MAKING WITH UNDERSTANDING:

RESEARCH ON STUDIES FROM A CONSTRUCTIONIST LEARNING ENVIRONMENT

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

OF EDUCATION AND THE COMMITTEE ON

GRADUATE STUDIES

OF STANFORD UNIVERSITY

IN PARITIAL FULFILLMENT OF THE

REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Marcelo Worsley

August 2014

Abstract

"Making" has become a 21st century education buzzword. Educators, entrepreneurs, researchers and politicians all seem to galvanize around the possibility for "making" to launch an educational revolution. However, to date, there has been little research that systematically investigates what and how students are learning in the context of "making." This dissertation serves to shed light on these questions, while also providing recommendations for how to improve the culture and practices of "making." Specifically I use data from three studies to document common strategies students use in constructionist learning, and demonstrate how those strategies result in differences in project quality, student learning, and the overall process in which students engage.

Preface

This dissertation is being written amidst a national shift towards promoting "making." After Dale Dougherty, founder, President and CEO of Maker Media, declared 2014 as the "Year of the Maker," President Barack Obama followed suit by issuing a proclamation that June 18 would henceforth be the National Day of Making. This announcement was made during the first ever White House Maker Faire which drew hundreds of "makers" from around the country to the venerated building that serves as one of the icons for the United States of America. Needless to say, there has been a tremendous amount of momentum, press and funding being directed towards expanding the practices of "making." Given the contemporary fascination with "making," this dissertation serves as a foray into some of the practical and theoretical factors associated with the current Maker Movement. More specifically, this dissertation can be viewed as a resource for teachers, practitioners and researchers who are currently using, or contemplating the use of "making" in their learning environments. It aims to offer practical, yet theoretically-grounded, and experimentally-justified, suggestions for studying and improving the use of "making" in educational settings. Because the intended audience spans both practitioners and researchers, the discussion involves both practical suggestions and in-depth descriptions of research methods and techniques. In recognizing the diversity of readers that may engage this text, I encourage readers to glean from the parts that they deem most applicable to their interests, but to also challenge themselves to venture into areas that may seem uncomfortable or unknown.

Table of Contents

List of Tables

List of Figures

Introduction.

This dissertation is situated around suggestions for improving "making." However, one cannot justifiable participate in the discussion of "making" without first considering the history and practices of constructionism.

Constructionism

Constructionism is a pedagogical approach based on the possibilities to make physical and computational artifacts. More specifically, Papert (1987a) defines Constructionism as:

> [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. (Abstract)

A later piece, Papert & Harel (1991), provides a complementary definition[1]:

> Constructionism--the N word as opposed to the V word--shares constructivism's connotation of learning as "building knowledge structures" irrespective of the circumstances of the learning. It then adds the idea that this happens especially felicitously in a context where the learner is consciously

[1] In the following excerpt the "V word" refers to constructivism, the "N word" refers to constructionism.

> engaged in constructing a public entity, whether it's a sand castle on the beach or a theory of the universe. (p. 1)

These quotations describe constructionism as being the idea that knowledge is constructed as learners interact with the world in ways that allow them to create personally meaningful, shareable objects. While Papert coined the term, constructionism was shaped by a deep history of progressive educational advocates who contributed to the emergence of self-directed, personally meaningful, hands-on, technology-enhanced, student-centered learning. I do not undertake to recount this history in whole, but will, instead, provide glimpses of several of the individuals and technologies involved.

Influential Scholars in Progressive Education

Jean-Jacque Rousseau[2]. The first author that I discuss is Jean-Jacque Rousseau, whose influential text, *Émile*, provided a strong contrast to the traditional view of education, which involved students memorizing content without ample justification or use for that knowledge. Rousseau advocated for an education in which a student is free to learn through self-directed inquiry, and in such a way that their learning is directly applicable to their everyday experiences. Moreover, *Émile* describes a student learning through their senses, which provide natural connections to science and mathematics, important disciplines at the time.

[2] By starting with Rousseau I am not overlooking the fact that prior works by of Hobbes, Plato, Locke, Nietzsche, Machiavelli and several other philosophers greatly influenced Rousseau's writing. I have simply determined that for the sake of brevity, and based on the direct connection between Rousseau's work and constructionism, that this is an appropriate starting point.

Friedrich Froebel. Froebel's contribution to the creation of kindergarten, both in terminology and in practice, has had a profound impact on the fabric of education and on creative luminaries such as Frank Lloyd Wright. Furthermore, Papert, and many of his predecessors were deeply inspired by Froebel's emphasis on hands-on, free play, and the "gifts" used to enable student-centered exploration.

Maria Montessori. Montessori, in building on the work of Froebel, proposed a somewhat more liberal model of learning that stressed even greater student individuality, the development of the senses, and engaging students in a wider variety of authentic, as opposed to simply imaginative, practices.

John Dewey. Dewey has a long list of influential texts that span the areas of philosophy, democracy, psychology and education. Among his texts on education, Dewey described the importance of hands-on learning and individual interests. For example, Dewey (1913) brings to the forefront the shortcomings of having students complete tasks that lack personal significance and that feel mechanical. Promoting an education system that leveraged students' individual interests was one of the ways that Dewey helped advance progressive education.

Lev Vygotsky. Vygotsky is typically most well-known for the theory of the "zone of proximal development." At the heart of this theory is the idea that individual development is influenced by various sociocultural factors. In a similar vein, Papert's constructionism recognizes that the social and cultural practices that surround a student influences their rate of development. Furthermore, he argues that the materials that one has access to can have a significant impact on learning and development – which echoes some of Vygotsky's claims about the role of language in thought.

Jean Piaget. Piaget was one of Papert's mentors. Hence, Piaget's views on constructivism, the idea that individuals construct knowledge based on their prior experiences, as opposed to absorbing knowledge into a tabula rasa, is one of the central tenets of constructionism. Additionally, Papert's view on the role of invention in the context of science and mathematics education bears great similarity to Piaget's idea that inventing is a form, or demonstration, of understanding (Piaget 1973).

While one cannot discount the influence of various other contemporary and historical cognitive psychologists, the aforementioned are among those whom Papert directly attributes many of his ideas for constructionism.[3]

The Rise of Technology in Education

Even amidst the contributions made by his predecessors, constructionism was also greatly enabled by a growing interest in using computers to support student learning that took place during the 1980s. The availability of relatively inexpensive microcomputers burgeoned a widespread proliferation in the ways that educators and researchers sought to make use of computers. Again, I will not exhaustively present every pertinent development that took place during that time, but will highlight a few in order to provide a loose historical portrait for the reader.

[3] Papert also makes references to other notable individuals such as Claude Levi-Strauss, Andrea diSessa, Marvin Minsky, Jerome Bruner and Rene Descartes, to name a few. For more information about these individuals' contributions to constructionism, see Papert (1980).

Intelligent Tutoring Systems. Intelligent Tutoring Systems (ITS) represent a type of computer-assisted instructional technology that was growing in popularity during the 1980s. These systems were motivated by prior work from the 1950s on Computer Assisted Instructors (CAI). CAI systems took a behaviorist view of learning and presented students with several problems to be solved. As the students worked through those problems, they would be assisted with small hints that would make salient mistakes in their computations or strategies, for example. ITS differed from CAI in that ITS included better models of human reasoning strategies that could more accurately pinpoint the nature of student errors.

Computer-based Learning Environments. The other paradigm of educational technology that was prevalent during the 1980s was Computer-based Learning Environments. These environments were much more exploration-oriented than ITS and described by Lawler as:

> [A]n environment in which other people can exercise their own creativity. This focus on sub-creation makes computer-based microworlds specifically useful for education. Sub-creation permits a collaboration between the didactic, structure intentions of society and the inescapable fact that people learn by their own choice; further, it recognizes that learning most frequently derives from people expressing what is important to them within the confines of a medium which both enables and constrains their expression. (Lawler, 1987, p. 14)

Whereas the design of ITS has clear connections to behaviorism, microworlds, as described above, are a hallmark of Piagetian constructivism. Microworlds were seen as an important tool to make scientific and mathematical concepts and their relationships more visible. For example, many of the early examples of LOGO usage provided students with a new frame for doing mathematical computations. But Papert is quick to note that microworlds are not a new idea, but are analogous to how infants, children and adults learn. For example, construction kits and blocks represent microworlds that many children utilize while growing up. Instead, Papert suggests that what's new is

> [T]he possibilities of microworlds that can be made from the computer are vast, beyond anything that one could do with any other materials. So the computer has opened up a new technology of being able to do things that are not so different in themselves – but in terms of how much you can do with it. (Papert, 1987b, p. 92)

Relative to CAI, Computer-based Learning Environments, like virtual microworlds, represented a new paradigm that enabled students to be in control of the computer in ways that were empowering and afforded personal expression.

LOGO. LOGO is the programming language that Papert and his colleagues first developed to allow students to interact with computer-based microworlds (Papert, 1987b). The language is a variant of Lisp, and centers around a turtle that moves based on the commands provided. LOGO was eventually developed into several other variants that added capability for concurrent programming (MultiLogo) (Resnick, 1990), complex and dynamic systems (StarLogo) (Resnick, 1996), agent-based modeling and networking (NetLogo) (Tisue & Wilensky, 2004), object-oriented programming (ObjectLogo) (Drescher, 1987) and also impacted later programming environments like Scratch (Maloney et al., 2004).

Programmable Bricks. The 1980s also saw the birth of programmable bricks. Fred Martin, Seymour Papert and Mitch Resnick, worked to combine LOGO programming with Lego blocks in a way that would allow students to not only control virtual microworlds, but to also create robots with sensors and actuators that could intelligibly interact with the world . This technology would eventually become the basis for LEGO Mindstorm, and would give rise to a community of inventors and do-it-yourself builders that designed and created new technological devices. Within this trend, several lower-cost variants (GoGo Board (Sipitakiat, Blikstein, & Cavallo, 2002), PicoCricket, Arduino, MakeyMakey (Silver, Rosenbaum & Shaw, 2012)) were made available that looked to extend beyond the capabilities of the LEGO Mindstorms (Lawler & Yazdani, 1987; Martin & Resnick, 1993; Martin, 1988; Papert, 1987b; Mitchel Resnick, Berg, & Eisenberg, 2000).

The technology advances of the late 1900s created an environment that fostered significant optimism in the area of technology and education. As such, constructionism converged on prior work that emphasized, self-directed, interest-driven, personally-meaningful learning, and the affordances of computer technology, in order to offer an alternative to the contemporary emphasis of computer-assisted instructional technologies. The pedagogical strategy, software, and hardware associated with constructionism, coupled with the availability of low-cost digital fabrication technology and Fablabs allowed for the development and expansion of the current label, “making” (Blikstein & Krannich, 2013; Blikstein, 2013).

Improving Making

As previously noted, this dissertation is centered on suggestions for improving the quality of "making." These suggestions are based on three studies that were designed following the constructionist paradigm.

To begin this discussion for improving "making," Chapter 1 describes the current state of "making" in education, and identifies various dimensions of the contemporary Maker Movement that deserve greater attention and analysis. Chapter 2 then transitions into the first suggestion for studying and improving "making." This first suggestion is towards the use of student reflection to identify how students' change over time. However, instead of simply proposing student reflection as a general practice, I focus on having students reflect on their strategy for solving or completing a given project. In order to substantiate this suggestion I first identify a series of common strategies that students utilize. I then draw upon prior literature and a combination of studies to justify why studying student reasoning strategies has relevance for documenting and promoting student development. As part of this justification I show that the different approaches students use have an impact on the success of their projects, and that this difference emerges across multiple learning contexts. Furthermore, I show how to identify student strategies through a pair of coding schemes. This first suggestion, albeit coarse in granularity, provides a tractable way to examine and characterize student development in Makerspaces and FabLabs.

