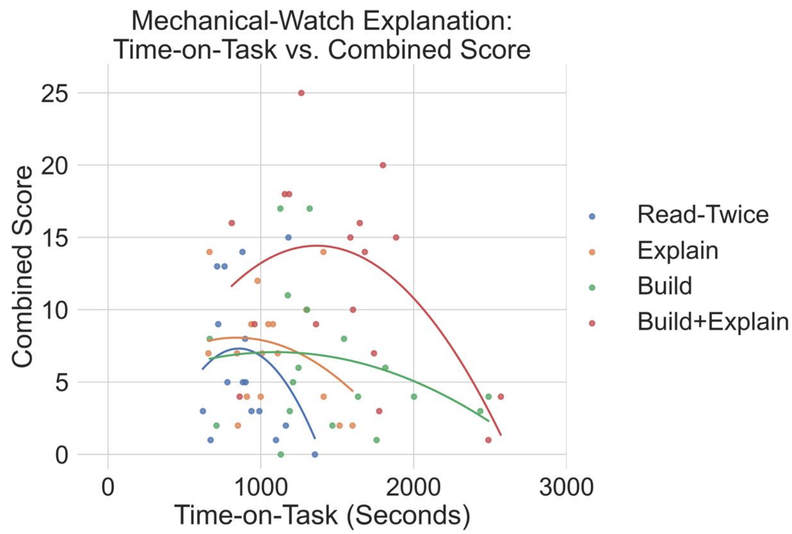


Figure 4.31: Relationship between Time-on-Task, Combined score on the Explain-a-Mechanical-Watch task, and experimental condition.

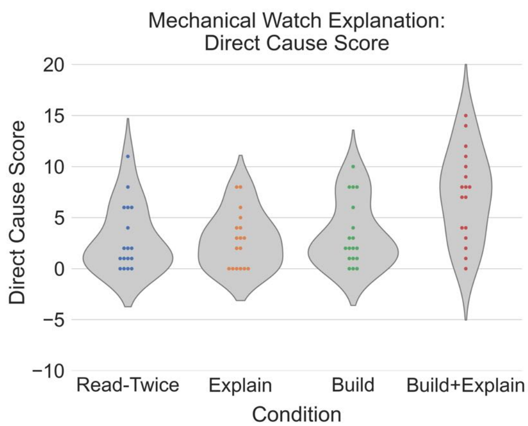
The difference on the Combined score was driven almost entirely by performance on the Direct Cause score. The results of a one-way ANOVA were highly significant $F(3, 65) = 6.23, p < 0.001$, and a follow-up two-way ANOVA found significant a main effect of building, $F(1, 65) = 8.08, p < 0.01$, a main effect of explaining, $F(1, 65) = 4.99, p < 0.05$, and an interaction effect, $F(1, 65) = 5.81, p < 0.05$.

To investigate this further post-hoc tests using the Holm-Bonferroni correction were conducted to compare pairwise means between groups. There were significant differences between the Build+Explain group and all three other groups. The Build+Explain group scored significantly higher than the Build group, $t(29.05) = -2.93, p < 0.05$, the Explain group, $t(32) = -3.49, p < 0.01$, and the Read-Twice group, $t(32) = 3.17, p < 0.05$. The means and standard deviations for all four groups can be found in Table 4.18. The distributions of Direct Cause scores can be found in Figure 4.32a, and a comparison of means can be found in Figure 4.32b.

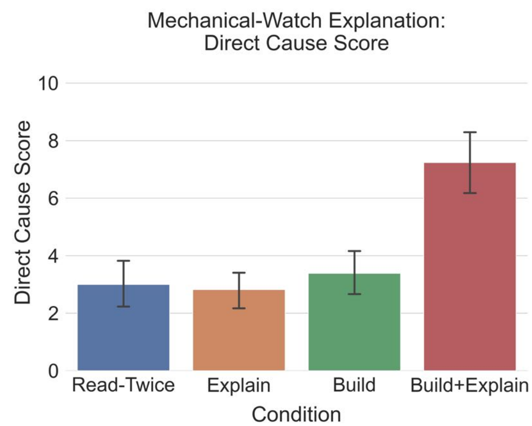
The only other significant effect was that self-explaining had a positive impact on the Purpose scale. A two-way ANOVA found a significant main effect of explaining on the

Table 4.18: Mean and standard deviations of Direct Cause scores on the Explain-a-Mechanical-Watch Task.

Condition	N	Mean	SD
Read-Twice	17	3.00	3.28
Explain	17	2.82	2.74
Build	17	3.39	3.20
Build+Explain	18	7.24	4.42



(a) Distribution of Direct Cause scores on the Explain-a-Mechanical-Watch task.



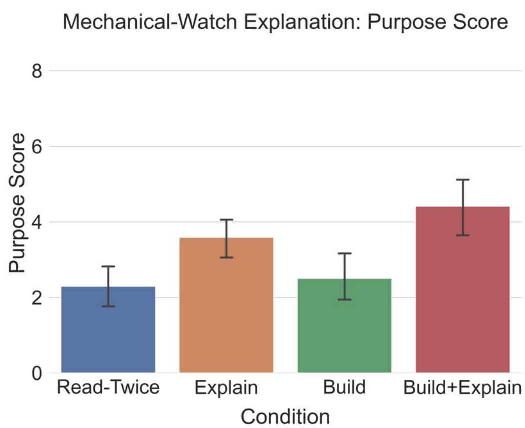
(b) Mean Direct Cause scores for the Explain-a-Mechanical-Watch task. Error bars show the standard error of the mean.

Figure 4.32: Direct Cause scores for the Explain-a-Mechanical-Watch task.

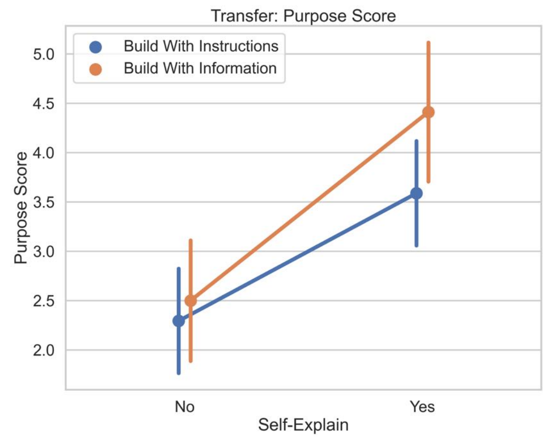
Purpose scale, $F(1, 65) = 7.16, p < 0.01$. This effect can be seen most clearly in the interaction plot (Figure 4.33b).

4.2.6 Transferring Knowledge to Reconstruct an Analogous Mechanism (Verge & Foliot Clock)

There was no relationship found between conditions on solving the verge & foliot clock problem, $\chi^2(4, N = 69) = 3.34, p = 0.50$. More details are provided in 4.19.



(a) Mean Purpose scores for the Explain-a-Mechanical-Watch task. Error bars show the standard error of the mean.

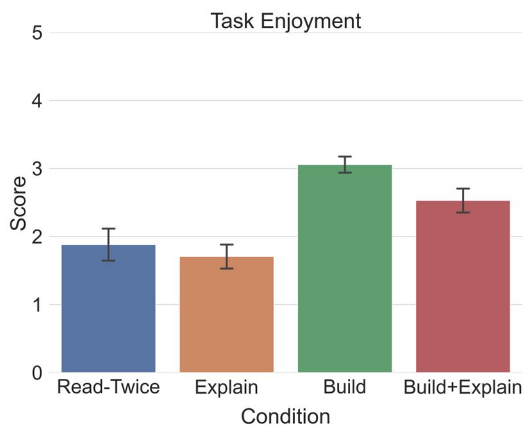


(b) Two-way interaction plot showing the effects of building and explaining on the Purpose score for the Explain-a-Mechanical-Watch task.

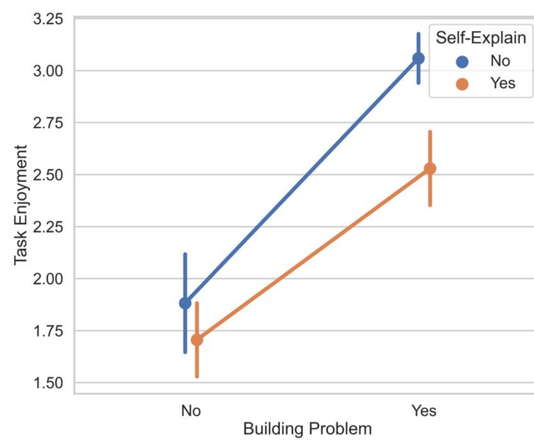
Figure 4.33: Purpose scores for the Explain-a-Mechanical-Watch task.

Table 4.19: Number of students who solved or failed to solve the verge & foliot problem split by condition.

Condition	Solved the Problem	Did Not Solve the Problem
Read-Twice	8	9
Build	10	8
Explain	9	8
Build+Explain	7	10



(a) Mean enjoyment scores. Error bars show standard error of the mean.



(b) Two-way interaction plot showing the main effect of building on enjoyment.

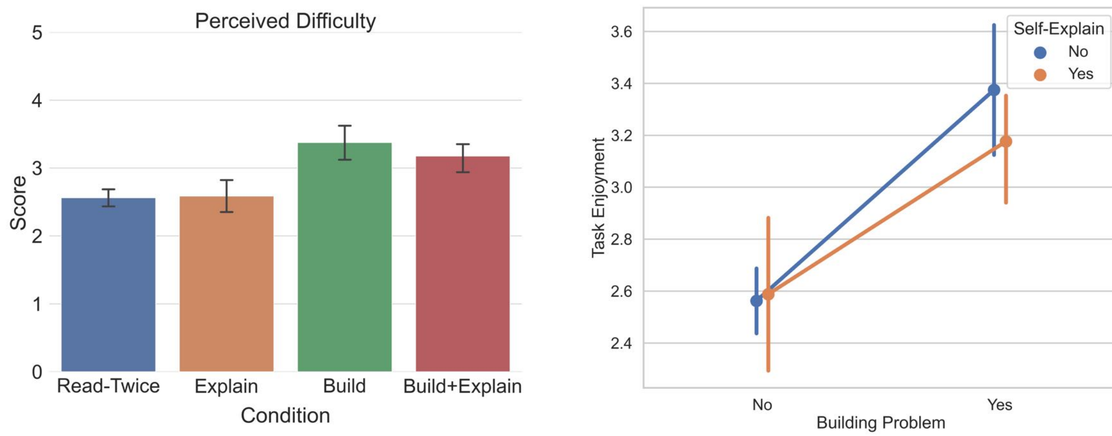
Figure 4.34: Scores for the question "How much did you enjoy learning about the pendulum clock?"

4.2.7 Enjoyment and Difficulty Learning about the Pendulum Clock

On the posttest, students were asked how much they enjoyed learning about the pendulum clock. There was a strongly significant main effect of building on enjoyment, $F(1,64) = 28.28, p < 0.001$, indicating that students who built the clock using information from the webpage enjoyed learning about the clock more than students who assembled the clock with instructions (Figure 4.34).

Students were also asked about how difficult it was to learn about the pendulum clock. There was a significant main effect of building on perceived difficulty, $F(1,64) = 9.88, p < 0.01$, with students who rebuilt the clock using information from the webpage reporting that it was more difficult than students who assembled the clock with instructions (Figure 4.35).

Thus, students who had to figure out how to build the clock found the task to be more enjoyable, and also more difficult, than students who assembled the clock using step-by-step instructions.



(a) Mean difficulty scores. Error bars show standard error of the mean.

(b) Two-way interaction plot showing the main effect of building on perceived difficulty.

Figure 4.35: Scores for the question "How difficult was it to learn about the pendulum clock?"

4.2.8 Summary of Findings

A summary of all the Findings discussed in this section can be found in Table 4.20.

4.3 Discussion

The goal in this study was to investigate the ways in which engaging in the engineering design process might support learning about mechanisms. This study focused on two aspects of the engineering design process—problem-solving and reflection—and attempted to understand the impact that each of these activities had on specific dimensions of mechanistic understanding.

The research questions in this study were as follows:

1. Is the problem-centered nature of engineering design important to supporting learning about mechanisms, or will other forms of making such as assembling with instructions also build mechanistic understanding?

Table 4.20: Summary of main and interactive effects in the study.

Groups	Type of Effect	Effect	Assessment
Explain, Build+Explain	Main effect of explaining	Stronger declarative knowledge, better able to reason about hypothetical situations	Verbatim questions, Inference questions
Explain, Build+Explain	Main effect of explaining	Better knowledge of the purpose of components, ability to transfer this knowledge	Explain-a-Pendulum Clock question, Explain-a-Mechanical-Watch question
Build, Build+Explain	Main effect of building	More structurally and organizationally complete mental models of the clock, better able to produce complete and coherent drawings of the clock from memory	Draw-a-Pendulum-Clock question
Build+Explain	Interaction effect of building and explaining	Much stronger understanding of how the components in the clock work together mechanically, best overall on transfer task	Explain-a-Pendulum-Clock question, Explain-a-Mechanical-Watch question

2. How does self-explaining support learning about mechanisms? Does it support learning about some dimensions of mechanistic knowledge more than other dimensions?
3. How does the combination of problem-solving and self-explaining support learning about mechanisms?
   (a) Do both activities support learning about the same dimensions of mechanisms, or does each activity foster a distinct form of understanding?

All of the students in this study had the same learning objective—to learn about how a pendulum clock worked—and worked with the same materials—a webpage containing information about pendulum clocks and a partially-disassembled LEGO model of the clock. However, the activities that students engaged in during the study differed between conditions. These conditions were determined using a 2x2 factorial design. The first factor compared a problem-centered making activity (figuring out how to reassemble a pendulum clock using information from a webpage) to a making activity with no problem (assembling a pendulum clock with step-by-step instructions). The second factor compared self-explaining the text on the webpage to reading it twice. Crossing these factors resulted in four different treatments, and students in the study were randomly assigned into one of four groups that received one of these treatments (Figure 4.1). By comparing these groups in different ways on a number of posttest questions, it was possible to answer each of the research questions listed above. A short description of each condition is provided in the following paragraph.

Students in the Read-Twice group read each section of the webpage twice. Each section described a different component in a pendulum clock. After reading each section twice, the students in this group were asked to locate the LEGO component on the table and place it into its correct location on the LEGO pendulum clock. Because they were provided with a photograph clearly showing where to place this component, this was a type of step-by-step assembly with instructions. Students in the Explain group read each section of the webpage once, and at the end of each section they were asked to explain what they had read in their own words. They were prompted to explain the name of the part, its purpose or role in the clock, and what parts it interacted with and how it interacted with them to do its job. After explaining each part, they were asked to identify and place the corresponding LEGO component back into the LEGO model with the help of a photograph showing the correct location. Students in the Build group were presented with the partially-deconstructed LEGO clock and told that their objective was to figure out how to put it back together and get it working. The webpage was presented as a source of information about pendulum clocks that they could use as a resource in whatever way was most helpful to them. They were also told that they could ask for help if they got stuck. Finally, students in the Build+Explain group started by going through the same activity as students in the Build group: figuring out how to reconstruct the clock using information from the webpage. However, after finishing this task, they were then instructed to read through the webpage, section-by-section, and to explain what they had read in their own words.

4.3.1 RQ1: Problem-Centered Making vs. Assembling with Instructions

The first research question in this study was "Is the problem-centered nature of engineering design important to supporting learning about mechanisms, or will other forms of making such as assembling with instructions also build mechanistic understanding?" To answer this research question, the two groups who engaged in a problem-centered making activity (figuring out how to reassemble the LEGO pendulum clock using information from the webpage) were compared to the two groups who did engage in problem solving (assembling the LEGO pendulum clock with step-by-step instructions).