The second suggestion, which is the focus of Chapter 3, is to look for evidence of student development from a multimodal perspective. Taking a multimodal perspective involves looking beyond success as the primary metric, and instead, searching among student actions, and the context of those actions, to identify student development. To validate this suggestion I present results from several analyses of student process data. Specifically, I show that adopting *principle-based reasoning* (Worsley & Blikstein, 2014) strategies is evidenced across the modalities of speech, action, gestures, and stress[4]. The analyses for this suggestion are grounded in a combination of qualitative and computational approaches. For ease of comprehension, I present the results from these analyses in Chapter 3, but withhold the details of the learning analytics and data mining techniques for Appendix A.

The final chapter of the dissertation includes a discussion of the implication and limitations of this work. The intent of this discussion is to ensure that readers have an appropriate awareness of the scope and applicability of the findings. Furthermore, this final chapter discusses the results with regard to prior literature from the learning sciences, and provides various interpretations of what these results may suggest about student learning in constructionist learning environments.

As its proponents would agree, there is significant potential for constructionism and "making" to have an increasingly important role in student learning. With the variety of tools currently available to support low-risk, experiential learning through digital fabrication and invention, one can reasonably expect to see changes in how and what students learn in both formal and informal environments.

[4] Stress is a modality that I am able to infer from electro-dermal activation sensors. Sensors for measuring electro-dermal will be discussed in more detail in Chapter 3 and in Appendix A.

However, as a part of supporting the growth of constructivist and constructionist-inspired pedagogies, the “maker” community needs to move beyond the essentialist view that making is fundamentally good and move towards a more concerted effort to study: (1) how students learn when they are “making;” (2) metrics that can better measure what students are learning; and (3) how to improve the learning experience. This dissertation is written to contribute to those conversations, and serve as a resource for researchers and practitioners interested in joining in on these discussions.

Chapter 1. The Current State of Making in Education

Over the past five years there has been an explosive increase in organizations and individuals interested in “making”, tinkering[5] and constructionism. While each of these terms have individual nuances, histories, and definitions, the overarching excitement for semi- or unstructured, hands-on, student-directed, learning permeates all of them. In this dissertation I will primarily use the term “making,” largely due to its prevalence in the media. However, my use of “making” should also be understood as making reference to popular terms such as tinkering, hands-on learning and learning-by-making. Furthermore, all of these terms should be viewed as being encompassed within the larger boundaries of constructionism. Having discussed some of the central tenets of constructionism in the Introduction, I begin this chapter by orienting the reader to several definitions of “making” and “makers.” I then describe the current state of “making” in education, by focusing on key literature from the Maker Movement. Within this discussion of contemporary “making,” I raise a number of concerns and challenges that need to be addressed in order to bridge “making” with the educational outcomes that are currently being touted by the Maker Movement.

What is Making? What are Makers?

Sylvia Libow Martinez and Gary Stager, prominent constructionists and authors of *Invent to Learn*, state that

> [M]aking is about the active role construction plays in learning. The maker has a product in mind when working with tools and materials…. Making is about

[5] Tinkering is defined as a mindset associated with playful problem solving, and is the approach typically associated with “making.”

> the act of creation with new and familiar materials … [it] is something powerful, a personal expression of intellect…. Making lets you take control of your life, be more active, and be responsible for your own learning. (Martinez & Stager, 2013, pp. 32-33)

Dale Dougherty, founder and CEO of Maker Media describes a "maker" as

> [s]omeone who is a builder, a creator, a producer, a developer, someone who has an active sense of taking an idea and developing it into something that's real and tangible and can be shared with other people. It could be as simple as a sketch, but it could be as fully realized as a fully manufactured object …. Making is a source of innovation. (Dougherty et al., 2013; McLaren, 2012)

President Obama, in his proclamation of the National Day of Making, refers to "makers" as including "students learning STEM skills to entrepreneurs launching new businesses to innovators powering the renaissance in American manufacturing" and "making" as the "discovery, experimentation, and innovation that has been the hallmark not only of human progress, but also of our Nation's progress (Obama, 2014)." These three quotations are among the most prevalent views of "making" (see Appendix C for other quotations about "making"). Across all three, the idea of "making" takes on a variety of roles. For Martinez and Stager, "making" is about bringing one's idea to fruition, and about personal empowerment. This idea is very much in line with Papert's constructionism. For Dougherty, "making" means anything from producing a "simple sketch" to a "fully manufactured object" with the only requirement being that the artifact be the result of "taking an idea and developing it into something that's real." Finally, Obama describes a broadened form of "making" that includes "students learning STEM" and the practices of "innovators powering the renaissance of American manufacturing." While each definition focuses on different aspects of "making," the three quotations have a clear similarity in articulating that the creation of tangible shareable objects has the potential to foster dramatic changes. These changes are expected to usher in a revolution at both the individual and societal levels. For the individual, "making" affords a new sense of agency and empowerment. At the societal level, there is the expectation that as individuals realize greater agency, they will collectively improve, while also developing the future innovations that will fuel social progress.

The Maker Movement

Despite the major contributions made by several important education researchers, the current excitement for "making" predominantly centers on Maker Media[6], MAKE magazine and the Maker Ed Initiative – which advertises the vision of turning every child into a "maker." The movement has overlooked much of the groundwork laid by the various education visionaries referred to in the Introduction. In the same way that the current movement has lost its connection to the true pioneers of "making," the movement, as a whole, has lost its connection to education research. While there are a number of things that the Maker Movement does well, an analysis of its practices and recommendations suggest that it discourages abstract thinking and straddles between outright dismissal of assessment and using assessments that are both

[6] Maker Media is the parent organization responsible for MAKE magazine and Maker Faire,

narrow and shallow in their scope. First, on the topic of abstract thinking, Dale Dougherty wrote:

> 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. (Dougherty, 2013, p. 9; Dougherty et al., 2013, p. 3)

This quotation is taken from an article entitled “The Maker Mindset” and also appears in the “Makerspace Playbook,” the primary resource distributed to individuals interested in creating Makerspaces. Without question, Dougherty’s concern about high stakes testing is justified, and frequently discussed within the education research community (e.g. Collins & Halverson, 2009; Honey & Kanter, 2012; Piaget, 1973). However, the outcry against abstract thinking seems to run contrary to the idea that students are experiencing meaningful learning in the context of “making.” Furthermore, it misunderstands Papert on several accounts. First, the implication that students should just play and have fun, misrepresents Papert’s goal of “hard fun.”

> My whole career in education has been devoted to finding kinds of work that will harness the passion of the learner to the hard work needed to master difficult material and acquire habits of self-discipline. But it is not easy to find the right language to explain how I think I am different from the ‘touchy feely ... make it fun make it easy’ approaches to education. (Papert, 2002, p. 1)

Papert is not opposed to students developing the skills of abstract thinking associated with mastering “difficult material.” Instead the tools that he developed were designed to bridge the physical and the cognitive, with the hope that students would develop skills that would allow them to apply robust reasoning strategies across content areas and contexts (e.g. see “Gears of My Childhood” in Papert, 1980).

Second, Dougherty also laments formal education as being too “serious.” This view that constructionism is not “serious” is at odds with the following quotation:

> My little play on the words construct and constructionism already hints at two of these multiple facets--one seemingly “serious” and one seemingly “playful.” (Papert & Harel, 1991, p. 1)

Constructionism is not interested in pitting serious against playful, but instead finds ways to live at the intersection of these two ideas. In fact, Piaget very openly writes about the laments of strictly fostering a playful environment,

> A few years ago the main trend, especially owing to the widespread influence of psychoanalysis, was carefully to avoid frustrating the developing child in any way. This led to an excess of unsupervised liberty which ended in generalized play without much educational benefit. (Piaget, 1973, pp. 6-7)

Third, Dougherty, like many others, appears to be following a “standard view” of the concrete versus abstract comparison. Taking the traditional layman view of concrete, as necessarily representing something that is physically tangible, does not truly accentuate the goals of constructionism or what it means to make something concrete. Specifically, Wilensky writes:

> The pivotal point on which the determination of concreteness turns is not some intensive examination of the object, but rather an examination

> of the modes of interaction and the models which the person uses to understand the object. This view will lead us to allow objects not mediated by the senses, objects which are usually considered abstract - such as mathematical objects - to be concrete; provided that we have multiple modes of engagement with them and a sufficiently rich collection of models to represent them.
>
> Keeping in mind that the adjective concrete applies not to things, not to concepts or ideas or physical objects, but rather to relationships between people and these things, it follows that what we would like to achieve in the schools by revaluing the concrete is not a restriction of children's knowledge to a smaller but more “concrete” domain, but rather an enrichment of the child's relationship to the whole panorama of human intellectual endeavor. The lesson we take from Piaget is not that the child develops by leaving behind the primitive world of concrete operations and leaping into the enlightened world of adult formal operations. Rather what we desire is that the child concretize his or her world by engaging in multiple and complex relationships with it. (Wilensky, 1991)

From Wilensky’s writing, it becomes clear that the objective of Turkle and Papert’s (1992) “revaluing the concrete” is to help individuals develop a deeper and more widespread awareness and connection with objects, ideas and concepts. This is in contrast with the abstract, only in so much as the abstract has little to no representational meaning beyond the specific rules and formula included within its definition or description. By juxtaposing play, which most “makers” would argue is an important learning experience, and abstract thinking, Dougherty, and the authors of the “Makerspace Playbook” are suggesting that abstract thinking is not a part of play. In line with this view then, many Makerspaces take on a general ethos of more “doing,” and less “thinking.”

In addition to the erroneous dichotomizing abstract thinking and play, the Maker Movement tends to take a dismissive or shallow stance towards the analysis of student learning. One example of how assessment is being dismissed can be found in the following excerpt from *Tinkering*, a book published by Maker Media, Inc.:

> If you question how I know this learning took place in the course of that tinkering, I’ll have to confide that I have no proof beyond the following: most kids have learned oodles and oodles of stuff, including talking and walking, texting, and skateboarding, with just this hit-and-miss, trial-and-success, seat-of-the-pants approach. I believe this is called “proof by inspection.” (Gabrielson, 2013, Preface)

Gabrielson is not concerned about demonstrating that students are learning while tinkering. The evidence of knowledge construction is embedded within the actions that students complete. From a theoretical perspective I can see some validity to this idea, as Toulmin (1999) similarly argues for the idea of “knowledge as shared procedures.” However Gabrielson is not truly employing the idea of knowledge being expressed through practice because Toulmin would still be interested in first identifying the common procedures for a given discipline, and then examining the extent to which a student is engaging in those repertoires. Gabrielson is unconcerned with studying student learning and fails to recognize the merits of its analysis, as the following quotation makes evident.

> 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. That's far and away more important than developing the next generation of fill-in-the-bubble exams. But I'm not so interested in that. I'm comfortable with my gut instinct, and I'm enormously interested in and committed to trying to get more kids tinkering. (Gabrielson, 2013, Preface)

This book is largely discrediting research that studies learning in Makerspaces, and, in so doing, implicitly encouraging others to do the same. As such, the text suggests that there is no need to spend time or energy convincing constituents outside of the "maker" community of the relevance of "making" for learning. While this may seem trivial, it is quite difficult to promote effective change within the "maker" community without first having an interest, and some techniques, for determining the usefulness of any proposed modification. Furthermore, given the current importance of high stakes testing, one can scarcely afford to completely overlook learning, and simply trust "gut instinct." Notwithstanding, similar arguments that doing is learning were made for project-based learning. Several studies showed that students engaged in completing their projects, but still weren't developing a conceptual understanding of what they were doing (e.g. Barron et al., 1998). To be fair, though, it seems like Gabrielson is aware of the varied impact that tinkering has on students. In the section preceding the previous quotation, he writes:

> When older students are exploring and tinkering in just the same manner, especially if it happens to be in an institution mandated to carry out education, one can hardly describe the scene without a chorus of glowering skeptics chiming in, as if on cue: "Well, yes, they're having fun, but are they learning anything?" Here's my answer, the answer of this book: ***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, not to mention good fun on their own time, *and* it may put them in the driver's seat of their own education by realizing their competence and ability to learn through tinkering, *and* they may begin to demand more of just this sort of learning opportunity. (Gabrielson, 2013, Preface)

The above quotation is repeatedly sprinkled with the word "may." All too often this has been the way that outsiders, and insiders, of the Maker Movement have viewed the merits of the environment. Most people will not question the fact that nearly all experiences *may* result in fostering a tremendous change within a given student. The real challenge, though, has been to know what those changes are, what fosters those changes, and when those changes are likely to take place.

Proponents of the Maker Movement have started to take some steps to answer these questions. For example, *Tinkering* and the "Makerspace Playbook" both provide suggestions on post-activity prompts that can be used to characterize how students have changed as a result of their experiences. Gabrielson's oversimplified recommendations include having students: (1) draw their projects; (2) describe the steps used to construct their project; (3) explain what they learned; and (4) answer some content questions. Students are instructed to select the questions that interest them, and are given time to jot down responses. Empirically, though, students tend to be very disinterested in answering these questions. Accordingly, students' responses are terse and often times fail to lend themselves to meaningful analysis (Gabrielson, 2013).

The suggestions from the "Makerspace Playbook" provide a richer set of questions through which to solicit feedback (Figure 1).

Questions we have asked in past surveys include...

- If a friend asked you to describe your Makerspace in 10 seconds or less, what would you say?
- What did you think of the project vision?
  ...the completed project?
  ...the experience exhibiting?
  ...meetings?
  ...workshops?
  ...plussing?
  ...shop facilities?
  ...overall: the whole program this year?
- For students: How much help did you get from your mentor(s)?
  - What part of your Makerspace was the most fun for you?
  - What was the least fun or most frustrating?
- For adult participants: How many projects did you help with?
  - Were any of the team members you helped your children?
  - How engaged were the project team members?
- If you could change one thing about the program, what would it be? This is the place to give more feedback that didn't fit any of the questions we've asked. Suggest changes would you like to see for next year, or ways to reduce any frustration you felt.
- Share your success stories! Tell us anything we might share when we try to get other kids and adults excited about the program. Students, you can tell us about things you learned or new skills you gained. You can even describe anything at Maker Faire that interested or inspired you.
- Do you think you'll use a Makerspace again in the future?
- Spreading the word: If you know someone who should hear about this program, please give us their email address(es) here.

In asking these questions and analyzing the results, your goal is two-fold: to continually improve the Makerspace, and also to gather great stories and data to help sustain the program.

Figure 1. "Makerspace Playbook" excerpt of recommended survey question

Interestingly, though, the suggested questions in the "Makerspace Playbook," appear to be much more self-serving and less focused on individual reflection than Gabrielson's. Specifically, the note at the end of Figure 1: "In asking these questions and analyzing the results, your goal is two-fold: to continually improve the Makerspace, and also to gather great stories and data to help sustain the program;"

does not convey a great concern for deeply understanding what students learned, or what facilitated that learning. Instead the phrasing and questions give the sense that documentation is more about fundraising and recruiting, and not about fostering student development. To be fair, the "Makerspace Playbook" does suggest asking about the amount of time spent getting help from mentors, but this seems quite shallow for trying to understand *how* the mentors helped each student. Furthermore, one would hope that a question like this could be answered by a mentor, or through the use of technology.

Particular opinions of the Maker Movement aside, the concerns about assessment and promoting deep conceptual understanding through "making" also arise from many other prominent resources on "making." For example, among the various case studies listed by the New York Hall of Sciences - one of the preeminent organizations that is advocating for "making" - only one case study provides any information about evidence of student learning. Specifically, the website states:

> Feedback from teachers, parents, special education assistants, and educators all indicate that children more thoroughly understood technical concepts after creating and building, as well as unbuilding and repairing. (Windsor, 2013)

This is yet another example of a shallow assessment that provides little in the way of truly identifying how "making" is supporting student learning, or even how these case studies enable others to measure the efficacy of their programs.

Similar instances of shallow assessments are littered throughout the literature on "making." Not surprisingly, then, trying to develop ways to improve the quality and culture of assessment within Makerspaces was one of the central suggestions of the Design-Make-Play workshop that I participated in at the New York Hall of Sciences in January 2012. Specifically, researchers raised the following concerns:

1. How can the field integrate design-make-play activities into formal learning environments?

2. Making is likely to remain an untapped learning tool because the current techniques used for assessment do not enable strong accountability measures.

3. Feature-rich assessment and documentation techniques are needed to expand the use of making in STEM education. (Honey & Kanter, 2012)

The three concerns raised are all tied to the ability to assess "making" which is likely to be a pre-requisite for expanding the use of "making" in schools. The current dismissal and lack of awareness around how to analyze constructionist learning is the context in which I write this dissertation.

From a more theoretical perspective, the history of educational technologies and education reform (Collins & Halverson, 2009; Tyack & Cuban, 1995) has repeatedly demonstrated that the implementation of "revolutionary technologies" follows a well-known cycle of hype and disappointment. During the hype stage, proponents promise benefits that are said to be self-evident and often engage in an essentialist discourse that equates research, skepticism, and temperance with conservatism. However, "revolutionaries" seldom have a good sense of the realities of the education system (Tyack & Cuban, 1995). As a proponent of constructionism who has spent several years designing and facilitating "maker" workshops, I have encountered many of the realities and challenges of effectively enacting constructionist learning. From these experiences, and my review of other literature from the "making" community, there are an abundance of opportunities to improve the quality of "making," but these changes require acknowledging those challenges and embarking on well-designed research. Thus, I see research as a tool for balanced optimization rather than conservatism, and I believe that this could help "making" overcome the hype cycle that has limited innovative educational initiatives from the past few decades.

Chapter 2. Tracking Student Strategies

One of the compelling and intriguing aspects of constructionist learning is students' ability to go from a problem or challenge to a refined final product. How do students navigate this process? Where do their ideas come from? How do they overcome failed designs? Prior work has provided little in the way of defining how these processes are enacted, or about the common strategies that students use for approaching open-ended[7], or loosely constrained, engineering tasks. Work by Kafai (1995) on student learning and development in the context of building computer games provides an early analysis of student strategies along two different dimensions. First, Kafai identifies the ways that students transition from using a game format, which bears resemblance to Mario Brothers or PacMan, to adopting a more narrative format, in which the student becomes increasingly concerned with developing character personalities and an innovative story to motivate continued player participation. Second, Kafai chronicles instances of: new ideas; surface changes, title changes and the consideration of changes. Studying the emergence and interaction among old ideas and new ideas provides a rich lens for studying student learning. In a line of research that is more closely tied to studies of engineering education practices, Apedoe & Schunn (2012) conducted a micro-level analysis by studying usage of common science strategies: vary one thing at a time and hold one thing at a time; in the context of an engineering design project. Using this lens yielded weak results in

[7] Here and in other portions of this dissertation, I refer to open-ended tasks, as tasks that have no clearly defined 'best' answer. Instead, there may be an infinite number of possible ways for completing the task.

identifying a correlation between success and strategy. Instead Apedoe and Schunn found that design-oriented strategies (hold particular thing(s) constant and adaptive growth) are far more predictive of student learning and student success, than the common science strategies. But these results still tell us very little about the overall approach that students used in terms of how they came up with their ideas.

Later work by Berland and collaborators (Berland, Martin, Benton, Petrick Smith, & Davis, 2013) examined patterns in computer programming behavior that provided justification for extending the "tinkering" and "planning" categories (Turkle & Papert, 1992) to consist of tinkering, exploring *and refining*. In this case, the authors dissected chunks of each student's programming process to identify commonalities within the beginning, middle and final thirds of a programming course.

Berland, Baker, & Blikstein (in press) reviews some additional instances of computational analyses of constructionist learning environments (e.g. Blikstein, 2011; Blikstein et al., in press; Worsley & Blikstein, 2013). Throughout these analyses, Berland et al., (in press) identified that there is significant potential for using machine learning to analyze "making," and answer novel questions about student learning strategies and behavioral patterns.

One objective of this chapter will be to describe common strategies from a much more overarching perspective. Identifying these strategies, as well as their theoretical and experimental significance will contribute to the broader question of how to assess learning in these complex environments.

Part 1: Common Strategies

From a practical perspective, describing student learning in constructionist learning environments like Makerspaces and Fablabs, can be quite difficult. With the breadth of activities that students complete one has to question what and how to evaluate their progress. For example, students are likely to gain proficiency in a wide variety of areas based on their project and the resources used. As such there appear to be a dearth of metrics and content areas that one would expect to be applicable to all students. Even constructs like computational thinking and design thinking, may not be fully realized for all participants simply because of the nature of their projects. Despite the significant variance in the types of students, projects and tools, one way to start examining student learning is by studying changes in how students describe the origins of their design ideas. Specifically at the conclusion, and perhaps at set intervals throughout the activity, I suggest that students be asked to briefly explain where their most recent ideas came from. Adopting such a practice could easily be added to the current surveys advocated by prominent "makers" and may be an opportunity to build models of how students are developing transferable skills.

Engaging students in post-activity reflections about the origins of their designs is precisely the content for this section of the dissertation. Specifically, I identify four strategies that students of differing experience levels utilize to generate ideas for completing an open-ended task.

Study 1 Design

Participants. Identification of the four strategies is based on a study of thirteen participants with different levels of prior experience. The most experienced students were enrolled in Engineering Ph.D. programs at the time of the experiment and had engaged in engineering practices for several years. The least experienced were 9th grade students who had limited prior knowledge or training in engineering. As additional context, the high school students included in this study came from three San Francisco Bay Area charter schools.

Task Description. All participants worked individually and were presented with the challenge of supporting a small mass (< 1 kg) as high off of a table as possible. To complete this task students were provided basic household materials: four drinking straws, five wooden Popsicle sticks, a roll of tape and a paper plate; and were given an unlimited amount of time to reach a final structure that they were satisfied with (Figure 2).

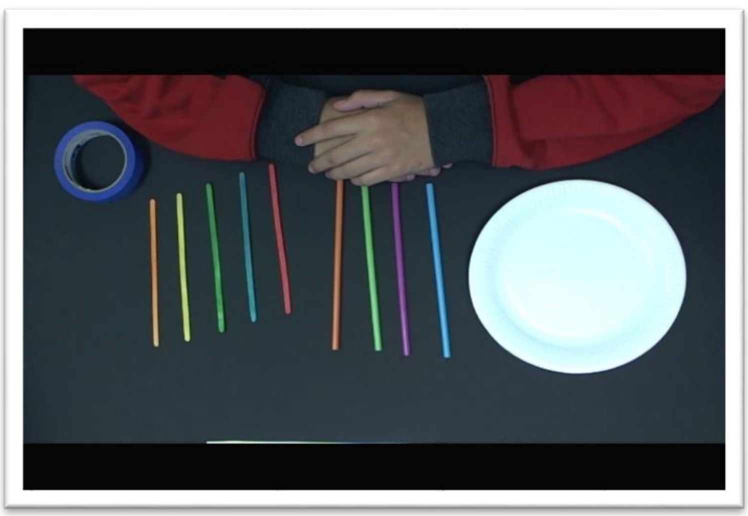


Figure 2. Materials provided for activity

At the conclusion of the activity students were asked about how they came up with the design of their structure. Responses to this question form the basis for identifying and describing the four common strategies referred to in this dissertation. Because my interest is about identifying the existence of these strategies I did not employ a detailed coding of each student's response. Instead, based on having conducted the interviews and watching the videos, I formulated the four strategies.

Strategies

1. *Unexplained spontaneous insight* – instances where the student is unable, or unwilling, to state where the idea came from.

2. *Materials-based reasoning* – instances where the student indicates that one or more properties of the materials provided the basis for their idea.

3. *Example-based reasoning* – instances where the student based their idea on a structure or item from a prior experience. A particular focus is on items from the student's home, community or school.

4. *Principle-based reasoning* – instances where the student based their design on principles or concepts from science or engineering.

In the sections to follow I provide detailed examples of each reasoning strategy based on the data from Study 1. I begin with principle-based reasoning, and then move to example-based reasoning, materials-based reasoning, and, finally, unexplained spontaneous insight.