The first finding was that students who engaged in problem-centered making produced drawings of the pendulum clock from memory that were more complete, coherent, and correct than students who assembled with instructions. Students in the problem-centered making groups drew more components, and the components were drawn in ways that were more structurally-accurate, making them easier to distinguish from other components in the clock. Additionally, they drew the organization of the components—their locations in the clock and the ways they interacted with each other—in ways that were more complete and correct.

One of the four dimensions of mechanistic understanding is knowledge of the organization of components in the system. Craver writes:

> To provide a mechanistic explanation, one shows how the different features of the phenomenon depend upon the organizational features of the underlying mechanism... it is insufficient merely to describe components and their activities. One has to further describe how those entities and activities are organized (e.g., spatially and temporally) into a mechanism. 2006, p. 373

The finding that students who engaged in problem-centered making produced more coherent, complete, and accurate drawings of the pendulum clock suggests that problem-centered making is a more effective way of learning about the organizational structure of a mechanism than other types of making such as building something with instructions.

One explanation for this effect is that the students in the problem-centered making condition had to use structural and organizational information in the task—thinking about the shape of components and how they worked together—and that this was a deeper form of processing than assembling with instructions. Because of this, students who worked on the building problem may have encoded more structural and organizational information in their long-term memory, which was reflected in the higher-quality drawings of the clock. An alternate explanation for this effect is that students in the two problem-centered making conditions simply interacted with the LEGO clock more than the students in the step-by-step assembly conditions, and that this made them more familiar with its parts and their structural organization. A follow-up study could investigate this hypothesis by comparing a Build group like the one in this study to a Start-From-Scratch condition that would fully assemble the LEGO clock from individual LEGO pieces using the official instructions. These groups would be likely to spend the same amount of time interacting with the clock, but only the Build group would be engaged in problem-solving as well. Comparing these two groups on the drawing task could provide more insight into the reason that the two Build groups build stronger organizational knowledge of the clock.

The second finding was that there were no main effects of problem-centered making on (a) declarative knowledge about the pendulum clock, (b) knowledge of the purpose of the components, and (c) knowledge of the direct, causal relations between components. Additionally, there was no main effect of problem-centered making on ability to transfer knowledge of the pendulum clock, as measured by the ability to explain the workings of a novel, analogous mechanism (the mechanical watch). Together, these findings suggest that the problem-solving aspect of engineering design is not sufficient to support the construction of complete mechanistic understanding about a system, and that it is also insufficient to build knowledge that transfers.

This finding might be related to the nature of the problem used in the study. While learning about the purpose and causal relations between components would have helped students reassemble the clock, it was possible to figure out how to solve the problem by only learning about the structural relations between components. For example, the simplest way of determining where the anchor belonged was to note that it physically interacted with the escape wheel, and then to find the location in the clock where the anchor could be attached so that the two parts could physically interact. A similar procedure could be followed for the rest of the parts. This strategy would not require students to pay attention to the purpose or causal relations between the components.

This explanation suggests that other types of problems might have been more effective at building knowledge about the purpose of parts or their causal relations. For example, if students had been required to build the missing parts from scratch, as opposed to working with pre-built modules, this might have supported learning about other aspects of the clock mechanism. This type of problem would also have been closer to the type of problem encountered in the engineering-design process. However, this type of problem is considerably harder than the one used in this study, and would have taken considerably more time. Nevertheless, exploring how different types of problem-centered making activities such as design and debugging might support the construction of mechanistic understanding is an interesting future direction.

4.3.2 RQ2: Self-Explaining vs. Reading

The second research question in this study was “How does self-explaining support learning about mechanisms? Does it support learning about some dimensions of mechanistic knowledge more than other dimensions?” To answer this question, the two groups who self-explained after reading each section of the webpage were compared to the two groups who did not self-explain. There were two main effects of self-explaining on learning about the pendulum clock mechanism. First, students who self-explained received high scores on the declarative knowledge questions than students who did not self-explain. Second, students who self-explained were more likely to transfer their knowledge about the purpose of each of the components in the clock to an analogous mechanism. However, students who self-explained were not more likely to learn about the direct, causal relations of the clock, indicated that self-explaining alone was not sufficient for the construction of complete mechanistic understanding.

The first effect of self-explaining, that students who self-explained built more declarative knowledge about the clock, was assessed using a set of six posttest questions. Three of these questions were designed to assess verbatim knowledge, and three were designed to measure the students' ability to make inferences or predictions about the pendulum clock in novel situations. The verbatim questions were based on information that was explicitly stated in the text, and the ability to answer these questions was based on the students' ability to recall information from the webpage. The inference questions required students to integrate information from multiple places in the text, and in some cases to generate new knowledge by bringing together prior knowledge with information in the text. These questions required students to explain and make predictions about the behavior of the pendulum-clock mechanism in novel, hypothetical situations. These question types were chosen based on prior work that demonstrated their effectiveness at picking up on the effects of self-explanation (Chi et al., 1994).

The students' answers to these questions were combined into a single score. On this combined measure, students who self-explained outperformed students who did not. This finding suggests that self-explaining was more effective at supporting the construction of declarative knowledge about the pendulum clock, and that these students were also better able to link these pieces of declarative knowledge together to draw inferences about the behavior of the pendulum clock in novel, hypothetical situations. This finding was a replication of prior work on self-explanation (Chi et al., 1994). Self-explaining is theorized to support learning about mechanisms because it supports the construction of more coherent mental models. During the process of generating an explanation, contradictions between a student's mental model and the information in the text are surfaced and brought to the student's attention. Additionally, during the process of self-explaining students must make inferences to fill in information that is missing in the text (VanLehn et al., 1992). Both of these processes are theorized to support the construction and refinement of mental models. Students who self-explain score higher on verbatim and inference questions because they possess higher-quality mental models (i.e., more systematic understanding).

The second effect was that students who self-explained were better able to transfer their knowledge about the purpose of components from the pendulum clock to explain the mechanical watch. A related finding was that students who self-explained learned more about the purpose of components in the pendulum clock, though this was only a trending relationship. This can be most-clearly seen in the interaction plots for the Explain-a-Pendulum-Clock task (Figure 4.27b) and the Explain-a-Mechanical-Watch task (Figure 4.33b). Whether this effect should be considered distinct from the previous one is debatable. The purpose of each of the components in the clock was explicitly stated in the text and videos, meaning that knowledge about the purposes of components was a form of verbatim, declarative knowledge. On the Explain-a-Pendulum-Clock task, knowledge of the purpose of components in a mechanism was a type of declarative knowledge, since these were all explicitly stated in the webpage that students read during the learning phase. However, because this effect was also present in the Explain-a-Mechanical-Watch task, a mechanism that students had not learned about, this choice was made to differentiate it from the previous effect. This finding suggests that self-explaining supported the construction of declarative knowledge about the purpose of components, and that students were able to transfer this knowledge to an analogous mechanism.

There were two aspects of mechanistic understanding where self-explaining provided no additional benefit: organizational knowledge; and direct, causal knowledge.

There was no main effect of self-explaining on the Draw-a-Pendulum-Clock task. This implied that self-explaining did not support the construction of structural, organizational knowledge about the pendulum clock. Recall that the effects of self-explanation are theorized to support the construction of more coherent, complete mental models about a mechanism. This finding suggests that the mental models that students construct through self-explanation alone may be incomplete, lacking structural and organizational information. Self-explaining is known to have boundary conditions (i.e., limits on its effectiveness), such as being ineffective in domains where there are no logical inferences or justifications (Wylie & Chi, 2014). This finding suggests that another boundary condition of self-explaining may be that it is not particularly effective at supporting learning about the structure and organization of a mechanism. It is important to note that the third self-explanation prompt used at the end of each section asked students, "What other parts does this part touch or directly interact with, and how does this part work together with those parts to do its job?" Thus, students did include organizational information in their self-explanations, but this did not provide additional support in the construction of organizational knowledge as measured in the drawing question.

There was also no main effect of self-explaining on learning about the direct, causal relations between components in the pendulum clock. Students who self-explained were not more likely to describe the direct, causal relations between components in the pendulum clock or in the components in the mechanical watch. This, too, implied that the mental models that students constructed during self-explaining were incomplete. This finding is more difficult to reconcile with the finding that self-explaining supported declarative knowledge about the pendulum clock. Presumably, the ability to answer inference questions is due to deeper understanding about how the clock works. This finding suggests that self-explaining may have supported the construction of abstract declarative knowledge—a network of propositions—that was not grounded in the concrete structure and organization of the clock itself. This might explain the lack of a differential effect on the ability to explain the direct, causal relations of the mechanisms when referencing a diagram.

Taken together, these findings suggest self-explanation alone is not sufficient to support the construction of complete understanding of a mechanism, as there were no main effects of self-explaining on knowledge about the structural or causal organization of components in the clock.

4.3.3 RQ3: The Combination of Problem-Centered Making and Self-Explaining

The third research question in this study was about how the combination of problem-solving and self-explaining might support learning about mechanisms. The findings indicated that the students in the Build+Explain condition were the only ones in the study to build complete mechanistic understanding about the pendulum clock, and that they were the only ones able to transfer that knowledge to figure out how a novel, analogous mechanism worked. In other words, the students in the Build+Explain group were the only ones who demonstrated understanding of how the pendulum clock worked.

The previous sections showed how problem-centered making and self-explaining each supported learning about complementary types of mechanistic knowledge. Problem-centered making supported learning about the structural organization of a mechanism, and self-explaining supported the construction of declarative knowledge, including knowledge of the purpose of the components. However, there was no main effect of either problem-centered making or self-explaining on learning about the direct, causal relations between components. This type of knowledge is one of the four dimensions of mechanistic understanding, and has been found to be the type of knowledge that experts possess and novices do not (Hmelo-Silver et al., 2007; Keil et al., 1999). So while self-explaining and problem-centered making each supported learning about specific dimensions of the pendulum-clock mechanism, neither could be said to support the construction of full mechanistic understanding. However, these two activities did support the construction of full mechanistic understanding when combined.

The Build+Explain group experienced the benefits of both problem-based making and self-explaining. Their declarative knowledge was the same as the Explain group, as evidenced by their scores on the Combined question scale, and the quality of their drawings was the highest overall, indicating that they possessed the most coherent and complete knowledge of the structural organization out of all four groups. Additionally, on both explanation posttest questions, students in the Build+Explain group scored as high on the Purpose scale as the Explain group. However, the Build+Explain group stood apart on three tasks in particular: the Draw-a-Pendulum-Clock task, the Explain-a-Pendulum-Clock task, and the Explain-a-Mechanical-Watch task. On each of these three tasks, what set the Build+Explain group apart was their performance on the Direct Cause scale. The strength of this effect across both explanation questions can be seen in the barplots for the Explain-a-Pendulum-Clock task (Figure 4.28a) and the Explain-a-Mechanical-Watch task (4.32b). This effect was especially pronounced on the transfer task, Explain-a-Mechanical-Watch, where the Build+Explain group scored nearly twice as high as any of the other three groups. On this task, students in the Build+Explain group scored significantly higher on the Direct Cause scale than all three of the other groups, all with large effect sizes (Cohen's $d > 1$). This was the largest and most dramatic effect found in the entire study. These findings suggest that constructing complete, transferable, mechanistic understanding required the combination of both problem-based making and self-explaining.

The impact of problem-based making and self-explaining was not an additive effect. That is, it was not the case that there was a main effect of either building or explaining that could explain the higher scores on the Direct Cause scales. Rather, this appeared to be an interactive effect; something that was due to the combination of both treatments. Furthermore, the type of knowledge supported by this interactive effect was knowledge of the direct, causal relations between components, which has been identified in previous work as type of knowledge that differentiates novices from experts. Understanding the reason for this interactive effect may provide some insight into how expert-like knowledge of mechanisms is constructed.

The findings suggest an explanation for this effect: one must first understand a mechanism's structure and organization before they are able to learn about its direct, causal relations. In the case of the pendulum clock, this would mean that only once students had learned about the parts, their shapes, and how they were organized in the clock were they ready to learn about how those parts worked together, causally. During the problem-centered making activity, students learned about the organization of the clock. This was suggested by the fact that the Build and Build+Explain groups both received similar scores on the Draw-a-Pendulum-Clock question. However, the Build+Explain condition learned significantly more about the ways the parts were causally related than the Build condition. This suggested that they constructed this direct, causal knowledge through self-explanation. However, the Explain group (who did not engage in a problem-centered making activity) also did not learn about the direct, causal relations in the clock. This suggests that self-explanation is better able to support learning about causal relations when there is prior knowledge about the parts and their structural organization. Problem-based making supported students in learning about the parts and their organization, and self-explaining helped put those parts into motion.

The Build+Explain group was the only group to build knowledge of the pendulum clock along all of the dimensions of mechanistic understanding, and was also the only group that managed to transfer this knowledge to infer how an analogous mechanism (the mechanical watch) worked. This suggests that complete mechanistic understanding (as defined by Craver (2006) and others) is necessary in order for transfer of this kind to occur. The ability to transfer knowledge to novel situations is a marker of expertise (Council, 2000). This finding suggests that supporting the construction of mechanistic knowledge along all of its dimensions—parts, purpose, organization, and causal relations—is important if the goal is to foster expertise. Further implications of this are discussed in the final chapter.

4.3.4 Limitations

4.3.4.1 Why Learn About Mechanical Clocks?

At first glance, the pendulum clock may seem like a strange, anachronistic topic to learn about. Pendulum clocks are an outdated technology that have been eclipsed by electronic clocks with quartz or cesium oscillators. From this perspective, there is little reason to care about interventions that foster deeper learning about pendulum clocks, since students are unlikely to ever use this knowledge in any meaningful way. However, the pendulum clock was not chosen because of its relevance to students' lives; rather, it was chosen because it exemplified a mechanistic system. This study was not intended to provide insight into how instruction might be improved for a specific topic; it was designed to provide insight into the nature of mechanistic understanding, and to identify interventions that fostered the construction of complete, transferable mechanistic knowledge. From this perspective, the pendulum clock was an ideal choice. Because pendulum clocks are uncommon in modern society, it was highly unlikely that the students who participated in the study would possess any prior knowledge about how they worked. This was confirmed during piloting. Additionally, pendulum clocks are an excellent example of a mechanistic system. A pendulum clock is a mechanism, in that it "consists of entities (or parts) whose activities and interactions are organized so as to be responsible for [a] phenomenon" (2017, p. 17). Understanding how the clock keeps time (the phenomenon) requires understanding the purpose or function of each of these parts, understanding how they are organized, and understanding how the direct, causal interactions between the parts are what allows the clock to keep time. Furthermore, none of these parts or interactions between them are hidden from view, as is the case in some systems (e.g., electronics); in fact, the parts are all clearly visible, and they directly interact in tangible, mechanical ways. This is important because it means making sense of the mechanical interactions does not require conceptual background knowledge or specialized tools, as would be the case in other systems (e.g., understanding an electrical system requires an understanding of voltage, current, Ohm's law, and so on). Finally, the pendulum clock is a system that can be readily disassembled, reassembled, and modified, which is not a characteristic of all types of mechanistic systems (e.g., biological organ systems). While there are virtual technologies which make this sort of interaction possible across a wide range of systems, learning about virtual systems would have introduced other factors into the study that were not of interest. In summary, the pendulum clock was chosen because of its suitability as a model mechanistic system, similar to how stack-based calculators have been used in the literature on problem-solving with mental models (Halasz & Moran, 1983; Norman, 2014). If this was a valid choice, it means that findings related to how students learn about the pendulum clock mechanism should apply to learning about other mechanistic systems as well.