Principle-based reasoning. Within this specific study, students whose strategies were classified as being principle-based, commonly used triangles and circles throughout their design. This was the case for the structure pictured in Figure 3 and Figure 4, which contain the underside of his structure.

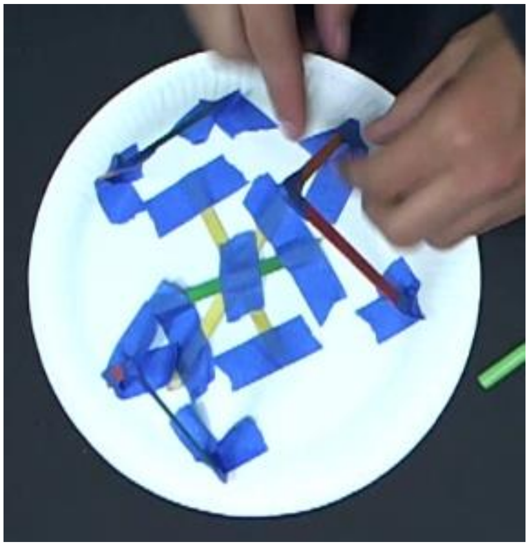


Figure 3. Sample principle-based design

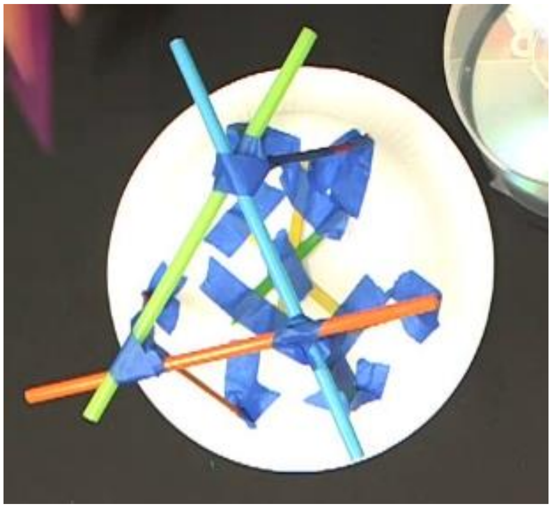


Figure 4. Sample principle-based design

In Figure 3 we see the early stages of the base. This base features two levels of triangles. The first is the shape of each leg. The second is the triangular base that the three legs define. Figure 4 makes the second level of triangles more explicit with the addition of three straws that form a triangle. When asked what inspired his design the student responded: “Well triangles are strong. And so, I decided to use as many triangles as I could.” Upon further probing about the importance of triangles the student offered the following explanation: “[i]t’s the most secure shape because, uhh, none of the angles can change once you have three sides in place. Whereas a lot of other shapes, they can tilt around and change.”

The student is very confident in his reasoning, and provides a succinct justification for his design. This was a 9th grader at a local high school who spent considerable time on engineering related projects outside of school. His structure was among the best of all of the participants in terms of stability.

However, principle-based reasoning does not always result in success. Figure 5 depicts an image of a structure that failed to meet the requirements. Namely, the weight fell off the structure. This structure was made by a Ph.D. student in Mechanical Engineering. When asked about what motivated the design, the student described the central role that triangles played in the structure.

> Interviewer: So what, what motivated your design?
>
> Student: Triangles…
>
> Interviewer: Triangles?
>
> Student: Uhh, the strongest structure is triangles. So I started with these triangles (pointing to the base of the structure) and figured out how to do more triangles… kind of like building a truss system.

We again see a principled approach to designing and building. She later reconfirms her principled orientation in describing why the structure failed and how she went about testing it. Specifically, she explains that the structure failed due to poorly constructed connections between the pieces of wood (Figure 6).

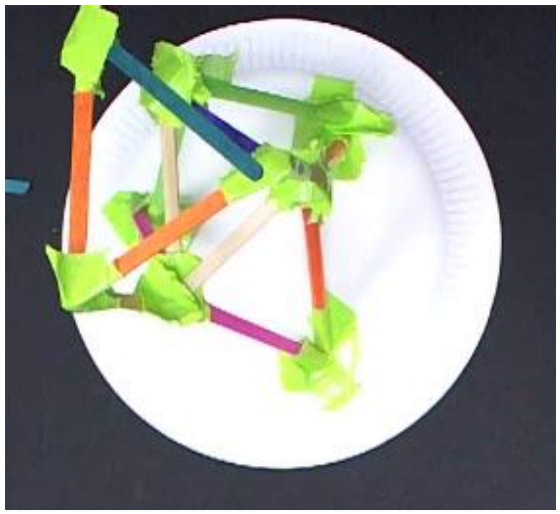


Figure 5. Unsuccessful principle-based design

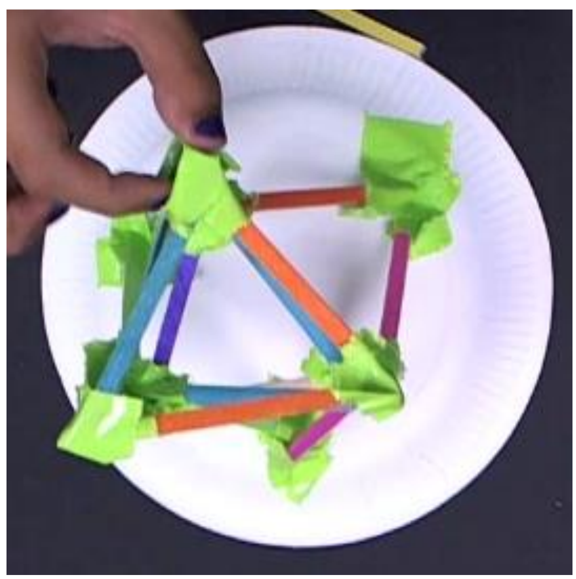


Figure 6. Student identifying weakness in their design

Furthermore, her approach for troubleshooting, which consisted of applying a uniform force to the top of the structure in order to identify weak points, also represents a focus on principles. From these excerpts we see that the student is using engineering principles at both the design and debugging stages. This depth of explanation is analogous to the constructs of mechanistic reasoning and causal reasoning (e.g. Lehrer & Schauble, 1998; Russ, Coffey, Hammer, & Hutchison, 2009).

Example-based reasoning. In example-based reasoning it is commonplace for students to create a structure that closely resembles a real-world object. More importantly, though, example-based designs are often modelled after a specific item that the student has frequently interacted with. This is certainly the case for the example that I present in the following dialogue with a 9th grade student.

> Interviewer: So how'd you come up with your design? What inspired it?
>
> Student: Uh…It's like the form of a chair. Plus. Yeah just like a chair… 'cause I have a chair like that… at home.

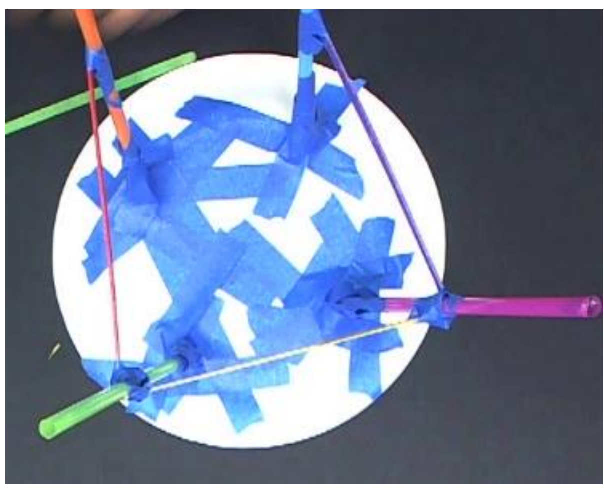


Figure 7. Successful example-based design

This student draws on similarities between a real-world object from his home and the engineering design task. He recognizes that a chair satisfies many of the same requirements as this particular task, and therefore elects to use a chair as the basis for his design (Figure 7). As the student continues to describe the motivation for the design, he indicates that he had briefly entertained another idea.

> Interviewer: Did you consider any other designs?
>
> Student: Nah. Oh. Oh yeah. And I'm like… Oh that's dumb.
>
> Interviewer: Why? What way was that?
>
> Student: That was um, just putting the straws straight up. And then putting the sticks inside. But like…let's use the sticks for something else.
>
> Interviewer: So what do you think, what would have happened if you had gone with the other design?
>
> Student: It would have fell, probably.

This other idea very closely resembled his current structure but did not have wooden sticks connecting the legs. When asked why he didn't pursue the other design he says that the other idea "was dumb." While he may have been hinting at principles in engineering design, he does not articulate this point. This line of reasoning is clearly distinct from the in-depth, principle-informed comments of the principle-based reasoning examples. However, the ability to transfer knowledge from a problem to a potential solution has advantages to being unable to articulate a response (unexplained spontaneous insight), or having ideas couched in the properties of the material (materials-based reasoning), which is described in subsequent sections.

Materials-based reasoning. Similarly, example-based reasoning, materials-based reasoning also provides a powerful tool for helping students start building their structures. However, instances of materials-based reasoning tended to occur alongside example-based reasoning. For example, when asked to describe the origins of his design one mechanical engineering graduate student remarked, “a table, I saw the plate and thought of making a table of some sort.” The second phrase, “I saw the plate and thought of…” captures the central idea of materials-based reasoning. The materials trigger the student to think in particular ways. This is in contrast to example-based reasoning because in example-based reasoning the tendency is to start by thinking of example structures that solve a similar problem as the one posed by the specific challenge. In materials-based reasoning the student *starts from the materials, instead of starting from the problem*. When comparing materials-based reasoning and example-based reasoning, the two may lead to the same overall design, but represent different initial motivations.

Materials-based reasoning is also distinct from principle-based reasoning. One example of this is the fact that principle based reasoning typically involves the student adapting or contorting the material to fit a principle. In materials-based reasoning, the student is trying to find the principle that “fits” the material, hence the directions of idea generation are opposite in tendencies.

To further ground the idea of materials-based reasoning, I present an exchange from a second 9th grade student.

> Interviewer: So, where’d your idea come from?
>
> Student: I saw how these [sticks] had the flat ends and started stabbing the thing to make it go in.

Here the student has taken an action based on an observed property of the material, namely the flat ends. He later goes on to describe that he put the sticks and straws together because the sticks fit inside the straws.

Student: But that didn't really like, stick in there, so I pulled the tape out and put the things in and wrapped it around it. And to make that taller, those (the sticks) fit inside these (the straws), so I just put it on top and made it.

The dialogue continues:

Interviewer: You mentioned earlier that the materials don't fit. What would have been better? What would have been ideal materials to build with?

Student: I mean that what I meant by that was like, I wanted something to have a fatter, what is it? Like a beginning, more like…

Interviewer: Tapered? Or fatter straws?

Student: Yeah so it would have been the, the beginning would have been more stable.

Interviewer: Oh at the bottom?

Student: Yeah. Cause my initial thing was to put the straws in like, a square and have these on top of it, like a flat like, boat type thing. Then have the tape and then that.

The student described the materials as not being a good "fit" for the activity. In this sense he was looking for the materials to dictate what he should do, as opposed to thinking about ways to use the provided materials to complete his idea. This approach encapsulates the materials-based reasoning strategy and highlights some of its limitations. Without appropriate cues from the materials this student probably would not have succeeded. And, in fact, this student significantly struggled until he found a clever way to use the roll of tape as a component in his final structure. Figure 8 shows the student's structure before this insight. The student eventually abandons that idea because it is unstable, and moves on to the design in Figure 9.