4.3.4.2 The Build Group Did Not Read-Twice

There was an issue with the treatment that the Build group received. In a true 2x2 factorial design, two groups would have self-explained, and two groups would have read the text on the webpage twice. In the current study, the Build group did not read the text on the webpage twice from top-to-bottom in the same way that the Read-Twice group did. They did read the content on the webpage, as it would have been impossible for them to figure out how to reassemble the clock otherwise. However, the number of times that the Build group read each section was not controlled, meaning that it is likely that students in this condition did not read each section on the webpage twice. This leaves open the possibility that the effects found in the Build+Explain group may not have been due to self-explaining after the problem-centered making task, but may have been due to reading the text on the webpage after working on the problem. This suggests a follow-up study that would compare a Build+Explain group to a Build+Read-Twice group. If the Build+Read-Twice group was not found to learn the causal relations between components, or not found to transfer their knowledge of the pendulum clock to infer the behavior of the mechanical watch, this would provide stronger evidence that the effect found in the current study was due to building, then self-explaining, as opposed to building, then reading the text again.

4.3.4.3 Reading-Twice Was Not a Proper Control Group

The Read-Twice group was used as a baseline condition that did not receive the treatment of either self-explaining or problem-centered making. However, a true baseline condition would have been no learning phase at all. The choice to use a read-twice condition in this way was based on Chi et al. (1994), who compared students in a self-explain condition to students in a read-twice condition. Chi et al. treated the read-twice condition as the control condition, and justified this choice in terms of time-on-task. They wrote that "reading [the passage] twice (as opposed to three or four times) engaged the unprompted students with the text for roughly the same amount of time as having the prompted students read and explain the text" (p. 451). This choice was likely made based on prior qualitative work by Chi et al. (1989). In that study, the authors observed students' strategies as they worked through physics examples in a textbook, finding that students who self-explained more frequently learned significantly more than students who simply read. Based on this prior work, it seemed sensible to include the read-twice group as a baseline, even though it was not a proper control group.

4.3.4.4 The Build+Explain Group May Have Benefited from Reduced Cognitive Load

The explanation provided above for why the Build+Explain condition learned more than the other conditions was based on the order in which they learned different types of knowledge about the mechanism. The idea was that the problem-centered making activity supported the construction of structural, organizational knowledge, which was then set into motion by self-explaining. However, an alternative explanation is that the order in which this information was presented did not matter, but what mattered is that students in the Build+Explain condition learned about different aspects of the mechanism in distinct, sequential phases. All of the other groups had to learn about the full mechanism during a single activity, which might have overloaded their working memory and made it difficult to build complete understanding of the clock. This suggests a follow-up study that would compare two groups: one that worked on the building problem, then self-explained; and one that self-explained, and then worked on the building problem. If both groups were found to learn the same types of knowledge and demonstrated the same capacity for transfer, this would suggest that the combined effect was not due to constructing knowledge in a particular order. But if differences were found, with the Build+Explain group learning more than the Explain+Build group, this would provide additional evidence for the hypothesis that learning about a mechanism should be done in stages. Additionally, if differences were found this would provide a possible explanation for prior findings that students who work on hands-on problems before an instruction phase learn more than students who receive instruction and then work on problems (Schneider et al., 2013).

4.3.4.5 Explaining With the Completed Clock Might Have Caused Transfer

Another explanation for the effect of problem-centered building followed by self-explaining is that the complete, working LEGO clock was present on the table during the self-explanation phase. This made it possible for students to map the content about the pendulum clock from the webpage onto the LEGO clock, which had the same deep, mechanistic structure as the pendulum clock described on the webpage but different surface structure. This type of activity has been shown to support analogical transfer (Gentner et al., 2003; Gick & Holyoak, 1983), which meant the presence of the complete LEGO clock during self-explanation might have been the cause of the Build+Explain group's higher scores on the transfer task. A fifth condition was designed to account for this possibility. This condition was the same as the Explain condition with one exception: the students were provided with a fully-assembled LEGO clock at the outset. At the end of each section, the students in this group were prompted to self-explain in the same way as the students in the Explain group. However, these students were explicitly asked to reference the working LEGO clock during their explanations, which differentiated them from the Explain group. Across all measures, this group performed similarly to the Explain group, indicating that the presence of the complete LEGO clock was not responsible for the effects of building, then explaining.

4.3.4.6 A Low Number of Explanations Might Have Suppressed the Self-Explanation Effect

Students in the Explain group did not learn as much about the pendulum clock mechanism as students in the Build+Explain group. There is a possibility that if students in the Explain group had produced a higher number of self-explanations, they would have learned more about the pendulum clock. Students in the current study were only asked to generate explanations at the end of each section, as opposed to generating an explanation after reading each section as was done in Chi et al. (1994). This reduced the frequency and granularity of explanations when compared to Chi et al. However, the decision to reduce the frequency of self-explanations was based on more-recent work by Hausmann and Chi (2002) which showed that there was still a positive effect on learning when self-explanations occurred less frequently than one-per-sentence. Additionally, the current study replicated the self-explanation effect on declarative knowledge, indicating that the number of self-explanations was sufficient to produce an effect. Still, a follow-up study that compared students who self-explained once at the end of each section to students who self-explained after each sentence might show that a higher number of self-explanations better supports the construction of complete mechanistic understanding.

Chapter 5

Discussion

The goal of this dissertation was to better understand how working on projects in makerspaces might affect the development of students' skills and knowledge. However, because makerspaces support an enormous variety of activities, it was necessary to restrict the focus to a specific type of activity and a specific set of outcomes. The activity was building mechanistic systems through the engineering design process, and the outcomes of interest were the development of mechanistic problem-solving skills and the construction of mechanistic understanding.

In the first study, high-school students took part in a year-long digital fabrication course in a makerspace. Over the course of the study, they worked on two long-term projects using the engineering design process. Both of the projects involved designing a single object that consisted of many interacting parts, and completing the projects required progressive refinement over the course of multiple iterations of the engineering design cycle. In the second project in the course, the students worked together to design and build a mechanistic system, the Rube Goldberg machine. After taking part in the course, the students received significantly higher scores on a set a hands-on problems involving mechanisms, and a fine-grained analysis of their problem-solving approaches indicated that they had become significantly more like expert mechanical engineers. This study suggested that designing and developing complex systems over multiple iterations of the engineering design process supported the development of mechanistic problem-solving skills in the high-school students.

The second study was designed to learn more about how building a mechanistic system through the engineering design process might support the construction of mechanistic knowledge. This study used a 2x2 factorial design to understand the impact of different aspects of the engineering design process—problem-centered making and reflection—on specific dimensions of mechanistic understanding. Students in this study were tasked with building a LEGO pendulum clock, and they all had access to a webpage containing text, diagrams, and videos about how pendulum clocks worked. The two factors in this study were problem-centered making [building the clock with step-by-step instructions vs. having to figure out how to build the clock using information on the webpage] and reflection [self-explaining the text on the webpage vs. not self-explaining]. Crossing these factors resulted in four groups: Read, Build, Explain, and Build+Explain. By comparing these groups on a set of posttest questions designed to measure different dimensions of mechanistic understanding, it was possible to learn more about how each of these activities supported learning about the pendulum-clock mechanism.

The first finding was that problem-centered making and self-explaining each supported learning about distinct dimensions of mechanistic knowledge. The students who self-explained learned more about what the purposes of the parts in the clock, and scored higher on a set of questions designed to assess declarative knowledge about the clock. The students who worked on the building problem (as opposed to the students who simply rebuilt the clock using instructions) learned more about the structural organization of the parts in the clock. However, neither of these activities was superior at supporting the construction of causal knowledge about the parts in the clock. The second finding was that the students in the Build+Explain group who engaged in both activities learned significantly more about the causal relations between components in the clock than any of the other groups. Additionally, this group of students were significantly more likely to transfer their knowledge to infer how an analogous, novel mechanism worked. These findings suggested that the combination of problem-centered making and reflection was most effective at supporting the construction of complete, transferable knowledge about a mechanism.

5.1 Limitations

Both of the studies in this dissertation were designed to "open the black box" and investigate how building mechanistic systems via engineering design impacted students mechanistic problem-solving skills and supported the construction of mechanistic understanding. However, each study had its own set of limitations, and neither fully achieved this goal.

The first study looked at how participating in a year-long course in digital fabrication impacted students' mechanistic problem-solving skills. Before and after the course, the students worked on a set of hands-on problems involving mechanistic systems. An analysis of differences in performance and approach between the pre-course and post-course students suggested that taking part in the course supported the development of the students' mechanistic problem-solving skills. A comparison to a group of mechanical engineering graduate students indicated that the post-course students were more significantly more like expert mechanical engineers than the pre-course students. All of these findings had to do with the effects of taking part in the course. However, this study did not to investigate the causes of those effects, which was a major limitation. No data about the students' experiences was collected during the course, and the study did not employ an experimental design. This meant that there was no way of knowing which elements of the course contributed to the development in problem-solving skill. While all of the students did work on projects over multiple iterations of the engineering design process, it is not possible to conclude that this played a role in bringing about those changes. However, this study does provide a strong suggestion that these two things are related, which is a motivation for future work in this area.

The second study was designed to investigate the ways that building a mechanism using the engineering design process might support students in learning about that mechanism. Because this was a laboratory study that employed an experimental design, it was possible to investigate the respective contributions of two different activities associated with the engineering-design process—problem-based making and reflection—on the construction of mechanistic understanding. This design was used to address one of the limitations in the previous study, which was the inability to associate any specific causes with the observed effects. The second study succeeded in uncovering a number of associations between problem-based making, self-explanation, and distinct dimensions of mechanistic knowledge. This addressed the limitation in the previous study, which was a failure to investigate how distinct aspects of the course were each associated with the observed outcomes.

However, a limitation of this study was that the activity that the Build and Build+Explain students engaged in was not a "true" engineering design project. At best the activity could be considered a highly-structured variation or simulation of a single cycle of engineering design. The task began with a problem, which was to build a pendulum clock. To solve this problem, the students had to learn about the clock from the webpage, and then use this information to generate hypotheses about where the various parts of the clock belonged. They were able to test their hypotheses by putting the clock together and testing whether the clock actually worked. Solving the problem typically required a number of cycles of hypothesis generation and testing. All of these things made this activity similar to the engineering design process. However, this project was different from a typical engineering design process in a number of ways. One difference was that the students did not have to design the parts that they were using to build the clock. The parts were given to them already-assembled, and the students had to figure out how to connect these parts in order to get the clock working. Another difference was in the nature of the problem. Engineering design is typically framed as a process that can be used to solve ill-structured problems, which "often possess conflicting goals, incomplete information, multiple solution methods, and multiple criteria for evaluating solutions" (Lee et al., 2011, p. 247). However, the building problem used in this study was fairly well-structured. There was a single goal state, all of the information needed to solve the problem was provided in the webpage, and there was a single criterion for evaluating the solution.

The Build+Explain treatment was designed to investigate the impact of first working on a building problem and then reflecting. The reflection activity was reading about the pendulum clock and explaining how each part of the clock worked. This activity design was based on a common troubleshooting practice: explaining the different aspects of a problem out-loud when stuck. The self-explanation phase in the Build+Explain group occurred after the students had already finished the building problem, and involved reading about the mechanism that they had constructed and explaining how it worked. However, this specific type of reflective practice is not a part of the typical engineering design process: Engineers do not typically read and self-explain text about a system after building it. This implies that the activity that was most representative of the typical engineering design process was that used in the Build group, who worked on the building problem but did not self-explain afterwards. Taken together, these two limitations suggest that the findings from the second study can only suggest that the typical engineering process supports learning about mechanisms.

Finally, in both of the studies in this dissertation studies worked with simple, engineered mechanisms like the differential and the pendulum clock. Learning about other types of mechanisms such as biological, computational, or astronomical systems is out of the scope of this work, but does suggest an avenue for future research.

5.2 Implications

5.2.1 Study 1: Theoretical Contributions and Implications

Prior research suggesting that students who work in makerspaces develop problem-solving skills has almost entirely relied on self-reports. Students in these studies are interviewed or surveyed after working in a makerspace and asked if they developed problem-solving skills. However, to date little work has attempted to measure or quantify these changes in problem-solving skills. This means little is known about the sorts of skills students develop, and even less is known about how those skills develop as a consequence of working on projects in a makerspace. The first study in this dissertation was designed to address this lacuna, and provided empirical evidence suggesting that working on engineering design projects in a makerspace supported the development of mechanistic problem-solving skills, making them more like experts.

The first contribution that this study offers is a specification of the theory that working in makerspaces supports the development of problem-solving skills. This study focused on a more-specific type of activity (engineering design) and a more-specific type of problem-solving (mechanistic), and the evidence suggested that these two may be related. In highlighting the possible relationship between these two activities, this study may help focus future research on problem-solving in makerspaces.

This study also provides new evidence about engineering expertise. By comparing the expert mechanical engineers to the pre-course students, it was possible to show that what set them apart was a difference in problem-solving approaches. On the Gearbox problem, experts performed a higher proportion of mechanical actions, which suggested that they had categorized the problem differently from novices. On the Flashlight-Repair problem, the experts used a more systematic troubleshooting strategy, which made it more likely that they would identify and correct all sources of error. Additionally, this study suggested that one of the ways that engineers develop problem-solving expertise is through working on and solving problems during the engineering design process. What suggested this inference was that after taking part in the course, the high-school students were significantly more likely to adopt expert-like problem-solving approaches.

None of these contributions would have been possible without the use of a set of novel, hands-on, mechanistic problems designed to assess mechanistic problem-solving skills. These assessments and the methods developed to analyze them are discussed in the following section.

5.2.2 Study 1: Methodological Implications

A novel set of hands-on problems were designed for this study to assess changes in mechanistic problem-solving skills. Two coding schemes were developed to quantify (a) how close each participant got to solving the problem and (b) the problem-solving approach each participant adopted during the problem. An unsupervised clustering algorithm was used to identify groups of participants with similar approaches, and by analyzing the process data from these groups it was possible to learn more about (a) the differences in approaches between experts and novices, and (b) how participation in the digital-fabrication course was associated with a significant, positive change in mechanistic problem-solving skills. This is the main methodological contribution of the first study.

This approach is based on a long history of research on problem solving that aims to understand expert-novice differences by having participants work on a specific problem and tracking their progress through the problem space (Newell & Simon, 1972). This work contributes two new problems and demonstrates that each can be used for measuring and understanding different types of mechanistic problem-solving skills. More generally, this study suggests that future work on problem-solving in makerspaces would benefit from using these problems and analysis techniques, or from adopting a similar approach. Using these hands-on problems provided much greater detail and insight into the changes in problem-solving skills than simpler methods such as interviews and surveys.

Additional evidence for the value of these methods is provided by comparing the students' performance on the hands-on problems to their performance on the written EFT assessment. The results from the EFT assessment suggested that the students became more confident in their knowledge about the internal components of mechanisms and more persistent in their approach to troubleshooting a broken device. Both of these results suggest a more general increase in confidence related to mechanisms. However, the survey failed to detect changes in the students' abilities to accurately identify components in a mechanistic system, and did not detect changes in their actual troubleshooting strategies, indicating an increase in confidence without an accompanying increase in competence. Without the use of the two hands-on problems, this would have been the conclusion of the study. This would have completely missed the fact that the students experienced an increase in problem-solving competence. There are practical implications of this advance that are discussed in the following section.

5.2.3 Study 1: Practical Implications

These findings have implications for educators who are interested in designing activities for makerspaces and in assessing the learning that occurs when students work on those activities.

The mechanistic problem-solving skills that the students developed during the course have practical value. It is possible that students who developed these skills are more capable of solving problems encountered in their everyday lives (e.g., fixing a bike, fixing a running toilet, assembling a piece of furniture). While future work would be needed to determine if the students were more capable of coping with everyday mechanistic problems after taking part in the course, the findings provide some reason to believe that this might be the case. Educators who are interested in fostering these kinds of problem-solving skills could design a digital-fabrication course around the same projects used in the first study (the Omni-Animal and the Rube Goldberg machine projects). Both of these projects are available online, and both are supported by the set of tools found in typical makerspaces (i.e., they do not require any specialized tools or materials).

This study may have future practical implications for educators who are interested in assessment in makerspaces. If the goal is to assess changes in mechanistic problem-solving, the hands-on problems used in this study could prove to be useful. As was demonstrated in the findings, there were strong correlations between progress towards the solution and the problem-solving approach adopted. This means that progress towards the solution, which is fairly simple to score, is a decent index of expertise. Much more work would need to be done to validate these tasks before they could be used as assessments. However, this may be a valuable direction for future research, since in the first study these hands-on problems were able to measure problem-solving skills which another, paper-based assessment failed to pick up on.

5.2.4 Study 2: Theoretical Contributions and Implications

Complete understanding of a mechanism is theorized to involve multiple dimensions of knowledge. This involves knowing about (a) what the mechanism does (i.e., the phenomenon), (b) what entities or parts it consists of, (c) how those parts are structurally organized, and (d) how the parts causally interact with one another to give rise to the phenomenon (Illari & Williamson, 2012). One of the core hypotheses in this dissertation was that building a mechanism through the engineering design process supports learning about the causal relations between parts in the mechanism being constructed. The second study was designed to investigate this hypothesis.

One of the main findings in the second study was that the Build+Explain students scored significantly higher than any of the other groups on a transfer question that required them to look at a diagram of a novel, analogous mechanism (a mechanical watch) and explain how it worked. Additionally, the Build+Explain group was found to have built the most complete body of mechanistic knowledge about the pendulum clock. The most consistent and dramatic difference between this group and the other groups was that students in the

Build+Explain group learned significantly more about the direct, causal relations between the components in the clock.

This finding has a number of implications. First, this finding provides some evidence for the validity of theories about what constitutes mechanistic understanding: that one can only be said to understand how a mechanism works if they possess knowledge of its parts, their structural organization, and their causal interactions. The students in the Build+Explain group possessed more knowledge about all of these dimensions of the pendulum clock mechanism, and they also were far more likely to transfer their knowledge to correctly infer the causal relations in the analogous mechanism. This suggests that there may be a relationship between this type of knowledge transfer and the possession of complete knowledge of the function, parts, purpose, and causal relations in a mechanism.

Second, this finding suggests that learning about the direct, causal relations between components in a mechanism is more likely to occur if the student already possesses mechanistic understanding along the other dimensions (i.e., prior knowledge or parts, their purpose, and their structural organization). This conclusion is based on the following analysis of the types of knowledge learned in each of the different groups.

First, what did people who self-explained learn? These students were able to make predictions about how the clock would behave in novel situations, and they understood the roles that each part played in the clock. However, when asked to explain how the clock worked, these students demonstrated weak understanding of how the parts causally interacted. Additionally, on the transfer task this group of students received high scores on the Purpose scale, but low scores on the Cause scale, indicating that they were able to infer the roles of the parts in the analogous mechanism, but did not understand how they worked together. Together, these findings suggest that self-explaining may have supported the construction of abstract declarative knowledge—a network of propositions—that was not grounded in the concrete structure and organization of the clock itself.

This is something that prior research on the self-explanation effect might have missed.

Prior work has found that self-explaining is an effective way of supporting learning about mechanisms (Chi et al., 1994). The explanation for this effect is that self-explaining is a process that unites fragmented knowledge into a complete mental model while also surfacing and resolving contradictions between the mental model and the information in the text (VanLehn et al., 1992). Evidence for this comes from students' answers on verbatim and inferential questions, where students who self-explain score higher than students who simply read the text. However, to the best of our knowledge, research on the self-explanation effect has not asked students to explain how the target mechanism works from a diagram, nor has it asked students to draw diagrams of the mechanism from memory, nor has it looked at whether students are able to transfer their knowledge to an analogous system. This means that there is a possibility that the students in prior work on the self-explanation effect did not actually build complete mechanistic understanding through self-explaining, but only constructed an abstract network of propositional knowledge. In fact, the results from the second study in this dissertation imply that this may be the case. This suggests that it would be worthwhile to replicate prior work on the self-explanation effect, but to include the use of other forms of assessment (e.g., drawing from memory, explaining with a diagram, transfer) to gain a more complete understanding of what students learned. This is discussed in more detail in the following section.

Second, what did the group who worked on the building problem learn? This group built better knowledge about the structural organization of the clock than the groups who reassembled the clock with instructions, however this activity was not effective at supporting learning about the causal relations between parts in the clock. A simple explanation for this is that the students only needed to use information about the structural organization of the clock (i.e., the shape of its pieces and which parts they interacted with) to solve the problem. Because of this, they had no reason to pay attention to the information on the webpage about how the clock worked, since this was not required for solving the problem. This suggests that the type of problems encountered during the engineering design process may determine what is learned, an implication that is discussed in more detail below.

Finally, why did the Build+Explain group learn so much more about the direct, causal relations between components in the clock? One possible explanation is that learning about the causal relations depended on prior knowledge about the structural organization of the clock. This is suggested by the fact that students in the Build+Explain group first worked on the building problem, and then self-explained the material on the webpage. First, during the problem-centered making activity, the students learned about the organization of the clock. Then, while self-explaining, instead of building an abstract network of propositional knowledge, they were able to build a more-concrete body of knowledge about the actual clock. Problem-based making supported students in learning about the parts and their organization, and self-explaining helped put those parts into motion.

In the Limitations section above, I argued that that the group activity that was most similar to the process of engineering design was the Build group, not the Build+Explain group. Engineers do not typically read and self-explain information about a system after successfully constructing it, which was what students in the Build+Explain group did. However, the students who did this learned much more about the pendulum-clock mechanism than any of the other groups, including the Build group. One implication of this finding is that there may be variations of the engineering design process that are better suited for use in education. This points to a direction for future research, to investigate forms of *educational* engineering design that are different from the typical process. One form of educational engineering design implied by these findings is to add an explicit self-explanation phase after students have worked on a design problem. Future work investigating this idea is described in the Next Steps.

Another limitation of this study was that none of the activities could be called true engineering design problems. This suggests two things. First, this suggests that future work is needed to investigate how working on proper engineering design problems supports mechanistic understanding. Second, this suggests that there may be useful types of educational activities that are inspired by the engineering design process, but that do not involve the full process of iteratively solving ill-structured problems over longer periods of time. The Build+Explain activity is one example. On average, this activity took 15-20 minutes. It is unlikely that a student with no engineering background would be able to design and build a working clock using the engineering design process in such a short period of time. However, if the goal is for students to build understanding of the mechanism that they are designing, the more-structured, problem-centered, reflective building activity used in the Build+Explain group may provide a useful template. More research is needed to investigate the effectiveness of this method for learning about other mechanisms, and to explore the educational potential of other types of structured, engineering-design inspired activities.

5.2.5 Study 2: Methodological Implications

The primary methodological implication of the second study has to do with the way that mechanistic knowledge is measured. This study used a variety of questions to assess the multiple dimensions of mechanistic understanding. The Draw-a-Pendulum-Clock task was designed to measure knowledge of the structural organization of the clock. The Explain-a-Pendulum-Clock task was designed to measure knowledge of the purposes of the parts in the clock as well as how they causally interacted to make the clock work. A set of declarative knowledge questions were used to measure how well students remembered what they read on the webpage, and to measure the students' abilities to make inferences about how the clock would behave in novel situations. Finally, the Explain-a-Mechanical-Watch question measured the ability to transfer knowledge of the pendulum-clock mechanism to infer how a novel, analogous mechanism worked.

Without measuring all of these dimensions of mechanistic knowledge, a number of effects would have been missed. If the drawing question had not been included, the effect of working on the building problem would have been missed. If the declarative questions had not been included, some of the effects of self-explaining would have been missed. If the two explanation questions had not been included, the interaction effect of building then explaining would have been missed. What this suggests is that future work on mechanistic understanding should take care to measure knowledge along of these dimensions.

Additionally, this implies that much of the prior work on how students learn about mechanisms may be incomplete. In the second study, the effect of self-explaining was associated with higher scores on declarative knowledge, which replicated prior work on the self-explanation effect (Chi et al., 1994). In that study, the authors concluded that students who self-explained built better mental models of the system, which was determined by analyzing the answers given on a set of inference questions. However, students in that study were not asked to draw a diagram of the circulatory system from memory, they were not asked to explain how the circulatory system worked from a diagram, and they were not asked to transfer their knowledge of the circulatory system to infer the behavior of an analogous mechanism (e.g., the circulatory system of an octopus, which has three hearts). A future study could attempt to replicate this work with the addition of these types of questions to determine whether self-explaining supported the construction of complete mechanistic understanding. This is one example of how a more-complete set of assessments might be used in research on mechanistic understanding.

One area of research in particular where this method of measurement may be useful is in research on learning about mechanisms by building computational models (Goel et al., 2009; Helms et al., 2010; Jonassen & Strobel, 2006; Mulder et al., 2016; Wilkerson et al., 2018). In cases where these models are designed in an environment that is abstracted from the phenomenon (as is the case with stock-and-flow models, functional-flow models, and models described using code), students may fail to learn about the structural organization of the system. This suggests a direction for future research that compares more abstract forms of modeling such as stock-and-flow (“SageModeler – Systems Modeling Tool”, n.d.) to forms of modeling that are more concrete such as bifocal modeling (Blikstein, 2012). It may be the case that these more-concrete approaches better support learning about the structural organization of the system, a type of knowledge which the second study in this dissertation showed was important to developing complete mechanistic understanding.

5.2.6 Study 2: Practical Implications

The primary practical implication of this study has to do with assessing knowledge about mechanisms. This study provided evidence that there is value in designing assessments to measure knowledge along multiple dimensions of mechanistic understanding. However, some of these forms of assessment are easier to use than others. For example, the drawing assessment is simple to implement and grade, while it is unclear how the two explanation questions could be used in practice. Designing question types that can be used in practice to assess multiple dimensions of mechanistic understanding is research that may be worth pursuing.

A secondary, tentative implication is for educators who are interested in designing activities in makerspaces that support learning about mechanisms. The activity used in the Build+Explain condition may be a useful template for the design of other activities that support student learning. For example, to support learning about electricity, students could be provided with a set of electrical components that could be assembled into a working circuit. Students could be provided with text about the circuit and its parts, but not provided with schematics or step-by-step assembly instructions. After figuring out how to put the circuit together, the students could then be asked to read and self-explain the text describing the circuit and its parts. Both of these implications should be considered tentative without future research. Until then, the implications from this study should remained limited to the domain of mechanical clocks.

5.3 Next Steps

The findings from the first study in this dissertation suggested that taking part in the year-long digital-fabrication course supported the development of mechanistic problem-solving skills in a group of high-school seniors. However, this study was unable to offer any suggestions about what aspects of the course might have contributed to this change. One possibility is that the type of project that students worked on played a role. Students in the first study worked on two projects: the Omni-Animal and the Rube-Goldberg machine. The Omni-Animal project involved designing and building a static figurine consisting of many smaller pieces, and the Rube-Goldberg project involved building a dynamic, mechanical system. A within-subjects crossover design could be used to isolate the effects of project-type on the development of mechanistic problem-solving skills. In the first phase of the study, one group of students would work on the Omni-Animal and the other would work on the Rube-Goldberg machine. In the second phase, the students who worked on the Omni-Animal would work on the Rube-Goldberg machine, and vice versa. By having students work on the hands-on problems at three points in the course—before, between the two phases, and after—it would be possible to learn more about whether the development of mechanistic problem-solving skills was an outcome of encountering and working on problems during the design of a mechanism, or whether it was a more general outcome of working on any project through engineering design.

The second study suggested a number of future directions. One is to investigate variations of the engineering design process with more explicit pedagogical support (i.e., forms of *educational* engineering design). Recent work has explored how inventing a principle from a set of contrasting cases during the engineering design process supports learning and transfer (Chase et al., 2019), which could be considered a form of educational engineering design. In that study, students were randomly assigned into one of three groups. All of the groups worked on an engineering design project (building a cantilever), but only one of the groups invented a principle with contrasting cases. A similar approach could be used to learn more about the effects of a self-explanation phase during or after working on an engineering design project. The second study suggests that adding a self-explanation phase after working on an engineering design problem may prove to be an effective way of supporting the construction of mechanistic understanding.

Another direction is to investigate educational activities that are inspired by the engineering design process, but that can be done in a shorter period of time or with fewer resources. One example of this type of activity is the one that the Build+Explain group worked on in the second study. This activity involved building a mechanism using information on a webpage, then reading and self-explaining the text after completing the task. This should be considered a type of problem-centered, reflective making, as opposed to a form of engineering design. If this task could be shown to support learning about other mechanisms, not just mechanical clocks, this would have the practical effect of providing educators with a template that they could use to design effective learning activities in makerspaces. This could be determined by replicating the second study in this dissertation with other mechanisms, such as electrical circuits, robots, and gearboxes.

This suggests another future research direction, which would involve investigating how problem-centered, reflective making might support learning about other types of mechanisms. Some examples of mechanisms that are difficult or impossible to make in this way are plant cells, organ systems, neurons, factories, airplanes, nuclear reactors, neural networks, and ecosystems. It is obviously difficult or impossible for students to learn about these mechanisms by building them. However, it is possible to create virtual simulations of these mechanisms that can be taken apart and put back together. An exciting future direction would be to explore how *virtual*, problem-centered, reflective making activities might support learning about these mechanisms. Students working with these virtual simulations could engage in the same process of figuring out how to build a mechanism using a set of resources that explained the parts of the system, their purposes, and how they causally interacted.

5.4 Conclusion

Together, the studies in this dissertation suggest that building mechanisms through a problem-centered, reflective process may support the development of mechanistic problem-solving skills as well as the construction of mechanistic understanding. This work has implications for researchers who are interested in how people learn about mechanisms, and for practitioners who are interested in supporting students in learning about how things work. Future work is needed to better understand the mechanisms through which skill and knowledge develop during the process of problem-centered, reflective making. These findings point to future directions for research, including the investigation of educational engineering design processes and the use of virtual, problem-centered, reflective making activities to support learning about a larger number of mechanistic systems.

Appendices

Appendix A

Detailed Analysis of the Mechanical-Reasoning, Cognitive-Reflection, and GPA Covariates in Study 2

In the section Pre-Analysis: Measuring Possible Confounds Between Conditions, it was reported that only one out of 16 potentially confounding factors varied between grounds: duration of the learning phase (i.e., time-on-task). However, three additional covariates were included in analyses of learning in the pendulum-clock study: mechanical-reasoning scores, cognitive-reflection scores, and GPA. None of these variables differed significantly between groups. However, they were included in the covariate analyses because of their theoretical relationships to learning about mechanistic systems. Details about the distribution of these three variables between groups is presented below.

A.1 Mechanical-Reasoning Scores

Mechanical reasoning scores did not vary significantly between conditions, $\chi^2(3, N = 69) = 1.64, p = .65$. The distribution of mechanical reasoning scores between each group can be found in Figure A.1, and the proportion of students who got the mechanical reasoning question correct can be found in Table A.1.

Table A.1: Mechanical reasoning scores in the pendulum-clock study.