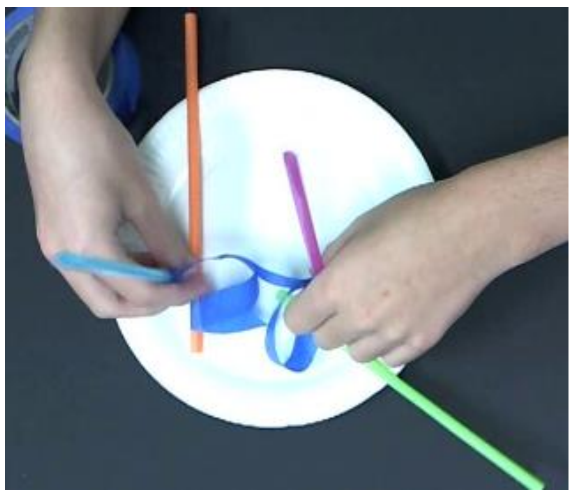


Figure 8. Student design before material-based insight

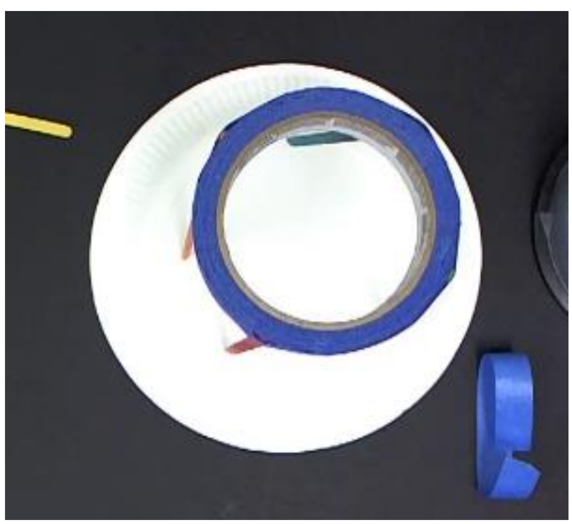


Figure 9. Student design after material-based insight

As an additional note, the above reference to a “boat type thing” and “make the bottom stable”, is another instance of how a student may invoke several of the reasoning strategies within the same explanation. Though I do not make an argument about the direction of causation, the student is describing a way to arrange the straws such that the structure will have a wider base, and likens this idea to a boat. Accordingly, it should be clear that the reasoning strategies do provide bridges between one another that can emerge over relatively short time-scales[8].

Unexplained Spontaneous Insight. Of all of the strategies, unexplained spontaneous insight is likely the easiest to recognize. Below are two dialogues with 9th graders that provide concrete instances of unexplained insights followed by the final structure for the student in the second dialogue (Figure 10).

Dialogue 1.

Interviewer: How’d you come up with that idea?

Student: I don’t know. I really don’t.

Interviewer: Have you seen anything that looks like that?

Student: No. It kind of looks like [turns it over]… I don’t know.

Dialogue 2.

Interviewer: So what inspired your design?

Student: I don’t know, I just started doing it.

[8] Following the presentation of the remaining reasoning strategies, I will discuss their non-mutual exclusivity in more detail.

Interviewer: But you had no initial thoughts?

Student: Not really.

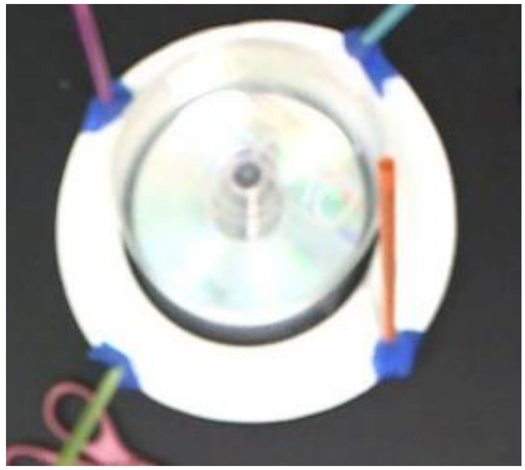


Figure 10. Unexplained spontaneous insight example

These two students are either unable or unwilling to articulate the origins of their ideas. While this may seem inconsequential, this behavior may be indicative of a lack of engagement, or a lack of a clear direction, and could easily create challenges for the student. For example, following an unknown approach can result in the students having considerable trouble in debugging or fixing their structures. Without a clear understanding of the intermediate impact that a given action will have, students might find it difficult to complete the task and/or effectively interpret the feedback that they get from testing their structure. This is in contrast to principle-based reasoning, which gives the student some foundational ground rules for analyzing their design; and example-based reasoning where students are at least working towards a template. Similarly, with materials-based reasoning, there is typically at least a portion of the design that the student can recognize as being "correct" in the sense that the material is operating in accordance with its analog. In the case of unexplained, spontaneous insight, this does not occur.

Relationship and Connections among Reasoning Strategies

As briefly alluded to in the description of material-based reasoning, the reasoning strategies need not be viewed as mutually exclusive from each other. In one of the materials-based reasoning examples, the plate caused the student to recognize an entire example structure that helped him complete the task. This provided a clear instance in which one reasoning strategy prompted the student to leverage a second strategy. Theoretically, one can imagine that any reasoning strategy can give raise to any other (Figure 11).

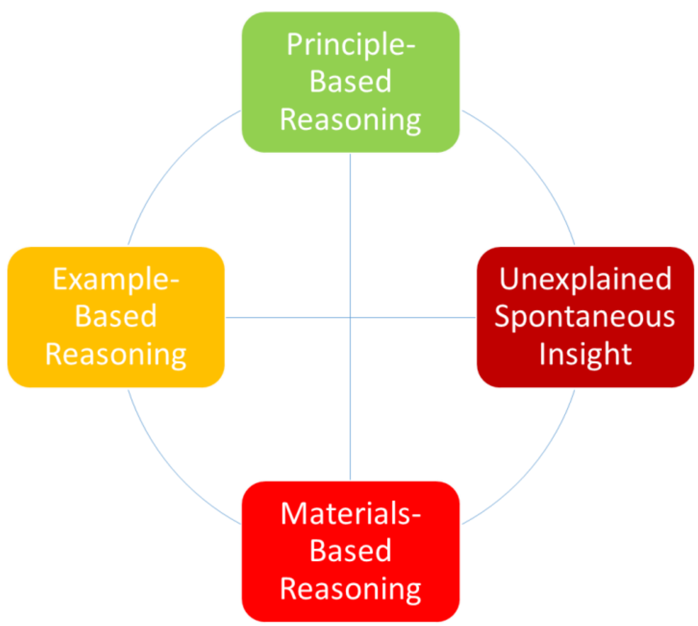


Figure 11. Theoretical Relationship among Reasoning Strategies

However, I hypothesize that this is not the case. Instead, I would expect that if we accept the proposed ordering of reasoning strategies: (1) unexplained spontaneous insight; (2) materials-based reasoning; (3) example-based reasoning; (4) principle-based reasoning; that individuals are most likely to move between consecutively numbered strategies (Figure 12).

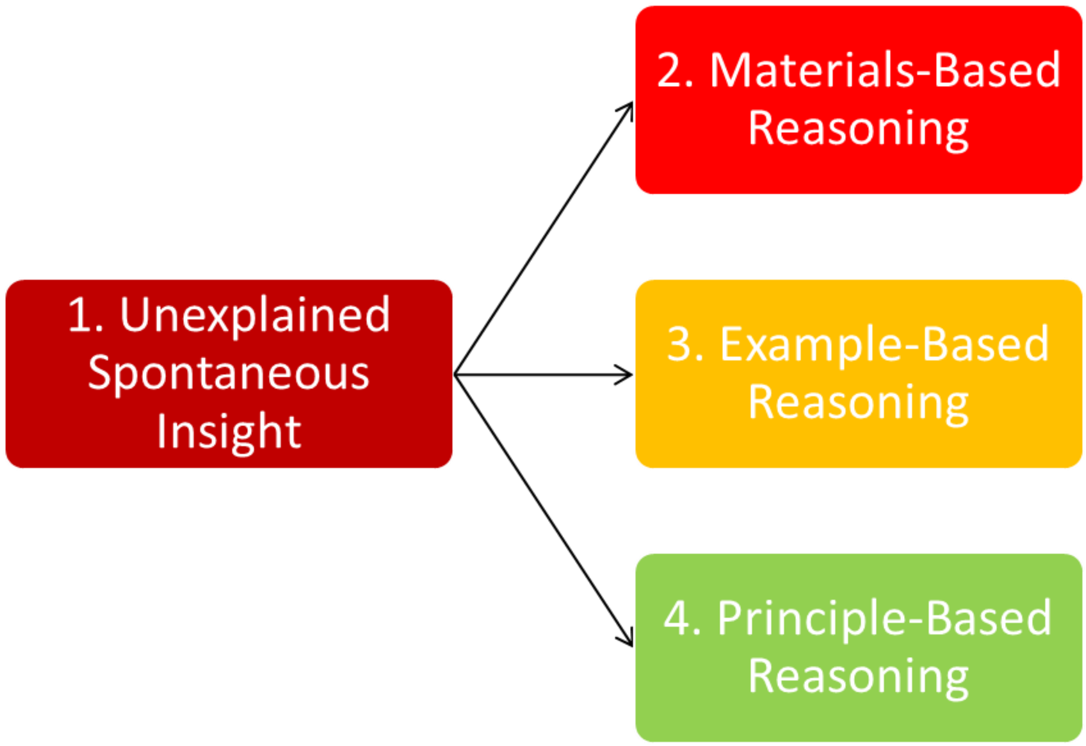


Figure 12. A representation of the relationships among the four common strategies.

The previously noted transition from materials-based reasoning to example-based reasoning would fit this requirement. Similarly, a student who is able to extract the principles from an example, would also adhere to this hypothesis. Figure 12 provides a representation of this hypothesized relationship among the reasoning strategies. In this figure, the arrows represent relationships that I theorize occur either spontaneously, or through coaching. From this perspective, one would expect to see instances that involve principle-based reasoning and example-based reasoning giving rise to one another, and instances where materials-based reasoning and example-based reasoning give rise to one another.

Having spent the past two sections describing the strategies and their hypothesized co-occurrence, I use the next section to outline two possible coding schemes for categorizing student respondents. After defining the coding schemes, I then move on to apply both coding schemes to a separate dataset.

Coding Schemes

Because I emphasize that practitioners and researchers study changes in reasoning strategies over time, I devote the current section to a pair of coding schemes that can be used for tracing student development.

Single Strategy Assignment. There are multiple approaches for categorizing student responses. In the first approach that I describe, a student response is categorized by a single strategy. Because the strategies are not mutually exclusive, this coding scheme describes a systematic way to check for each strategy. Specifically, a student's response is categorized based on the highest level of reasoning that it employs. [9] Accordingly, the coder can follow four quick questions:

1. Does the response mention engineering or science principles, even if they aren't described in complete or correct terms?

[9] Justification for suggesting that a given reasoning strategy is "higher" than others will be taken up in significant detail in Part 2 and Part 3 of Chapter 2. Specifically, Part 2 offers theoretical justification for this claim, while Part 3 provides empirical justification for this claim.

    a. If yes, label principle-based reasoning.
    b. If no, proceed to the question 2.

2. Does the response mention one or more example structures from a prior experience?
    a. If yes, label example-based reasoning.
    b. If no, proceed to question 3.

3. Does the response mention a comparison, or analogy, between one of the materials provided and another entity or property?
    a. If yes, label materials-based reasoning.
    b. If no, proceed to question 4.

4. Does the response indicate that the student is unaware of, or unwilling to state, the origin of their idea?
    a. If yes, label unexplained, spontaneous insight.
    b. If no, label as other.

Multiple Strategy Assignment. The first coding scheme above provides a simple way for classifying an entire response, without being concerned about interactions between reasoning strategies. The second coding scheme provides an additional level of complexity by identifying the presence or number of instances of each strategy type across one of more utterances. Thus, instead of assigning a single strategy to a student response, a response is described as a combination of different strategies. In this case the coder will label each individual phrase or sentence. For each phrase or sentence the coder answers the following questions:

1. Does the phrase mention engineering or science principles, even if they aren't described in complete or correct terms?
    a. If yes, then increase the count of principle-based statements by 1.
    b. If no, proceed to the question 2.
2. Does the phrase mention one or more example structures from the student's prior experiences?
    a. If yes, then increase the count of example-based statements by 1.
    b. If no, proceed to question 3.
3. Does the phrase mention a comparison or analogy between one of the materials provided and another entity?
    a. If yes, then increase the count of example-based statements by 1.
    b. If no, proceed to question 4.
4. Does the response indicate that the student is unaware of, or unwilling to state, the origin of their idea?
    a. If yes, increase the count of unexplained, spontaneous insight by 1.
5. Move on to the next phrase or sentence.