Condition	N	Proportion of Students Who Answered Correctly
Read-Twice	17	0.71
Explain	17	0.65
Build	17	0.83
Build+Explain	18	0.71

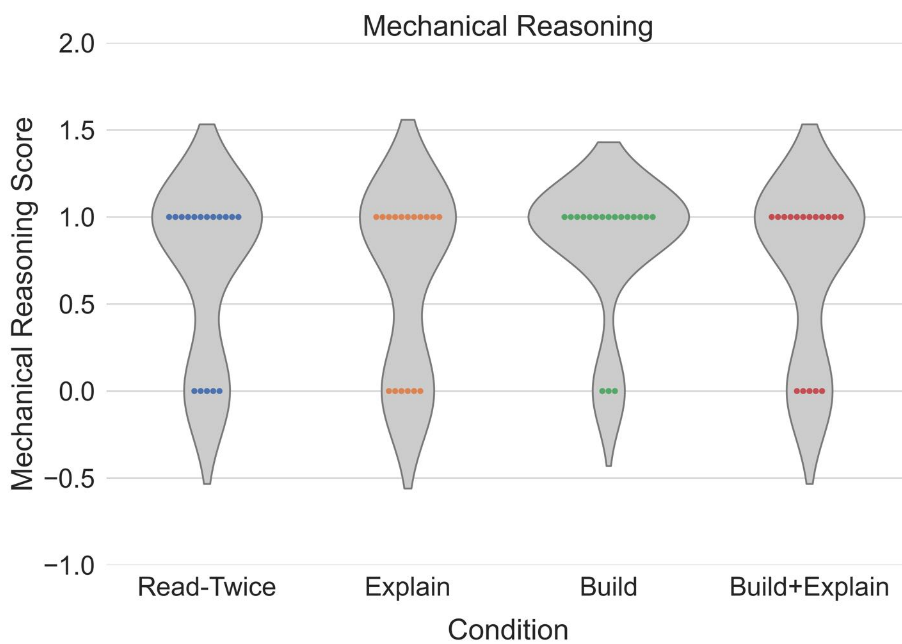
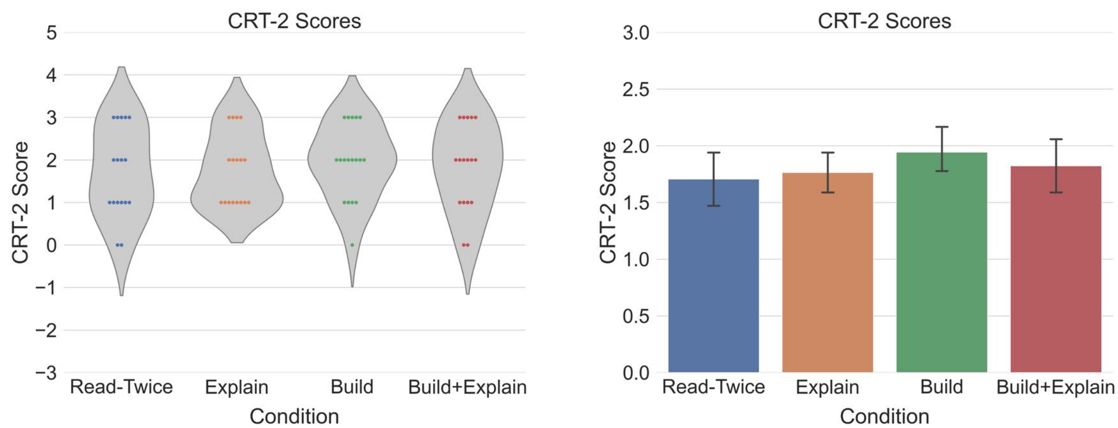


Figure A.1: Distribution of mechanical reasoning scores for students in each condition. Each point represents one student's score.



(a) Distribution of CRT-2 scores for students in each condition. Each point represents one student's GPA.

(b) Mean CRT-2 scores for each group. Error bars show the standard error of the mean.

Figure A.2: Cognitive reflection scores for each condition.

A.2 Cognitive-Reflection Scores

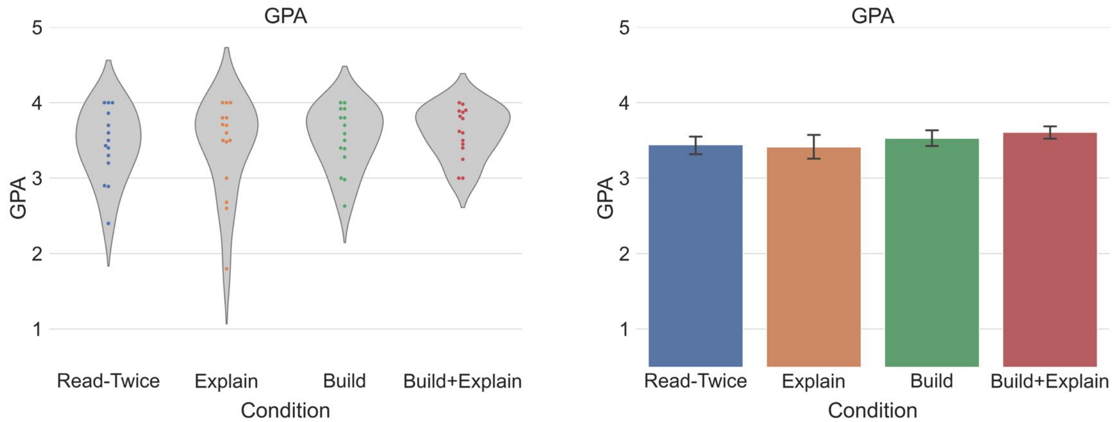
CRT-2 scores, a measure of miserly processing, did not vary significantly between conditions, $F(3, 65) = 0.20, p = 0.89$. The distribution of CRT-2 scores can be found in Figure A.2a, and the mean and standard deviation of GPAs between groups can be found in Table A.2 and Figure A.2b.

Table A.2: CRT-2 scores between conditions in the pendulum-clock study.

Condition	N	Mean	SD
Read-Twice	17	1.71	1.05
Explain	17	1.76	0.83
Build	17	1.94	0.87
Build+Explain	18	1.82	1.01

A.3 GPAs

Self-reported GPAs did not vary significantly between conditions, $F(3, 55) = 0.50, p = 0.68$. The distribution of GPAs between each group can be found in Figure A.3a, and the mean



(a) Distribution of GPAs for students in each condition. Each point represents one student's GPA.

(b) Mean GPA for each group. Error bars show the standard error of the mean.

Figure A.3: GPAs each condition.

and standard deviation of GPAs between groups can be found in Table A.3 and Figure A.3b.

Table A.3: Student self-reported GPAs between conditions in the pendulum-clock study.

Condition	N	Mean	SD
Read-Twice	17	3.44	0.48
Explain	17	3.41	0.63
Build	17	3.53	0.42
Build+Explain	18	3.60	0.33

Appendix B

Detailed Coding Schemes for Study 2

B.1 Coding Schemes for the Draw-a-Pendulum-Clock Task

In the Draw-a-Pendulum-Clock task, students were asked to draw a diagram of a pendulum clock from memory. Students in all conditions saw a number of diagrams of the pendulum clock during the learning phase, some of which depicted specific components and some of which depicted the entire clock. In this task, they were not allowed to view the webpage, nor could they see the LEGO clock. The drawings were generated entirely from the students' memories of the pendulum clock. The drawings were coded along three dimensions: Parts, Mechanisms, and Misconceptions. The full coding schemes are provided in the tables below.

B.2 Coding Scheme for the Pendulum-Clock Explanation

In the Pendulum-Clock explanation task, students were shown a diagram of a pendulum clock (Figure 4.6) and asked to explain how it worked in their own words. These explanations were coded along seven dimensions. There were four types of statements that were coded as correct–Parts, Purpose, Mechanism, and Causal—and three types of statements that

Table B.1: Parts coding scheme used to analyze the drawings from the Draw-a-Pendulum-Clock task.

Code	2 points	1 point	0 points
Weight	Weight clearly distinct from Pendulum	Weight present but not visually distinct from Pendulum	No Weight in the drawing
Gears	Gears clearly meshing with one another	Gears drawn as disconnected circles	No Gears in the drawing
Escape Wheel	Escape wheel clearly distinct from Gears	Escape Wheel and other Gears are visually identical	No Escape Wheel in the drawing
Anchor	Anchor present	N/A	No Anchor in the drawing
Pendulum	Pendulum clearly distinct from the Weight	Pendulum present but not visually distinct from Weight	No Pendulum in the drawing
Hands	Hands present	N/A	No Hands in the drawing

Table B.2: Mechanism coding scheme used to analyze the drawings from the Draw-a-Pendulum-Clock task.

Code	2 points	1 point	0 points
Weight to Gears	Connection from Weight through center of Gear	Any line between Weight and Gears	No connection
Gears to Escape Wheel	Escape Wheel on axle with Gear that meshes with other Gears	Gears and Escape Wheel touch or mesh	No connection
Escape Wheel to Anchor	Escape Wheel clearly interacts with Anchor	Escape Wheel is near the Anchor	No connection
Anchor to Pendulum	Anchor mounted on axle that clearly connects to Pendulum	Anchor mounted on axle	Anchor floating in space
Gears to Hand	Connection present	N/A	No connection

Table B.3: Misconception coding scheme used to analyze the drawings from the Draw-a-Pendulum-Clock task.

Code	2 points	1 point	0 points
Weight to Escape Wheel	Connection from Weight through center of Escape Wheel	Any line from Weight to Escape Wheel	No connection
Weight to Pendulum	Connection clearly rendered	Connection vaguely rendered	No connection
Pendulum to Escape Wheel	"	"	"
Anchor to Hand	"	"	"
Pendulum to Hand	"	"	"
Weight to Hand	"	"	"
Pendulum to Gears	"	"	"

were coded as incorrect—Purpose-Misconceptions, Mechanism-Misconceptions, and Causal-Misconceptions. All seven of these coding schemes are provided in the tables in this section.

Table B.4: The Parts coding scheme used to analyze the pendulum-clock explanations.

Code	2 points	1 point	0 points
Weight	Part explicitly mentioned by name	Part mentioned, but not by name (e.g., "this")	No mention
Gears	"	"	"
Escape wheel	"	"	"
Anchor	"	"	"
Pendulum	"	"	"
Hand	"	"	"

B.3 Coding Scheme for the Mechanical-Watch Explanation

In the watch-clock explanation task, students were shown a labeled diagram of a mechanical watch (Figure 4.7) and asked to explain how it worked in their own words. These explanations were coded along six dimensions, three correct and three incorrect. The three correct dimensions were Purpose, Mechanism, and Causal; and the three incorrect dimensions were

Table B.5: The Purpose coding scheme used to analyze the pendulum-clock explanations.

Code	2 points	1 point	0 points
Weight = Energy	Weight clearly described as source of energy or power	Vague mention of weight falling and gears turning	No mention
Escapement = Energy Regulator	Escapement or its parts described as regulating energy	Mentions “not spinning out of control”	"
Escapement = Rate	Escapement described as helping maintain the clock’s constant rate	N/A	"
Pendulum determines rate	Pendulum described as part that determines rate	N/A	"
Hands = Indicators	Hands described as indicating time	Mentions “minute”, “hour”, or “second” hands	"

Purpose-Misconceptions, Mechanism-Misconceptions, and Causal-Misconceptions. The full coding schemes are provided in the tables below.

Table B.6: The Mechanism coding scheme used to analyze the pendulum-clock explanations.

Code	2 points	1 point	0 points
Weight moves the Gears	Mechanical relation between parts clearly stated in explanation	Mechanical relation vaguely stated in explanation	No mention
Gears move the Hand	"	"	"
Escape Wheel moves the Hand	"	"	"
Gears move the Escape Wheel	"	"	"
Escape Wheel releases Gears	"	"	"
Escape Wheel moves the Anchor	"	"	"
Anchor regulates Escape Wheel	"	"	"
Anchor moves the Pendulum	"	"	"
Pendulum moves the Anchor	"	"	"
Pendulum length determines rate	"	"	"

Table B.7: The Causal coding scheme used to analyze the pendulum-clock explanations.

Code	2 points	1 point	0 points
Weight powers Pendulum directly	Causal relation clearly stated	Causal relation vaguely stated	No mention
Weight powers Escape Wheel directly	"	"	"
Weight powers the Hand directly	"	"	"
Weight powers the Escapement directly	"	"	"
Gears power the Pendulum directly	"	"	"
Gears power the Escape Wheel and Anchor	"	"	"
Escape Wheel powers the Pendulum directly	"	"	"
Escapement regulates the Gears	"	"	"
Escapement powers the Pendulum	"	"	"
Escapement drives the Hand	"	"	"
Pendulum regulates the Escape Wheel	"	"	"
Pendulum affects the Escape Wheel and Anchor	"	"	"
Pendulum and Anchor regulate the Escape Wheel	"	"	"

Table B.8: The Purpose-Misconceptions coding scheme used to analyze the pendulum-clock explanations.

Code	2 points	1 point	0 points
Another source of energy	Another (incorrect) source of energy identified in explanation	N/A	No mention
Anchor regulates power from Pendulum to Escape Wheel	Purpose clearly stated	N/A	No mention

Table B.9: The Mechanism-Misconceptions coding scheme used to analyze the pendulum-clock explanations.

Code	2 points	1 point	0 points
Anchor pushes and turns the Escape Wheel	Mechanical misconception clearly stated	Mechanical misconception vaguely stated	No mention
Anchor rocks the Escape Wheel	"	"	"
Escape Wheel rocks Anchor	"	"	"
Escape wheel turns the Gears	"	"	"
Escape wheel slows the Anchor	"	"	"
Gears release the Weight	"	"	"
Hand rocks the Weight	"	"	"

Table B.10: The Causal-Misconceptions coding scheme used to analyze the pendulum-clock explanations.

Code	2 points	1 point	0 points
Pendulum moves the Hand directly	Causal misconception clearly stated	Causal misconception vaguely stated	No mention
Escapement drives the Gears	"	"	"
Pendulum moves the Escape Wheel	"	"	"
Pendulum and Anchor power the Escape Wheel	"	"	"
Escapement causes the Pendulum to oscillate at a constant rate	"	"	"

Table B.11: Purpose coding scheme used to analyze the mechanical-watch explanations.

Code	2 points	1 point	0 points
Mainspring = Energy	Mainspring identified as the source of energy	N/A	No mention
Escapement = Energy Regulator	Escapement or its parts described as regulating energy	N/A	"
Escapement = Rate	Escapement described as helping maintain the watch's constant rate	N/A	"
Balance Wheel determines rate	Balance wheel described as part that determines rate	N/A	"
Hands = Indicators	Hands described as indicating time	Mentions "minute", "hour", or "second" hands	"

Table B.12: The Mechanism coding scheme used to analyze the mechanical-watch explanations.

Code	2 points	1 point	0 points
The Winding Pinion coils the Mainspring	Mechanical relation between parts clearly stated in explanation	Mechanical relation vaguely stated in explanation	No mention
The Mainspring moves the Gears	"	"	"
Gears move the Hand	"	"	"
Gears move the Escape Wheel	"	"	"
Escape Wheel releases Gears	"	"	"
Escape Wheel moves the Lever	"	"	"
Lever regulates Escape Wheel	"	"	"
Lever moves the Balance Wheel	"	"	"
Balance Wheel moves the Lever	"	"	"

Table B.13: The Causal coding scheme used to analyze the mechanical-watch explanations.

Code	2 points	1 point	0 points
Mainspring powers Balance Wheel directly	Causal relation clearly stated	Causal relation vaguely stated	No mention
Mainspring powers Escape Wheel directly	"	"	"
Mainspring powers the Hand directly	"	"	"
Gears powers Balance Wheel directly	"	"	"
Escape Wheel powers the Balance Wheel directly	"	"	"
Escapement and Hands are in sync	"	"	"
Lever regulates Escape Wheel and Gears	"	"	"
Lever and Escape Wheel regulate Gears	"	"	"
Balance Wheel regulates Escape Wheel	"	"	"
Balance Wheel regulates Lever and Escape Wheel	"	"	"
Balance Wheel and Lever regulate Escape Wheel	"	"	"

Table B.14: The Purpose-Misconception coding scheme used to analyze the mechanical-watch explanations.

Code	2 points	1 point	0 points
Another source of energy	Another (incorrect) source of energy identified in explanation	N/A	No mention
Winding Pinion sets the time	Purpose clearly stated	"	"
Winding Pinion acts as Pendulum	"	"	"
Mainspring displays time	"	Purpose vaguely stated	"

Table B.15: The Mechanism-Misconception coding scheme used to analyze the mechanical-watch explanations.

Code	2 points	1 point	0 points
Winding Pinion continually rotates the Mainspring	Mechanical misconception clearly stated	Mechanical misconception vaguely stated	No mention
Mainspring moves the Winding Pinion	"	"	"
Winding Pinion continually rotates the Gears	"	"	"
Winding Pinion continually rotates the Escape Wheel	"	"	"
Lever moves the Escape Wheel	"	"	"
Mainspring oscillates and rotates the Gears	"	"	"
Escape Wheel moves the Gears	"	"	"
Balance Wheel is the source of power that moves the Lever	"	"	"
Hands move the Gears	"	"	"
Gears move the Mainspring	"	"	"

Table B.16: The Causal-Misconception coding scheme used to analyze the mechanical-watch explanations.

Code	2 points	1 point	0 points
Winding Pinion powers the Gears	Causal misconception clearly stated	Causal misconception vaguely stated	No mention
Winding Pinion winds up the Balance Wheel	"	"	"
Escape Wheel moves the Hands	"	"	"
Hands move the Escape Wheel	"	"	"
Escape Wheel regulates the Hands	"	"	"
Escape Wheel slows down the Balance Wheel	"	"	"
Balance Wheel moves Hands directly	"	"	"
Escapement powers the Gears	"	"	"
Balance Wheel powers the Escape Wheel	"	"	"
Balance Wheel rocks the Escape Wheel	"	"	"
Escapement rocks the Balance Wheel at a constant rate	"	"	"
Escape Wheel and Lever regulate power released from Balance Wheel	"	"	"
Escape Wheel powers the Mainspring	"	"	"

Appendix C

Study 2 Pretest

REP ID: ____

Date and Time: __________________________________

Please be honest in answering the following questions.

1. Have you taken any programming or engineering courses in high school or college? If so, how many?

2. Have you ever used LEGO to build machines with gears, motors, or pulleys? If so, what have you built?

3. Do you have any prior design experience? For example, have you worked on projects in a maker space or taken courses on design thinking?

4. If you're running a race and you pass the person in second place, what place are you in?

5. A farmer had 15 sheep and all but 8 died. How many are left?

6. Emily's father has three daughters. The first two are named April and May. What is the third daughter's name?

The following questions ask you about mechanical systems. Some examples of mechanical systems are bicycles, automobiles, and clocks. They contain moving parts like gears, pulleys, motors, and wheels.

7. How confident are you that you could fix a problem with an unknown mechanical system?
    a. Not at all confident
    b. Slightly confident
    c. Somewhat confident
    d. Confident
    e. Very confident

8. How confident are you that you could figure out how an unknown mechanical system works?
    a. Not at all confident
    b. Slightly confident
    c. Somewhat confident
    d. Confident
    e. Very confident

9. How confident are you that you could put together an unknown mechanical system?
    a. Not at all confident
    b. Slightly confident
    c. Somewhat confident
    d. Confident
    e. Very confident

10. How much do you enjoy learning about and working with mechanical systems?
    a. Not at all
    b. Slightly
    c. Somewhat
    d. A good amount
    e. A great deal

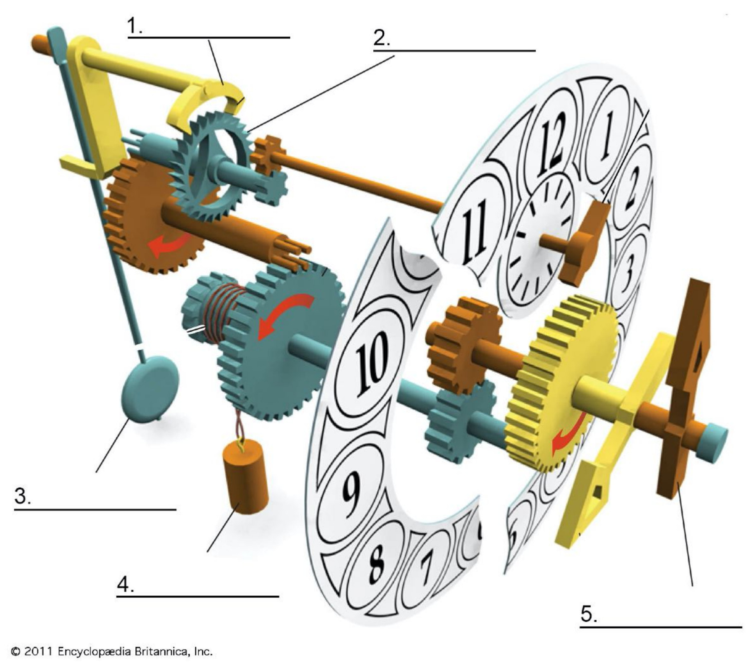
11. Would you be interested in future opportunities to learn about and work with mechanical systems?
    a. Not at all
    b. Slightly
    c. Somewhat
    d. A good amount
    e. A great deal

12. Do you have any prior experience working with mechanical systems?

13. What is the name of the system shown in the picture below?

14. How confident are you that you could **explain** how this system works?
    a. Not at all confident
    b. Slightly confident
    c. Somewhat confident
    d. Confident
    e. Very confident

15. How confident are you that you could **figure out** how this system works?
    a. Not at all confident
    b. Slightly confident
    c. Somewhat confident
    d. Confident
    e. Very confident



16. Explain what this system does and how it works. You can refer to each part using its name or its number in the diagram. If you don't know, write "I don't know."

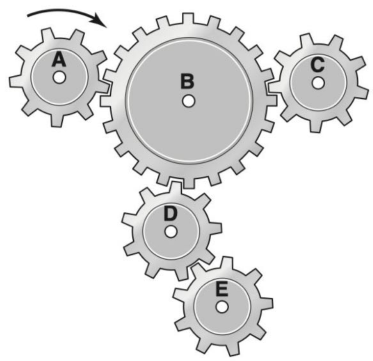
Appendix D

Study 2 Posttest

REP ID: ______________________________________

Date and Time: ________________________________

1. Five gears are shown in the diagram. If gear A turns as shown, what are the other gears that turn in the same direction?

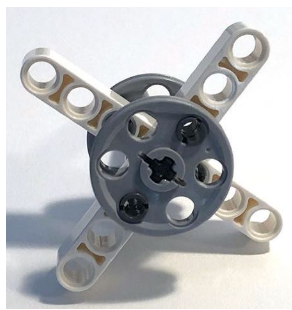


2. Clocks with pendulums tend to run faster when cold. This is caused by
   a. the pendulum becoming longer when cold.
   b. the pendulum becoming shorter when cold.
   c. the air expands and slows the pendulum.
   d. the air contracts and speeds up the pendulum.

3. What would happen if you were to double the weight in the mechanical clock?

4. Which of these four parts would you remove to stop the motion of the clock and prevent the gears from turning and the weight from falling?
    a. The pendulum
    b. The anchor
    c. The escape wheel
    d. The hand

5. How does the escape wheel move in the clock?
    a. It rotates in a single direction
    b. It rocks back and forth



6. What keeps the escape wheel moving?
    a. The pendulum swinging
    b. The weight falling
    c. The anchor pushing
    d. The hand rotating

7. What is the purpose of the pendulum in the clock?
    a. To provide energy that keeps the escape wheel turning
    b. To regulate the gears from spinning out of control
    c. To provide the energy that keeps the anchor moving back and forth
    d. To keep the parts in the clock moving at a constant rate

The following questions ask you about mechanical systems. Mechanical systems contain moving parts like gears, pulleys, motors, and wheels. Some examples of mechanical systems are bicycles, automobiles, and clocks. Please be honest in answering the questions in this survey.

8. How confident are you that you could fix a problem with an unknown mechanical system?
    a. Not at all confident
    b. Slightly confident
    c. Somewhat confident
    d. Confident
    e. Very confident

9. How confident are you that you could figure out how an unknown mechanical system works?
    a. Not at all confident
    b. Slightly confident
    c. Somewhat confident
    d. Confident
    e. Very confident

10. How confident are you that you could put together an unknown mechanical system?
    a. Not at all confident
    b. Slightly confident
    c. Somewhat confident
    d. Confident
    e. Very confident

11. How much do you enjoy learning about and working with mechanical systems?
    a. Not at all
    b. Slightly
    c. Somewhat
    d. A good amount
    e. A great deal

12. Would you be interested in future opportunities to learn about and work with mechanical systems?
    a. Not at all
    b. Slightly
    c. Somewhat
    d. A good amount
    e. A great deal

13. How much mental energy did you put in to learn about how the clock works from the text, animations, and videos?
    a. No mental energy
    b. A slight bit of mental energy
    c. Some mental energy
    d. A lot of mental energy
    e. A great amount of mental energy

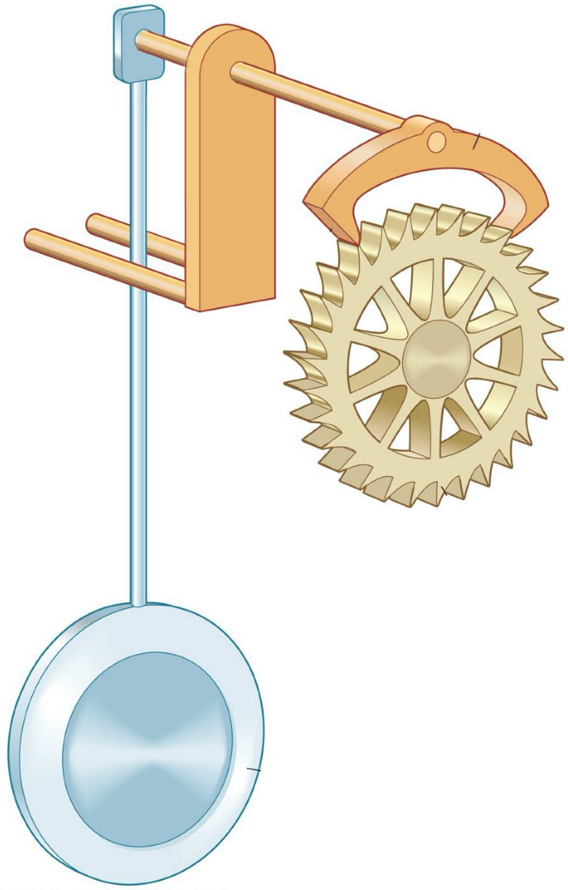
14. How much mental energy did you put in to actually put the clock back together?
    a. No mental energy
    b. A slight bit of mental energy
    c. Some mental energy
    d. A lot of mental energy
    e. A great amount of mental energy

15. How difficult did you find the task of putting the clock back together?
    a. Not difficult at all
    b. A slight bit difficult
    c. Somewhat difficult
    d. Difficult
    e. Very difficult

16. How much did you enjoy the task of putting the clock back together?
    a. I did not enjoy it at all
    b. I enjoyed it a little
    c. I enjoyed it somewhat
    d. I enjoyed it a lot
    e. I enjoyed it a great deal

17. In the space below, please draw a diagram of a mechanical clock.

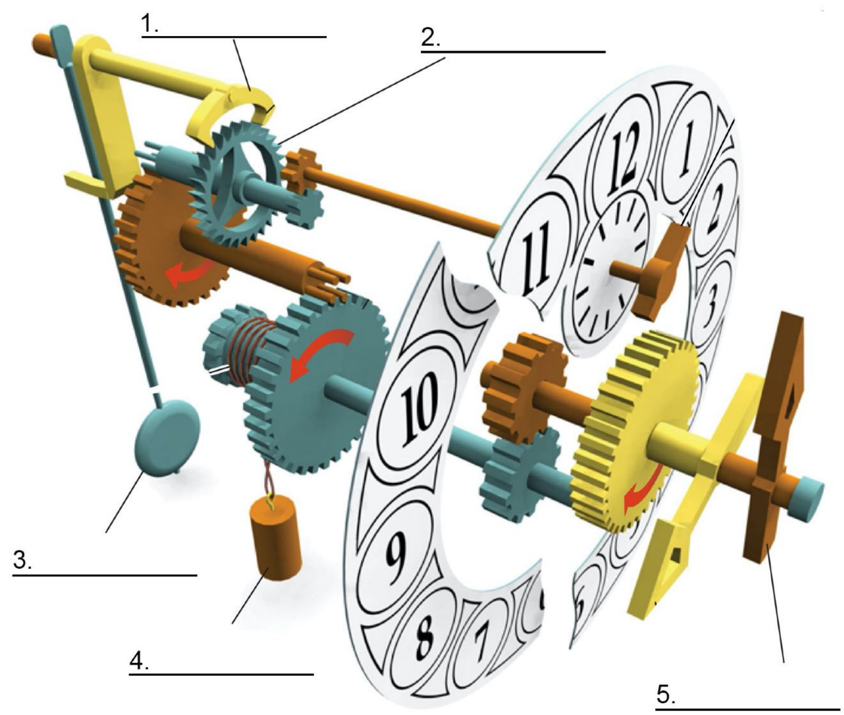
18. Please explain how this part of the clock works.



19. What is the name of the system shown in the picture below?

20. How confident are you that you could **explain** how this system works?
    a. Not at all confident
    b. Slightly confident
    c. Somewhat confident
    d. Confident
    e. Very confident

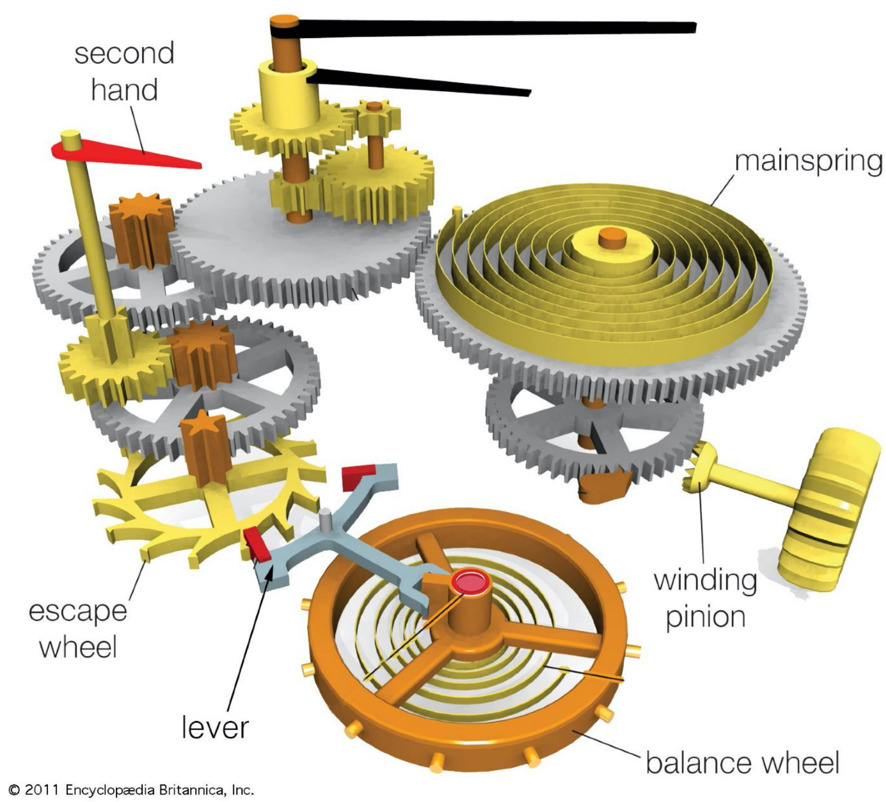
21. How confident are you that you **understand** how this system works?
    a. Not at all confident
    b. Slightly confident
    c. Somewhat confident
    d. Confident
    e. Very confident



22. Provide an explanation of what this system does and how it works. You can refer to each part using its name or its number in the diagram.