Through these coding schemes, researchers and practitioners should have a means for categorizing student responses at different levels of granularity and at different levels of complexity. To make this more explicit, in the next section I summarize aggregate results from applying the two coding schemes to a similar study (N=54) where students worked in dyads to complete a pair of engineering design challenges. As additional context, students were asked to identify principles or mechanisms in three example structures at three separate points during the study. Hence, these students have been primed to think about principles, and may not represent the expected frequency, of each reasoning strategy among a typical population of students. The statements codes were derived from student-provided hand-written responses to the following two questions:

1) How did you come up with your design?
2) Did something in particular motivate the design?

Results of the Coding Schemes

A single research assistant classified each participant's response using the two coding schemes mentions above. Using the simple coding scheme, in which a given response is characterized based on a single reasoning strategy, resulted in 63% considered as principle-based reasoning, 19% considered as example-based reasoning, and the remaining 18% equally split between materials-based reasoning and unexplained spontaneous insight (Table 1).

Table 1. Frequency of Strategy Usage using the Single Strategy Assignment Coding Scheme

Strategy	Percentage
Unknown	9%
Material	9%

Example	19%
Principle	63%

As a point of comparison, the second coding scheme provided a much more diverse picture of the student reasoning strategies. Principle-based reasoning was still the most frequently occurring (35%), but the combination of principle- and example-based ended up being the second most frequently occurring (18%). Hence the combination of strategies represents a large proportion of the overall set of strategies used. Table 2 contains a summary of the results from this coding scheme.

Table 2. Frequency of Strategy Usage using the Multiple Strategy Assignment Coding Scheme

Strategy	Percentage
Example-Unknown	2%
Principle-Material	2%
Principle-Unknown	2%
Example-Material	4%
Material-Unknown	4%
Material	6%
Principle-Example-Material	6%
Unknown	9%
Example	13%
Principle-Example	19%
Principle	35%

Of particular note is that only 6% of the responses involved reasoning strategies that I hypothesized to be uncommon. Namely, example-unknown, principle-materials and principle-unknown, are the three least frequently occurring strategies. Additionally, among the combinations of strategies, principle-example-material was the second most frequently occurring, which indicates that students may move between more than two reasoning strategies within the same explanation. These results provide initial confirmation of the hypothesized reasoning strategy relationships referred to above, where consecutively numbered strategies are more likely to co-occur than non-consecutively numbered reasoning strategies (Figure 12).

Summary

Despite the wealth of individuality that students bring to constructionist learning environments, there are commonalities in how they approach the design and implementation of a task. In this section I pinpointed four of the common strategies that students use, and provided examples of how these strategies differ from one another. As part of distinguishing the different strategies from one another, I presented a pair of coding schemes that may have utility in studying student development in constructionist learning environments and included results that are based on those coding schemes. The utility of the coding schemes, and tracking student strategies will become more salient across the next two sections, where I focus on the theoretical and experimental bases for suggesting that the reasoning strategies are hierarchically related. In referring to the strategies as representing a hierarchy, one aspect of hierarchy that I am highlighting is the increasing level of complexity and utility that one is able to draw upon when moving up from unexplained spontaneous insight to principle-based reasoning. Several prominent cognitive psychologists and education researchers have made a similar argument insomuch as moving towards principle-based reasoning is associated with the development of expertise, scientific reasoning, and strategies that have broad applicability (Atman, Chimka, Bursic, & Nachtmann, 1999; Brennan & Resnick, 2012; Bruner, 1960; Chi, Glaser, & Rees, 1981; Lehrer &

Schauble, 1998; Moss, Kotovsky, & Cagan, 2006; Piaget, 1973). To make this connection more concrete, I use the next section to describe prior work from cognitive psychology that can be used to justify why principle-based reasoning is more efficacious, and of greater complexity, than the other three forms of reasoning. I then conclude the chapter with a section that presents experimental evidence to empirically confirm the increased efficacy of principle-based reasoning relative to example-based reasoning.

Part 2: Theoretical Justification

All four strategies mentioned in Part 1 of this chapter bear resemblance to prior literature. While seldom cast with the same terminology, or within the engineering design or "making" domain, several studies support the existence, and the relative hierarchy associated with the four strategies identified.

Unexplained-Spontaneous Insight

Prior research on creativity contains examples of a phenomena, much like unexplained spontaneous insight. A foundational part of this research is based on the work of Maier (1931), Epstein (1999), and Kohler (in Epstein, 1999). Maier (1931) completed early work on the emergence of human insight in complex, novel situations through a task dubbed the "rope experiment" or the "two string problem." In this task, students were placed in a room in which two ropes hung from the ceiling. In addition to the two ropes, the room also had other materials (e.g. poles, clamps, chair, and an extension cord). The students were instructed to find a way to tie the two ropes together, which could not easily be completed since the student could not hold one string and reach the second string. Maier was particularly interested in understanding how conscious students were of their reasoning process. Based on post-task interviews he found that solutions typically occurred spontaneously, and that students were seldom able to articulate or express consciousness of the steps that enabled them to solve the task. Kohler did similar work by examining the emergence of insight among chimpanzees as part of a genre of work that anthropomorphized various animal species. Kohler used a box-and-banana problem in which the subject must learn to properly position a box in order to reach a banana that is otherwise out of reach. In Kohler's work, nearly all of the chimpanzees resorted to the same approach of fruitlessly jumping up to grab the banana, without making use of the box. This activity continued for several minutes without much indication of learning. However, after some time one chimpanzee experienced the momentary insight to move the box underneath the banana. To Kohler's observations this was a moment of unexplained insight (Epstein, 1999). Epstein (1999) pushes this idea further by proposing that individuals can be trained to behave in predictable ways when placed in new situations. In his box-and-banana study, he conditioned pigeons through various combinations of training regimens: (1) climbing a box; (2) freely pushing a box around the room; (3) directionally pushing a box around a room; (4) learning to avoid flying to grab a banana. After the training process, the configuration of the room was altered. The pigeons that had been trained in climbing and directional box pushing succeeded, while those pigeons who had not learned those skills failed. Similarly, even pigeons that learned to push the box, but never learned to do it in a directional fashion, took significantly longer than the other pigeons. If we extrapolate this research to humans, Epstein's work suggests that the types of novelty and creativity that people display in new situations should not necessarily be interpreted as involving higher-level reasoning. Instead, the behaviors of the humans, chimpanzees and pigeons may merely be the result of prior conditioning and can be elicited through the recommendations outlined in Generativity Theory (Epstein, 1999). In this theory, Epstein suggests that creativity can be fostered by promoting: (1) capturing and documenting new ideas; (2) seeking challenges; (3) broadening skills and knowledge; (4) regularly changing one's surroundings. For this reason, unexplained spontaneous insight can be classified as the least advanced of the four common strategies that I've presented, as there is no clear indication that the student was using unique human reasoning capabilities in order to complete the task. However, this should not be taken to devalue or discredit the spontaneous unexplained ideas that students generate, or their depth of knowledge. While Bruner, Goodnow, & Austin (1956) also emphasize that the ways humans make insights and inferences are highly dependent on one's environment, Bruner (1960) makes it evident that these intuitions are still quite useful. Specifically, Bruner makes the case that while formal education tends to stress the importance of analytic thinking, intuitions, or, as termed in this thesis, unexplained spontaneous insights, are of paramount importance in everyday life and in mathematics and science. Furthermore, using intuitive thinking often occurs only because an individual has deep knowledge of a given domain, and need not employ step by step analytical reasoning to arrive at the solution to a given problem. From this perspective, unexplained spontaneous insight can represent a very high-level of experience. Because of this, unexplained spontaneous insight represents an unknown category that can mean different things in different contexts. However, in the context of this dissertation, unexplained spontaneous insight tended towards instances where the student was not operating from the perspective of an expert with deep domain specific intuitions.

Materials-Based Reasoning

Because Epstein's work is centrally concerned with idea generation and creativity, some of his research also has relevance to materials-based reasoning. Specifically, in a series of studies that involved a modification to Maier's rope problem, Epstein (1999) discusses the observation of highly predictable behavior based on the properties of the object provided to the participants. In Epstein's studies, instead of providing various props for the subjects to use, he provided a single object, and indicated to the participants that they could use the object if they pleased. In these studies Epstein found that student performance and behavior was largely dependent on the properties of the object that was provided. For example, Epstein compared performance and behavior when the additional item provided varied in length. When the additional item was relatively long he found that participants experienced great difficulty in completing the task. He attributes this to the subject being convinced that the length of the long object was the key to solving the challenge. However, when the additional object was very short in length, participants had a much easier time solving the challenge. Namely, they realized that the object would need to be attached to one rope, and then placed in motion in order for the student to complete the task. This example from Epstein's work appears to coincide with materials-based reasoning, in that student reasoning is largely dictated by the properties of the material. The difference in my work is that I am examining student awareness of how the materials cued their behavior.

Materials-based reasoning also has ties to analogical problem-solving (Carbonell, 1982; Gick & Holyoak, 1980; Polya, 1945). This is a theory that will have connections to materials-based reasoning, example-based reasoning and principle-based reasoning, though with significant variation across each category. Among these three reasoning strategies, students construct analogies at different levels of abstraction (Anderson & Greeno, 1981; Carbonell, 1982; Gick & Holyoak, 1980). Central to analogical problem solving is the ability to draw connections between a real-world analog and elements of the current problem. In the case of materials-based reasoning, students develop a component level analog between one of the materials and a part of a real-world structure. Constructing this analogy represents an important way to enable completion of a task. However, as shown in Epstein's work, following a materials-based reasoning approach has limits in its applicability and may indicate that the student is unable to develop a more generalizable solution. Hence, relative to the other strategies, it appears to be superior to unexplained spontaneous insight, but excludes the more widely applicable strategies of example-based reasoning and principle-based reasoning.

Example-Based Reasoning

As mentioned above, example-based reasoning can be viewed as falling within the category of analogical problem solving (Anderson & Greeno, 1981; Carbonell, 1982; Chi et al., 1981; Gentner & Holyoak, 1997; Gentner, Loewenstein, & Thompson, 2003; Gentner, 2004; Gick & Holyoak, 1980; Kolodner, 1992; Loewenstein, 2010; Polya, 1945; VanLehn, 1996). In this instance of analogical problem solving, a student's technique for solving a given problem is closely tied to that of an example problem that they have previously encountered. As one can imagine, a student's ability to successfully solve a new problem is closely related to the similarity between the reference problem and the new problem. Additionally, problems that maintain the same surface features, as opposed to having similar deep structural features are also easier for students to solve (e.g. Gick & Holyoak, 1980; Polya, 1945).

During the past decade, researchers have paid greater consideration to the cognitive factors that enable students to correctly use analogical problem solving. For example, researchers have compared the effects of students being exposed to multiple examples, multiple questions and the principles that underlie a given example (Gentner et al., 2003; Gentner, 2004; Loewenstein, 2010). Many of these studies were motivated by the desire to identify the importance of example-encoding, and problem-encoding[10]. More specifically, researchers wanted to explore the effectiveness of learning-by-analogy when students have not had a chance to strongly encode the example, and determine how this weak encoding can be overcome when applying the analog to a new task. Strength of encoding has particular relevance to the definition of example-based reasoning used in this paper. As described here, example-based reasoning involves real-world objects from the student's home, community or school. By focusing on objects with high familiarity, the assumption is that the analog will have been deeply encoded, and that this deeper encoding will aid in the retrieval process.