23. The picture below shows the inside of a mechanical watch. How confident are you that you could **explain** how this system works?
    a. Not at all confident
    b. Slightly confident
    c. Somewhat confident
    d. Confident
    e. Very confident

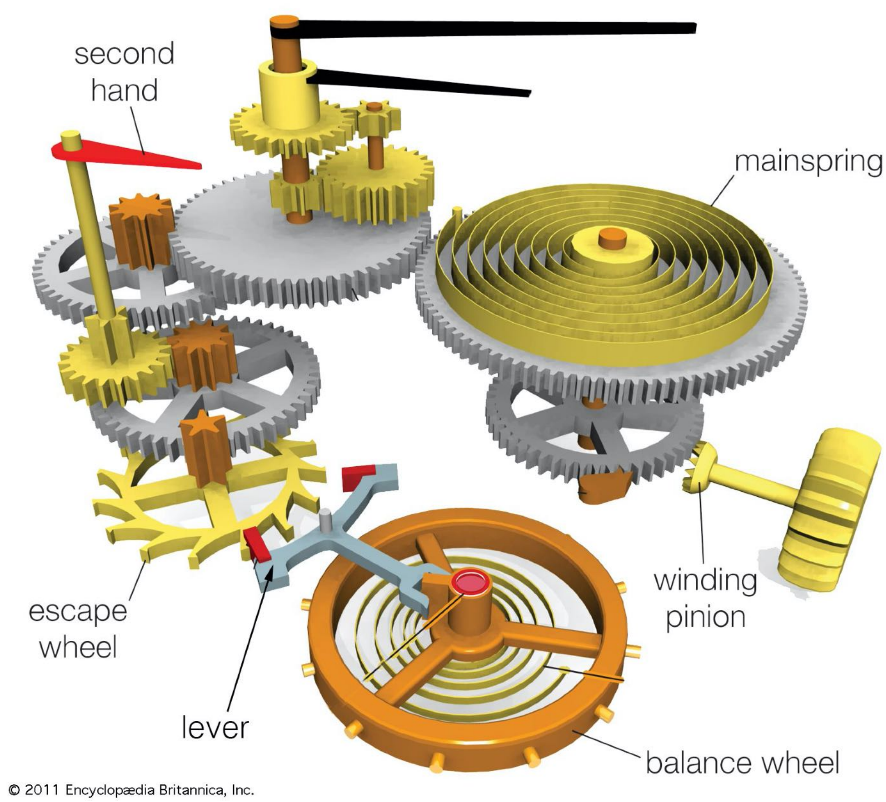
24. How confident are you that you could **figure out** how this system works?
    a. Not at all confident
    b. Slightly confident
    c. Somewhat confident
    d. Confident
    e. Very confident



25. Do your best to explain how you think this mechanical watch might work.

26. All of the parts in the mechanical clock have analogous parts in the mechanical watch. Draw lines connecting the parts from the mechanical clock to the analogous parts of the mechanical watch.

Mechanical Clock	**Mechanical Watch**
Pendulum	Second Hand
Weight	Mainspring
Escape Wheel	Balance Wheel
Anchor	Escape Wheel
Hand	Lever



27. Do your best to explain how you think this mechanical watch might work.

References

About Maker Faire. (n.d.). Retrieved June 28, 2020, from https://makerfaire.com/makerfairehistory/

About the Fab Foundation. (n.d.). Retrieved August 11, 2020, from https://fabfoundation.org/about/#about-intro

About Us. (n.d.). Retrieved August 3, 2020, from https://makered.org/about/

Anderson, C. (2012, October 2). *Makers: The New Industrial Revolution*. Crown.

Bamberger, J., & Schön, D. A. (1983). Learning as reflective conversation with materials: Notes from work in progress. *Art Education*, *36*(2), 68–73.

Bell, B. I. (1949). *Crisis in education: A challenge to American complacency*. Whittlesey House.

Bers, M. U., Flannery, L., Kazakoff, E. R., & Sullivan, A. (2014). Computational thinking and tinkering: Exploration of an early childhood robotics curriculum. *Computers & Education*, *72*, 145–157. https://doi.org/10.1016/j.compedu.2013.10.020

Berthold, K., Eysink, T. H. S., & Renkl, A. (2009). Assisting self-explanation prompts are more effective than open prompts when learning with multiple representations. *Instructional Science*, *37*(4), 345–363. https://doi.org/10.1007/s11251-008-9051-z

Bestor, A. E. (1955). *The Restoration of Learning: A Programm for Redeeming the Unfulfilled Promise of American Education*. New York: Knopf.

Bielaczyc, K., Pirolli, P. L., & Brown, A. L. (1995). Training in self-explanation and self-regulation strategies: Investigating the effects of knowledge acquisition activities on problem solving. *Cognition and instruction*, *13*(2), 221–252.

Blikstein, P. (2012). Bifocal modeling: A study on the learning outcomes of comparing physical and computational models linked in real time. *Proceedings of the 14th ACM International Conference on Multimodal Interaction*, 257–264. Retrieved November 10, 2014, from http://dl.acm.org/citation.cfm?id=2388729

Blikstein, P. (2013). Digital fabrication and 'making' in education: The democratization of invention. *FabLabs: Of Machines, Makers and Inventors* (pp. 1–21). Retrieved February 15, 2015, from http://www.academia.edu/download/30555959/Blikstein-2013-Making_The_Democratization_of_Invention.pdf

Blikstein, P., Kabayadondo, Z., Martin, A., & Fields, D. (2017). An Assessment Instrument of Technological Literacies in Makerspaces and FabLabs: Assessment of Technological Literacies in Makerspaces and FabLabs. *Journal of Engineering Education*, *106*(1), 149–175. https://doi.org/10.1002/jee.20156

Blikstein, P., & Krannich, D. (2013). The makers' movement and FabLabs in education: Experiences, technologies, and research. *Proceedings of the 12th International Conference on Interaction Design and Children - IDC '13*, 613. https://doi.org/10.1145/2485760.2485884

Buchholz, B., Shively, K., Peppler, K., & Wohlwend, K. (2014). Hands On, Hands Off: Gendered Access in Crafting and Electronics Practices. *Mind, Culture, and Activity*, *21*(4), 278–297. https://doi.org/10.1080/10749039.2014.939762

Buechley, L. (2006). A Construction Kit for Electronic Textiles. *2006 10th IEEE International Symposium on Wearable Computers*, 83–90. https://doi.org/10.1109/ISWC.2006.286348

Chase, C. C., Malkiewich, L., & S Kumar, A. (2019). Learning to notice science concepts in engineering activities and transfer situations. *Science Education*, *103*(2), 440–471.

Chi, M. T. H., & Bassok, M. (1989). Learning from examples via self-explanations. *Knowing, learning, and instruction: Essays in honor of Robert Glaser*, 251–282.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive science*, *13*(2), 145–182.

Chi, M. T. H., De Leeuw, N., Chiu, M.-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive science*, *18*(3), 439–477.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*(2), 121–152. https://doi.org/http://dx.doi.org/

Chin, C., & Brown, D. E. (2000). Learning in Science: A Comparison of Deep and Surface Approaches. *Journal of Research in Science Teaching*, *37*(2), 109–138. https://doi.org/10.1002/(SICI)1098-2736(200002)37:2<109::AID-TEA3>3.0.CO;2-7

Chinn, C. A., & Brewer, W. F. (1993). The Role of Anomalous Data in Knowledge Acquisition: A Theoretical Framework and Implications for Science Instruction. *Review of Educational Research*, *63*(1), 1–49. https://doi.org/10.3102/00346543063001001

Chu, S. L., Quek, F., Bhangaonkar, S., Ging, A. B., & Sridharamurthy, K. (2015). Making the Maker: A Means-to-an-Ends approach to nurturing the Maker mindset in elementary-aged children. *International Journal of Child-Computer Interaction*, *5*, 11–19. https://doi.org/10.1016/j.ijcci.2015.08.002

Clements, D. H., & Battista, M. T. (1989). Learning of Geometric Concepts in a Logo Environment. *Journal for Research in Mathematics Education*, *20*(5), 450–467. https://doi.org/10.2307/749420

Clements, D. H., & Gullo, D. F. (1984). Effects of computer programming on young children's cognition. *Journal of Educational psychology*, *76*(6), 1051.

Conati, C., & Vanlehn, K. (2000). Toward Computer-Based Support of Meta-Cognitive Skills: A Computational Framework to Coach Self-Explanation, 28.

Council, N. R. (2000). *How people learn: Brain, mind, experience, and school: Expanded edition*. National Academies Press.

Craver, C. F. (2006). When mechanistic models explain. *Synthese*, *153*(3), 355–376. https://doi.org/10.1007/s11229-006-9097-x

Cremin, L. A. (1961). The Transformation of the School: Progressivism in American Education, 1876-1957. *British Journal of Educational Studies*, *10*(1), 106–106.

Davis, R., Kafai, Y. B., Vasudevan, V., & Lee, E. (2013). The education arcade: Crafting, remixing, and playing with controllers for scratch games. *Proceedings of the 12th International Conference on Interaction Design and Children*, 439–442. Retrieved April 1, 2015, from http://dl.acm.org/citation.cfm?id=2485846

Dewey, J. (1901). *The place of manual training in the elementary course of study*.

Dewey, J. (1930, January 2). *The Quest for Certainty: A Study of the Relation of Knowledge and Action*.

Dewey, J. (1997). *How we think*. Courier Corporation. (Original work published 1910)

Dewey, J. (2004). *Democracy and education*. Courier Corporation. (Original work published 1916)

DiSessa, A. A. (2001). *Changing minds: Computers, learning, and literacy*. Mit Press.

Eberbach, C., Hmelo-Silver, C., Jordan, R., & Sinha, S. (2012). Multiple Trajectories for Understanding Ecosystems.

Engineering in K-12 Education: Understanding the Status and Improving the Prospects. (2009, September 8). National Academies Press. https://doi.org/10.17226/12635

FabLearn Supporters. (n.d.). Retrieved August 11, 2020, from https://fablearn.org/about/supporters/

Feurzeig, W. (2006). Educational technology at BBN. *IEEE Annals of the History of Computing*, *28*(1), 18–31. https://doi.org/10.1109/MAHC.2006.8

Fields, D. A., Searle, K. A., Kafai, Y. B., & Min, H. S. (2012). Debuggems to Assess Student Learning in e-Textiles (Abstract Only). *Proceedings of the 43rd ACM Technical*

Symposium on Computer Science Education, 699–699. https://doi.org/10.1145/2157136.2157367

Fosnot, C. T. (1993). Rethinking science education: A defense of Piagetian constructivism. *Journal of Research in Science Teaching*, *30*(9), 1189–1201.

Fosnot, C. T., & Perry, R. S. (1996). Constructivism: A psychological theory of learning. *Constructivism: Theory, perspectives, and practice*, 8–33. Retrieved January 20, 2016, from http://randallsperry.com/fosnotandperry.pdf

Franklin, D., Conrad, P., Boe, B., Nilsen, K., Hill, C., Len, M., Dreschler, G., Aldana, G., Almeida-Tanaka, P., & Kiefer, B. (2013). Assessment of computer science learning in a scratch-based outreach program. *Proceeding of the 44th ACM Technical Symposium on Computer Science Education*, 371–376. Retrieved May 11, 2014, from http://dl.acm.org/citation.cfm?id=2445304

Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Gabadinho, A., Ritschard, G., Mueller, N. S., & Studer, M. (2011). Analyzing and Visualizing State Sequences in R with TraMineR. *Journal of Statistical Software*, *40*(4), 1–37. Retrieved November 6, 2016, from http://archive-ouverte.unige.ch/unige:16809?gathStatIcon=true

Gabrielson, C. (2013). *Tinkering: Kids learning by making stuff.*

Galaleldin, M., Bouchard, F., Anis, H., & Lague, C. (2016). The Impact of Makerspaces on Engineering Education. *Proceedings of the Canadian Engineering Education Association (CEEA)*. https://doi.org/10.24908/pceea.v0i0.6481

Garneli, B., Giannakos, M. N., Chorianopoulos, K., & Jaccheri, L. (2013). Learning by Playing and Learning by Making. In M. Ma, M. F. Oliveira, S. Petersen, & J. B. Hauge (Eds.), *Serious Games Development and Applications* (pp. 76–85). Springer. https://doi.org/10.1007/978-3-642-40790-1_8

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology*, *95*(2), 393.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*(1), 1–38. https://doi.org/10.1016/0010-0285(83)90002-6

Glennan, S. (2017). *The new mechanical philosophy* (First edition). Oxford University Press.

Goel, A. K., Rugaber, S., & Vattam, S. (2009). Structure, behavior, and function of complex systems: The structure, behavior, and function modeling language. *Ai Edam*, *23*(1), 23–35.

Gray, S. H. (1990). Using protocol analyses and drawings to study mental model construction during hypertext navigation. *International Journal of Human-Computer Interaction*, *2*(4), 359–378.

Halasz, F. G., & Moran, T. P. (1983). Mental models and problem solving in using a calculator. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '83*, 212–216. https://doi.org/10.1145/800045.801613

Hardiman, P. T., Dufresne, R., & Mestre, J. P. (1989). The relation between problem categorization and problem solving among experts and novices. *Memory & Cognition*, *17*(5), 627–638. https://doi.org/10.3758/BF03197085

Harnett, C. K., Tretter, T. R., & Philipp, S. B. (2014). Hackerspaces and engineering education. *2014 IEEE Frontiers in Education Conference (FIE) Proceedings*, 1–8. https://doi.org/10.1109/FIE.2014.7044395

Hartry, A., Werner-Avidon, M., Hsi, S., Ortiz, A. J., & Quigley, K. C. (2018). TechHive: Team-based, Real-world Engineering Challenges for Teens.

Hausmann, R. G., & Chi, M. T. H. (2002). Can a computer interface support self-explaining. *Cognitive Technology*, *7*(1), 4–14.

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d'. *Behavior Research Methods, Instruments, & Computers*, *27*(1), 46–51.

Helms, M., Vattam, S., & Goel, A. (2010). The effects of functional modeling on understanding complex biological systems. *In Proc. ASME IDETC/CIE Conference on Design Theory and Methods*, 107–115.

Hlubinka, M., Dougherty, D., Thomas, P., Chang, S., Hoefer, S., Alexander, I., & McGuire, D. (2013). Makerspace playbook: School edition. *Retrieved from Maker Media website: http://makerspace. com/wpcontent/uploads/2013/02/MakerspacePlaybook-Feb2013. pdf.*

Hmelo-Silver, C. E., Marathe, S., & Liu, L. (2007). Fish Swim, Rocks Sit, and Lungs Breathe: Expert-Novice Understanding of Complex Systems. *Journal of the Learning Sciences*, *16*(3), 307–331. https://doi.org/10.1080/10508400701413401

Houchins, J., Tatar, C., & Oliver, K. (2020). Computational Thinking and other Emergent Practices in Student Reflections on Makerspace Activities, 29–37. Retrieved June 8, 2020, from https://www.learntechlib.org/primary/p/215727/

Illari, P. M., & Williamson, J. (2012). What is a mechanism? Thinking about mechanisms across the sciences. *European Journal for Philosophy of Science*, *2*(1), 119–135. https://doi.org/10.1007/s13194-011-0038-2

Jee, B. D., Uttal, D. H., Spiegel, A., & Diamond, J. (2015). Expert–novice differences in mental models of viruses, vaccines, and the causes of infectious disease. *Public Understanding of Science*, *24*(2), 241–256. https://doi.org/10.1177/0963662513496954

Jonassen, D. H., & Strobel, J. (2006). Modeling for meaningful learning. *Engaged learning with emerging technologies* (pp. 1–27). Springer.