In the context of expertise development, analogical problem solving, in its

[10] Encoding can be seen as the depth of a student's ability to recall and reconstruct a given example. Hence a strongly-encoded object is easy to recall in detail, whereas a weakly encoded object is more difficult to recall in detail.

most basic form, is seen as having relevance across several levels of experience. However, as individuals develop greater expertise, they are expected to spend less time building from analogies, and spend more time building from principles, which they would have inferred from interacting with several previous analogies (Anderson & Greeno, 1981; Chi et al., 1981; VanLehn, 1996). For example, VanLehn (1996) describes a process of generalization where the student no longer needs to pay attention to the surface information. Anderson & Greeno (1981) describe a similar transition from basing solutions on analogous problems, to eventually being able to work directly from general methods, strategies and concepts.

Finally, example-based reasoning is typically viewed as the default approach that students use (Hartley, Wilke, Schramm, D'Avanzo, & Anderson, 2011; Reeves & Weisberg, 1994; Roll, 2009; VanLehn, 1996). As the default approach, it certainly has its affordances. For instance, example-based reasoning is superior to strategies devoid of analogical reasoning and may have greater generalizability than materials-based reasoning, since materials-based reasoning is couched in the materials. Additionally, the process of identifying an analog means that the student has begun to identify features from which to draw comparisons between the problem space and previous experiences, and this can be a useful approach for tackling new problems.

Principle-based reasoning

Principle-based reasoning also has ties to analogical problem solving, and is often times referred to as learning-by-analogy (e.g. Colhoun, Gentner, & Loewenstein, 2008; Gentner & Holyoak, 1997; Gentner et al., 2003; Gentner, 2004; Kolodner, 1992; Kurtz & Loewenstein, 2007; Moss et al., 2006). Whereas materials-based reasoning involves drawing component-level analogs, and example-based reasoning involves surface features or system analogs, principle-based reasoning involves identification of deep structural features. These deep structural features are typically used for describing mechanisms, principles or phenomena at a deeper level than surface features. As an example of this Moss, Kotovsky, & Cagan (2006) showed that more experienced engineering students had greater facility with identifying and effectively using knowledge of the relationship between structure and function, than did less experienced students. The participants in their studies included college freshmen and college seniors. Each student was shown component diagrams of three electromechanical devices for forty seconds at a time. After the forty second period, students had 3 minutes to reconstruct the component diagram that they had just observed. Between the freshmen and the seniors, there were no differences in the rate of diagram recall. However, Moss et al. did find significant differences in the way that students went about reconstructing the diagram. Specifically, the way that the students grouped components, and the error rate in connecting groups of components significantly differed between the freshmen and the seniors. This effect was largely attributed to the seniors' ability to recognize and utilize more principle-based reasoning strategies in organizing what they observed in each system. In specific terms, the seniors were more likely to base their representations and reconstructions on the flow of energy through the three different systems. The freshmen, on the other hand, appeared to use surface similarities between the different subcomponents in order to reconstruct the diagrams. A second study that required students to generate written descriptions of each device confirmed the central role that function played in the different strategies students used for reconstructing and representing the electromechanical devices. Thus, even though both populations of students were shown the same representations, the seniors were able to extract and interpret the diagrams differently because they were using a more principle-based reasoning approach. The principles that they used were likely the result of having been exposed to other electromechanical devices through their prior coursework. From these prior examples, students were able to construct a set of principles that they used and applied to the new task of reconstructing diagrams for three electromechanical devices. In other words, they drew analogies between the principles used in the previous devices and the devices pictured, hence, principle-based reasoning can be interpreted as a learning-by-analogy strategy.

As was seen in Moss, Kotovsky, & Cagan, (2006), principle-based reasoning is commonly associated with individuals of high expertise, and seldom associated with novices. However, it should be understood that experts do not only use principle-based reasoning. When faced with new problems, expert students tend to fluctuate between example-based reasoning and principle-based reasoning strategies (Ahmed & Wallace, 2003; Chi et al., 1981; Cross & Cross, 1998; Mosborg, Adams, & Kim, 2005).

Principle-based reasoning also bears resemblance to mechanistic or causal reasoning (Bolger, Kobiela, Weinberg, & Lehrer, 2012; Lehrer & Schauble, 1998; Russ et al., 2009). A central theme of this line of research is the importance of having students address questions of how and why different mechanical devices and machines work as they do. Through advocating for students to engage in creating in-depth explanations of science, as opposed to mere recognition of surface features or components, researchers have documented students' ability to gradually build general rules out of the observations and explanations that they form (e.g. Hammer, 2004).

Discussion

Based on the prior literature, there appears to be a clear hierarchy in the level of complexity of the four reasoning strategies. Principle-based reasoning involves identifying deep structural features, whereas example-based reasoning tends to be involve surface features. Similarly, example-based reasoning involves drawing analogies between an entire structure and a given task, while materials-based reasoning is more constrained, and normally involves component level analogs. This hierarchy also coincides with prior work on expertise, and provides the theoretical basis for studying student reasoning strategies in more depth. Accordingly, in Part 3 I test the hypothesis that priming students using principle-based reasoning is associated with a higher rate of success than example-based reasoning. Testing this hypothesis is important because prior literature suggests that example-based reasoning, in so much as it involves students focusing on surface features, is the default strategy that students use.

While this prior research did not explicitly discuss example-based reasoning as the default strategy in constructionist learning environments, a survey of various tools used in the "making" community confirm this. For example, one of the ways that students are encouraged to gain familiarity with "making" is by finding a project on makezine.com or instructables.com (Dougherty et al., 2013; Gabrielson, 2013). These sites provide step-by-step instructions on what a student needs to do in order to complete a given project. However, in examining the typical set of instructions, I found that the focus is on creating an artifact that mirrors the example images provided on the website. Furthermore, the project instructions seldom contain explanations around why the materials used were well suited for the task, or any description of challenges that the poster encountered when trying to make the project work. The emphasis is on having samples that future students and “makers” can follow and replicate.

Lego Mindstorms follows a similar approach in which students are directed how to build a given robot through step-by-step instructions that are complete with pictures. However, these instructions fail to include any justification or explanation around what the student is doing, or how a given piece functionally fits into a larger robotic device. Instead, the emphasis is on having students complete their design, and then having them make slight modifications to the example in order to enhance their robot.

The Scratch community works in much the same way, with the vast majority of users remixing other students’ content. Remixers are, quite literally, using examples as the basis for their projects. Given this apparent salience of example-based reasoning in constructionist learning environments, demonstrating that principle-based reasoning yields better results has the opportunity to further motivate tracking student reasoning strategies, and suggests that constructionist learning environments should find ways to authentically promote principle-based reasoning strategies.

However, before beginning the analysis, it is important to acknowledge that there is also prior research that would put to question the hypothesis that principle-based reasoning should produce better results than example-based reasoning. One such argument is rooted in the expertise literature which emphasizes that to be an expert involves having extensive domain knowledge and the ability to properly organize that knowledge (Nokes, Schunn, & Chi, 2010). If this is uniformly the case, one could assume that there will be no benefit to principles based reasoning, since all of the students in the study that I describe in Part 3 are non-experts. Since, the students lack both expert knowledge and expert organization of that knowledge, the students in the principle-based reasoning condition will not perform any better than the students in the example-based reasoning condition.

Additionally there is significant research that highlights the challenges of fostering principle-based reasoning. For example, Hammer (2004b) and Wilson et al. (2006) both find that students in K-16 contexts have a difficult time using causal, or principle-based, reasoning in the domains of biology and physics. Specifically, Hammer (2004b) shows many instances where students' causal reasoning quickly deteriorates, and where students fall into the trap of relying on formulas, instead of building on and considering the principles associated with those formulas. In similar fashion researchers have found that students have a propensity to fall back on surface features even after being instructed about principles, or mechanisms (Hammer & Russ, 2008; Roll, 2009). Accordingly, one can question how well students would be able to infer principles in an unguided fashion. Another common challenge within principle-based reasoning is students getting caught up with the wrong details for a given problem (Chue & Lee 2013). All of these prior finding put to question how well students will be able to effectively use principle-based reasoning strategies.

Beyond the considerations associated with the difficulty of effectively enacting principle-based reasoning, there remain a number of other considerations that I briefly mention to further justify the relevance of this work. First, asking students to generate principles may be beyond their current level of domain expertise. If this is the case, one can imagine that the students are unable to effectively use the activity to identify strategies for approaching the task, and potentially encounter two deleterious outcomes. The first is that students experience stereotype threat as they perceive of their own lack of knowledge and lack of preparation for completing the task. This will cause the students to not only perform worse than their peers, it will also cause them to perform worse than if they had not completed the intervention (Steele, 1997). Another possible outcome is that students will develop scientifically incorrect principles, and then use these inaccurate principles in the design of their structures. Put differently, students may identify principles that do not apply to the situation of interest (e.g. diSessa, 1993; Lehrer & Schauble, 1998). This, again, has the potential to cause the students to perform worse than their peers.

As an additional consideration, prior work on cognitive load suggests that having students participate in exploratory tasks, when they do not have sufficient prior knowledge, is a significant drain on cognitive load, and will ultimately detract from their ability to complete the task (Kirschner, Sweller, & Clark, 2010; Kirschner, 2002; Sweller, 1988). In the example-based condition there is less propensity for this since the intervention is one that presumably requires less prior knowledge.

Based on these various arguments one could reasonably argue that principle-based reasoning will not prove to be significantly beneficial relative to example-based reasoning. Accordingly, there are reasons to go beyond mere speculation from prior research and complete an empirical study that directly compares principle-based reasoning and example-based reasoning.

Part 3: Empirical Justification

While the previous section provided a theoretical justification for why principle-based reasoning has benefits not achieved through example-based reasoning, this section features empirical results to test this hypothesis. Specifically, I conducted two studies in which non-expert students completed an example-based reasoning, or principle-based reasoning intervention, before participating in a building activity. Both of these studies show that the principle-based reasoning condition outperforms the example-based condition in terms of the quality of the designs that students make.

Study 2 Participants

For Study 2 the population of students included forty high school students from around the United States who were participating in a summer program at Stanford University. This implementation was largely akin to that of a classroom in that the entire population of students worked on the task at the same time. Each student received a worksheet with instructions for their specific intervention. The worksheet that the students completed can be found in Appendix D.

Study 3 Participants

For Study 3, the population of students included local high school students and Stanford undergraduate students. The distribution of high school students and undergraduate students was the same across the two conditions. This study followed a semi-structured clinical interview protocol with each pair of students completing the activity at different times, and a research assistant closely watching the process.

Study Descriptions

Study 2 and Study 3 involved dyads of students working to complete an engineering design challenge. The challenge mirrored that of Part 1 in which students were asked to build a structure that could support a 0.5 lb. weight as high above a table as possible. However, unlike the task from Part 1, students were not given tape, and were not given unlimited time. The sequence of events (Figure 17) completed for the activity included:

1. Baseline Sensor Data Collection – before beginning the actual study, students complete some baseline electro-dermal activation activities to determine their baseline in both stressful and non-stressful situations (this step was only completed for the Study 3).
2. Pre-test (Figure 13) – students were asked to generate as many ways as possible to make an unstable structure more stable. The goal of the pre-test was primarily to account for any differences in prior experience, as well as serve as a reference point for assessing how each student's conceptual intuitions changed as a result of the experiment.

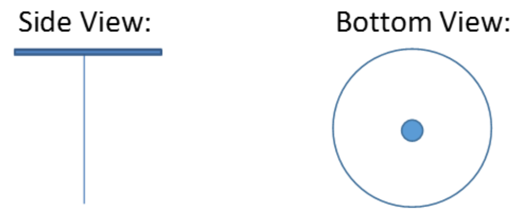


Figure 13. Diagram presented for pre- and post-test

3. Intervention – students participated in either an example-based reasoning intervention or a principle-based reasoning intervention. During both interventions students were first shown a picture of a ladder (Figure 14), a bridge (Figure 15) and an igloo (Figure 16). In the example-based condition students were asked to generate three ideas of relevant structures from their home, community or school that would be useful in thinking about completing the current task. In the principle-based condition students were asked to generate three mechanisms, or engineering principles, that cause one or more of the three items pictured (Figure 14, Figure 15, and Figure 16) to be structurally sound. The intervention task was three minutes in duration for both conditions.