Jonassen, D. H., Strobel, J., & Lee, C. B. (2006). Everyday Problem Solving in Engineering: Lessons for Engineering Educators. *Journal of Engineering Education*, *95*(2), 139–151. https://doi.org/10.1002/j.2168-9830.2006.tb00885.x

Kafai, Y. B., Fields, D. A., & Searle, K. A. (2014). Electronic Textiles as Disruptive Designs: Supporting and Challenging Maker Activities in Schools. *Harvard Educational*

Review, *84*(4). Retrieved December 29, 2015, from http://search.proquest.com/docview/1642662289/abstract

Kafai, Y. B., Lee, E., Searle, K., Fields, D., Kaplan, E., & Lui, D. (2014). A Crafts-Oriented Approach to Computing in High School: Introducing Computational Concepts, Practices, and Perspectives with Electronic Textiles. *Trans. Comput. Educ.*, *14*(1), 1:1–1:20. https://doi.org/10.1145/2576874

Kafai, Y. B., Searle, K., Kaplan, E., Fields, D., Lee, E., & Lui, D. (2013). Cupcake Cushions, Scooby Doo Shirts, and Soft Boomboxes: E-Textiles in High School to Promote Computational Concepts, Practices, and Perceptions. *Proceeding of the 44th ACM Technical Symposium on Computer Science Education*, 311–316. https://doi.org/10.1145/2445196.2445291

Kafai, Y. B., & Vasudevan, V. (2015). Hi-Lo tech games: Crafting, coding and collaboration of augmented board games by high school youth. *Proceedings of the 14th International Conference on Interaction Design and Children*, 130–139.

Kahneman, D. (2011, October 25). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.

Keil, F. C., Levin, D. T., Richman, B. A., & Gutheil, G. (1999). Mechanism and explanation in the development of biological thought: The case of disease. *Folkbiology*, 285–319.

Kirsh, D. (2009). Problem Solving and Situated Cognition. *The Cambridge Handbook of Situated Cognition*, 264–306.

Lajoie, S. P., & Lesgold, A. M. (1992). Dynamic Assessment of Proficiency for Solving Procedural Knowledge Tasks. *Educational Psychologist*, *27*(3), 365–384. https://doi.org/10.1207/s15326985ep2703_6

Lane, H. C., Cahill, C., Foutz, S., Auerbach, D., Noren, D., Lussenhop, C., & Swartout, W. (2013). The Effects of a Pedagogical Agent for Informal Science Education on Learner Behaviors and Self-efficacy. In H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik (Eds.). D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos,

D. Tygar, M. Y. Vardi, & G. Weikum (**typeredactors**), *Artificial Intelligence in Education* (pp. 309–318). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-39112-5_32

Lee, C. B., Jonassen, D. H., & Teo, T. (2011). The role of model building in problem solving and conceptual change. *Interactive Learning Environments*, *19*(3), 247–265. https://doi.org/10.1080/10494820902850158

Litts, B. K., Kafai, Y. B., Lui, D., Walker, J., & Widman, S. (2017). Understanding High School Students' Reading, Remixing, and Writing Codeable Circuits for Electronic Textiles. *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*, 381–386. https://doi.org/10.1145/3017680.3017740

Litts, B. K., Kafai, Y. B., Lui, D. A., Walker, J. T., & Widman, S. A. (2017). Stitching Codeable Circuits: High School Students' Learning About Circuitry and Coding with Electronic Textiles. *Journal of Science Education and Technology*, *26*(5), 494–507. https://doi.org/10.1007/s10956-017-9694-0

Looking Back at the Launch of the Maker Movement: The First Maker Faire | Make: (2015, April 16). Retrieved June 28, 2020, from https://makezine.com/2015/04/16/looking-back-launch-maker-movement-first-maker-faire/

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about Mechanisms. *Philosophy of Science*, *67*(1), 1–25. https://doi.org/10.1086/392759

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2016). *Cluster: Cluster Analysis Basics and Extensions.*

Maker Ed's Sponsors. (n.d.). Retrieved August 11, 2020, from https://makered.org/about/supporters/

Maker Education Initiative Launches at Maker Faire. (n.d.). Retrieved August 2, 2020, from https://makered.org/blog/maker-education-initiative-launches-at-maker-faire/

Martin, L. (2015). The Promise of the Maker Movement for Education. *Journal of Pre-College Engineering Education Research (J-PEER)*, *5*(1). https://doi.org/10.7771/2157-9288.1099

Melo, M. M., & Rabkin, A. (2019). Makerspaces in US State Colleges and Universities.

Miłkowski, M. (2013). A mechanistic account of computational explanation in cognitive science. *Cooperative Minds: Social Interaction and Group Dynamics. Proceedings of the 35th Annual Meeting of the Cognitive Science Society, Eds M. Knauff, M. Pauen, N. Sebanz, and I. Wachsmuth (Austin, TX: Cognitive Science Society)*, 3050–3055.

Millner, A., & Daily, S. B. (2008). Creating an Educational Ecosystem for Design, Personal Fabrication, and Invention. *Communities of Practice Vol. 1: Creating Learning Environments for Educators* (p. 403). IAP.

Mulder, Y. G., Bollen, L., de Jong, T., & Lazonder, A. W. (2016). Scaffolding learning by modelling: The effects of partially worked-out models. *Journal of research in science teaching*, *53*(3), 502–523.

Newell, A., & Simon, H. A. (1972). *Human problem solving* (Vol. 104). Prentice-Hall Englewood Cliffs, NJ.

Norman, D. A. (2014). Some observations on mental models. *Mental models* (pp. 15–22). Psychology Press.

O'Reilly's Make and Otherlab Win DARPA MENTOR Award to Bring Making to Education: Program Will Help Establish Makerspaces in High Schools. (n.d.). Retrieved June 28, 2020, from https://www.oreilly.com/pub/pr/2962

Papavlasopoulou, S., Giannakos, M. N., & Jaccheri, L. (2017). Empirical studies on the Maker Movement, a promising approach to learning: A literature review. *Entertainment Computing*, *18*, 57–78. https://doi.org/10.1016/j.entcom.2016.09.002

Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas.* Basic Books.

Papert, S. (1986). *Constructionism: A new opportunity for elementary science education.* Massachusetts Institute of Technology, Media Laboratory, Epistemology and ...

Papert, S., & Harel, I. (1991). Situating constructionism. *Constructionism*, *36*(2), 1–11.

Patton, R. M., & Knochel, A. D. (2017). Meaningful makers: Stuff, sharing, and connection in STEAM curriculum. *Art Education*, *70*(1), 36–43.

Pea, R. D. (1987). Logo programming and problem solving. Retrieved May 21, 2015, from https://telearn.archives-ouvertes.fr/hal-00190546/document

Pea, R. D., & Kurland, D. M. (1983). On the Cognitive Prerequisites of Learning Computer Programming. Retrieved August 2, 2020, from https://telearn.archives-ouvertes.fr/hal-00190531

Pea, R. D., & Kurland, D. M. (1984). On the cognitive effects of learning computer programming. *New Ideas in Psychology*, *2*(2), 137–168. https://doi.org/10.1016/0732-118X(84)90018-7

Pea, R. D., Kurland, D. M., & Hawkins, J. (1985). Logo and the development of thinking skills. *Children and microcomputers: Research on the newest medium*, 193–317. Retrieved November 4, 2013, from http://telearn.archives-ouvertes.fr/docs/00/19/05/34/PDF/A23_Pea_etal_85.pdf

Peirce, C. S. (1878). How to make our ideas clear. *The nature of truth: Classic and contemporary perspectives*, *2001*, 193–209.

Peppler, K., & Glosson, D. (2013). Stitching Circuits: Learning About Circuitry Through E-textile Materials. *Journal of Science Education and Technology*, *22*(5), 751–763. https://doi.org/10.1007/s10956-012-9428-2

Peppler, K., Maltese, A., Keune, A., Chang, S., & Regalla, L. (2015). Survey of Makerspaces, Part I, 5.

Polanyi, M. (2005). *Personal knowledge: Towards a post-critical philosophy*. Retrieved August 11, 2018, from https://ebookcentral.proquest.com/lib/concordiaab-ebooks/detail.action?docID=179903

Qiu, K., Buechley, L., Baafi, E., & Dubow, W. (2013). A Curriculum for Teaching Computer Science Through Computational Textiles. *Proceedings of the 12th International*

Conference on Interaction Design and Children, 20–27. https://doi.org/10.1145/2485760.2485787

Rasmussen, J., & Jensen, A. (1974). Mental Procedures in Real-Life Tasks: A Case Study of Electronic Trouble Shooting. *Ergonomics*, *17*(3), 293–307. https://doi.org/10.1080/00140137408931355

Reiner, M., Pea, R. D., & Shulman, D. J. (1995). Impact of simulator-based instruction on diagramming in geometrical optics by introductory physics students. *Journal of Science Education and Technology*, *4*(3), 199–226.

Resnick, M., Maloney, J., Monroy-Hernández, A., Rusk, N., Eastmond, E., Brennan, K., Millner, A., Rosenbaum, E., Silver, J., Silverman, B., et al. (2009). Scratch: Programming for all. *Communications of the ACM*, *52*(11), 60–67. Retrieved March 26, 2015, from http://dl.acm.org/citation.cfm?id=1592779

Resnick, M., & Silverman, B. (2005). Some reflections on designing construction kits for kids. *Proceedings of the 2005 Conference on Interaction Design and Children*, 117–122.

Robertson, S. I. (2003). *Problem solving*. Psychology Press.

Russ, R. S., Scherr, R. E., Hammer, D., & Mikeska, J. (2008). Recognizing mechanistic reasoning in student scientific inquiry: A framework for discourse analysis developed from philosophy of science. *Science Education*, *92*(3), 499–525. https://doi.org/10.1002/sce.20264

SageModeler – Systems Modeling Tool. (n.d.). Retrieved July 19, 2019, from https://sagemodeler.concord.org/

Schneider, B., Wallace, J., Blikstein, P., & Pea, R. (2013). Preparing for Future Learning with a Tangible User Interface: The Case of Neuroscience. *IEEE Transactions on Learning Technologies*, *6*(2), 117–129. https://doi.org/10.1109/TLT.2013.15

Schneider, B., Bumbacher, E., & Blikstein, P. (2015). Discovery Versus Direct Instruction: Learning Outcomes of Two Pedagogical Models Using Tangible Interfaces.

Searle, K. A., Fields, D. A., Lui, D. A., & Kafai, Y. B. (2014). Diversifying high school students' views about computing with electronic textiles. *Proceedings of the Tenth Annual Conference on International Computing Education Research*, 75–82.

Sheppard, S., Colby, A., Macatangay, K., & Sullivan, W. (2006). What is engineering practice? *International Journal of Engineering Education*, *22*(3), 15.

Sheridan, K. M., Halverson, E. R., Litts, B. K., Brahms, L., Jacobs-Priebe, L., & Owens, T. (2014). Learning in the making: A comparative case study of three makerspaces. *Harvard Educational Review*, *84*(4), 505–531. Retrieved March 20, 2015, from http://her.hepg.org/index/BRR34733723J648U.pdf

Sipitakiat, A., Blikstein, P., & Cavallo, D. (2002). The GoGo Board: Moving towards highly available computational tools in learning environments. *Proceedings of Interactive Computer Aided Learning International Workshop*. Retrieved February 15, 2015, from http://www.blikstein.com/paulo/documents/papers/SipitakiatBlikstein-GoGoBoard-ICL2002.pdf

Smith, P. A. M., Leeman-Munk, S., Shelton, A., Mott, B. W., Wiebe, E., & Lester, J. (2019). A Multimodal Assessment Framework for Integrating Student Writing and Drawing in Elementary Science Learning. *IEEE Transactions on Learning Technologies*. https://doi.org/10.1109/tlt.2018.2799871

Solomon, C., Harvey, B., Kahn, K., Lieberman, H., Miller, M. L., Minsky, M., Papert, A., & Silverman, B. (2020). History of Logo. *Proceedings of the ACM on Programming Languages*, *4*, 1–66. https://doi.org/10.1145/3386329

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and brain sciences*, *23*(5), 645–665.

Telhan, O., Kafai, Y. B., Davis, R. L., Steele, K., & Adleberg, B. M. (2014). Connected messages: A maker approach to interactive community murals with youth. *Proceedings of the 2014 Conference on Interaction Design and Children*, 193–196. Retrieved April 1, 2015, from http://dl.acm.org/citation.cfm?id=2610450

The third industrial revolution [magazine]. (2012). *The Economist*. Retrieved June 28, 2020, from http://www.economist.com/leaders/2012/04/21/the-third-industrial-revolution

Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making*, *11*(1), 15.

Timotheou, S., & Ioannou, A. (2019). On Making, Tinkering, Coding and Play for Learning: A Review of Current Research. In D. Lamas, F. Loizides, L. Nacke, H. Petrie, M. Winckler, & P. Zaphiris (Eds.), *Human-Computer Interaction – INTERACT 2019* (pp. 217–232). Springer International Publishing. https://doi.org/10.1007/978-3-030-29384-0_14

Togou, M. A., Lorenzo, C., Cornetta, G., & Muntean, G.-M. (2020). Assessing the Effectiveness of Using Fab Lab-Based Learning in Schools on K–12 Students' Attitude Toward STEAM. *IEEE Transactions on Education*, *63*(1), 56–62.

Troxler, P. (2013). Making the third industrial revolution—the struggle for polycentric structures and a new peer-production commons in the FabLab community. *FabLab: Of machines, makers and inventors* (pp. 181–198). Transcript Bielefeld.

Tschan, F. (2002). Ideal Cycles of Communication (or Cognitions) in Triads, Dyads, and Individuals. *Small Group Research*, *33*(6), 615–643. https://doi.org/10.1177/1046496402238618

van der Meij, J., & de Jong, T. (2011). The effects of directive self-explanation prompts to support active processing of multiple representations in a simulation-based learning environment. *Journal of Computer Assisted Learning*, *27*(5), 411–423. https://doi.org/10.1111/j.1365-2729.2011.00411.x

VanLehn, K., Jones, R. M., & Chi, M. T. (1992). A model of the self-explanation effect. *The journal of the learning sciences*, *2*(1), 1–59.

Vattam, S. S., Goel, A. K., Rugaber, S., Hmelo-Silver, C. E., Gray, S., & Sinha, S. (2011). Understanding Complex Natural Systems by Articulating Structure-Behavior-Function Models, 17.

Vihavainen, A., Miller, C. S., & Settle, A. (2015). Benefits of Self-explanation in Introductory Programming. *Proceedings of the 46th ACM Technical Symposium on Computer Science Education*, 284–289. https://doi.org/10.1145/2676723.2677260

Vossoughi, S., & Bevan, B. (2014). Making and Tinkering: A Review of the Literature, 56.

What is Engineering Design? (n.d.). Retrieved August 11, 2020, from https://www.engineergirl.org/128119/engineering-design

Wilensky, U., & Resnick, M. (1999). Thinking in levels: A dynamic systems approach to making sense of the world. *Journal of Science Education and technology*, *8*(1), 3–19.

Wilkerson, M. H., Shareff, R., Laina, V., & Gravel, B. (2018). Epistemic gameplay and discovery in computational model-based inquiry activities. *Instructional Science*, *46*(1), 35–60. https://doi.org/10.1007/s11251-017-9430-4

Wissehr, C., Concannon, J., & Barrow, L. H. (2011). Looking Back at the Sputnik Era and Its Impact on Science Education. *School Science and Mathematics*, *111*(7), 368–375. https://doi.org/10.1111/j.1949-8594.2011.00099.x

Woodward, C. M. (1887). *The manual training school, comprising a full statement of its aims, methods, and results,with figured drawings of shop exercises in woods and metals.* Boston, http://hdl.handle.net/2027/mdp.39015074760466

Worsley, M., & Blikstein, P. (2013). Towards the development of multimodal action based assessment. *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, 94–101. Retrieved November 18, 2016, from http://dl.acm.org/citation.cfm?id=2460315

Worsley, M., & Blikstein, P. (2014). The Impact of Principle-Based Reasoning on Hands-on, Project-Based Learning. Retrieved June 30, 2020, from https://repository.isls.org//handle/1/952

Wylie, R., & Chi, M. T. H. (2014). The Self-Explanation Principle in Multimedia Learning.

Yu, Y., Hmelo-Silver, C. E., Jordan, R., Eberbach, C., & Sinha, S. (2016). Promoting Transfer of Ecosystems Concepts. *International Journal of Environmental and Science Education*, *11*(18), 11127–11148. Retrieved April 23, 2019, from https://eric.ed.gov/?id=EJ1120593