Figure 14. Ladder picture

Figure 15. Bridge picture

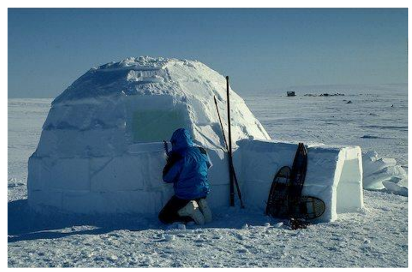


Figure 16. Igloo picture

4. Initial Design Drawing – students worked individually to create a quick sketch of what they thought their final structure would look like. This task was done as an intermediate step that would highlight if the intervention alone conferred noticeable advantages to one condition or the other.
5. Building Activity – students were given the materials and had fifteen minutes to complete their structure.
6. Post-Test – students repeated the pre-test task, and were given access to their pre-test data. The pre-test was made available to them in order to let them reflect on their prior designs (i.e. reuse them if they so pleased) and eliminate any concerns that some students may have forgotten their pre-test answers, while others memorized theirs.
7. Reflection – student were verbally asked the same post-interview questions as in Study 1, as well as some additional questions dependent on how their structure performed.

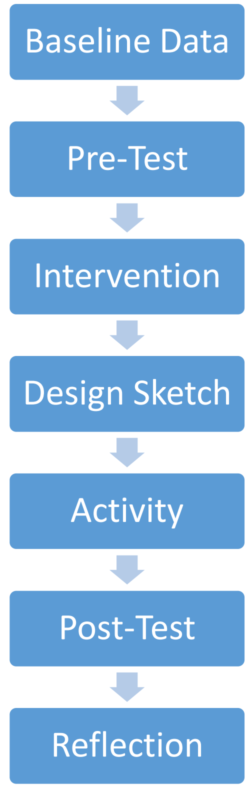


Figure 17. Overall design of the studies

Intervention Design

Before launching into a discussion of the analysis and the results, some attention should be paid to the quality and design of the interventions. Previous researchers have endeavored to develop taxonomies to describe different student-centered learning practices. In particular, two recent taxonomies provide important distinctions that are useful for categorizing active learning environments relative to passive learning environments (Menekse, Stump, Krause, & Chi, 2013) and for comparing different types of active learning environments (Alfieri, Brooks, Aldrich, & Tenenbaum, 2011). Menekse et al. (2011) describe applying the Differentiated Overt Learning Activities (DOLA) framework to students in an engineering education context. Specifically, they classify activities as being active, constructive and interactive. They demonstrate that when considering student learning, interactive is more effective than constructive, and constructive is more effective than active. Hence the three types exist in a hierarchy of efficacy.

Active learning is described as students participating in the learning experience through manipulation of materials or information. Prototypical examples of active learning include students having to copy information from the whiteboard, underline or highlight text, or play a video game that doesn't require making strategic decisions. Constructive learning activities challenge students to generate knowledge beyond that which was provided to them. In particular, the goal is that students combine information that they have recently been introduced to, with their prior knowledge in order to re-organize the mental models that they use to make sense of their surroundings and the course content. Examples of constructive activities include making concept maps, summarizing new information and engaging in self-explanations of subject matter specific content or phenomena. Finally, interactive learning is defined as the co-construction of knowledge within a group of one or more, other individuals. As students participate in knowledge co-construction they enter into a dialogue that challenges them to question their assumptions and both substantiate and refine their current understanding of the course content. Examples of interactive learning activities include reciprocal teaching, interactions that involve peers, experts and/or computer systems and structured argumentation (Menekse et al., 2013). Within Menekse et al's framework, both of the experimental conditions that I use represent interactive activities, the most effective of the active learning strategies.

However, when one considers Alfieri et al. (2011), the interventions that I designed reside among a class of strategies that are associated with poor learning gains, relative to direct instruction. Alfieri et al. (2011) conducted a meta-analysis of enhanced discovery learning experiences. Enhanced discovery learning involves the use of one of more activities that are designed to improve the quality of the discovery, or student-centered, learning environment. The specific types of enhancements that the authors identified include: generative tasks, elicited explanation; and guided discovery. Generative tasks are ones in which learners are required to generate information in response to experimenters' questions. Elicited explanations go beyond simple generation by challenging students to create explanations of the target task or the target materials. Finally, guided discovery includes providing students with direct instruction and/or feedback in order to help them proceed through each step of the learning activity. Within the meta-analysis, generative tasks were far less effective than elicited explanations and guided discovery. The two experimental conditions in these studies asked students to generate examples, or generate principles and/or mechanisms. This would put them both among the least effective of the enhancements strategies[11]. However, as I will show, despite being somewhat discredited as a positive form of enhancement, generative strategies can be quite effective. Furthermore, despite being aware of this research, my keeping the tasks at a relatively generative level seemed most appropriate for remaining within the current paradigm of "making."

[11] It is also possible that the principle-based reasoning condition is an instance of elicited explanations, in which case, Alfieri et al. (2011) would be providing additional justification for why the principle-based reasoning condition is more efficacious than the example-based reasoning condition.

In addition to the two taxonomies, there is an additional body of literature that proved to be essential in designing the two interventions. For example, in constructing the principle-based reasoning condition I carefully considered: 1) how to prompt students to generate principles; 2) the number of structures to show; and 3) the nature of the structures. Colhoun, Gentner, & Loewenstein (2008) demonstrated that extracting principles is significantly less likely to occur when the principle is not placed in the context of an example. Accordingly pictures were used to ground the discussion of principles. Three structures were included because several studies have demonstrated that students are much less likely to properly encode principles from a single example, than from multiple examples (Colhoun et al., 2008; Gentner et al., 2003; Kurtz & Loewenstein, 2007). Finally, the three sample structures were chosen to provide a variety of principles for students to identify. Specifically, students are likely to have very different levels of familiarity with each of the objects. I assume that many students will have interacted with and used a ladder; that all will have driven or walked over a bridge; and that none of them have extensively considered how to construct an igloo.

In order to keep things balanced, the same images were used for the example-based reasoning condition. Accordingly, one initial concern is that by using the same images, students in the example-based reasoning condition would start to look for principles. However various studies have shown that students have a tendency to focus on surface features even when presented with multiple items to compare (Hammer & Russ, 2008; Kurtz & Loewenstein, 2007; Roll, 2009). As such it seems unlikely that students in the example-based reasoning condition would make extensive use of principles. To make this more apparent to the reader I provide example artifacts from students in each condition and present their design drawings, the product of their intervention phase, and pictures of their structure.

Manipulation Check

As noted above, validating that the interventions were effective, will proceed by presenting artifacts from a pair from each condition. The example-based reasoning pair that I selected was chosen because these two students initially started by trying to identify principles, and had to be redirected back to generating examples. As such, if these students, who were initially bent on using principles still followed an example-based approach, this would suggest that the intervention likely took hold. The principle-based reasoning pair was selected because what the students ultimately describe in terms of principles is not extremely formal, or scientific, in terms of language or specificity. Nonetheless, these students were able to create a stable structure. In this respect principles need not be in the form of specific mathematical formulas in order to be effective. Instead having students identify principles and mechanisms in everyday language may be sufficient (Brown & Ryoo, 2008; Brown & Spang, 2008; Brown & Kloser, 2009).

Example-Based Reasoning. Figure 18 shows the example structures that the students came up with during the intervention phase. The students initially started by writing down "triangles" (crossed out). After reiterating that the objective was to think of three structures the students identified a swing, an easel and a dinner table. In interpreting the remainder of the artifacts and the designs that they created, it is important to understand that the students were specifically referring to a baby swing that has a “t” shaped cradle.

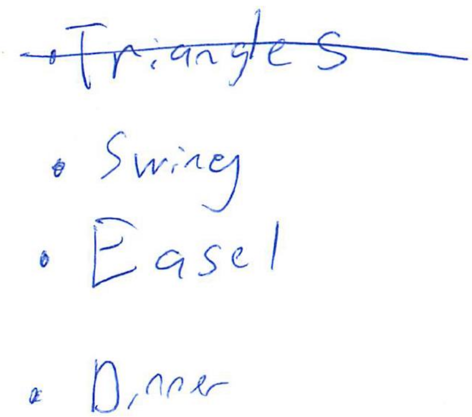


Figure 18. Extract of examples from pair of students in the example-based reasoning condition

Figure 19 contains one of the students’ initial design sketch (the design sketch was generated immediately after completing the intervention task). Of central importance to the image is the circled item at the bottom of their diagram. This student took the idea of a baby swing and tried to model their design after it. Specifically, the cross structure was intended to mirror the design of the baby cradle. This is explicitly noted by the text “create a cradle for the weight” (in the blue rectangle in Figure 19).

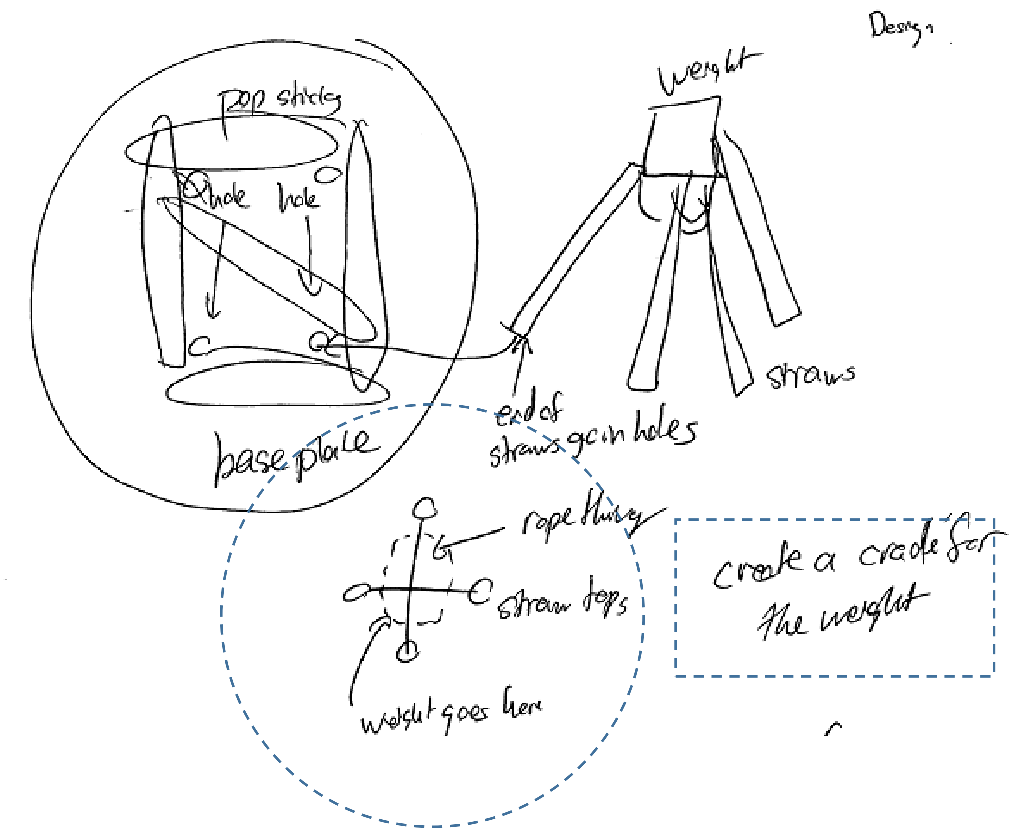


Figure 19. Design document from example-based reasoning student

When transitioning to the examination of the physical structure that the students created, the swing cradle remains the focal point of their design (Figure 20). Unfortunately, as they designed it, the cradle did not serve its purpose. In fact, had they abandoned the cradle, and simply placed the weight directly on top of the straws they would have succeeded. However, they were so centrally focused on using a baby swing as the model for their design, that they overlooked the actual utility of the cradle (Figure 20, Figure 21).

Figure 20. Side view of the final structure from an example-based reasoning pair

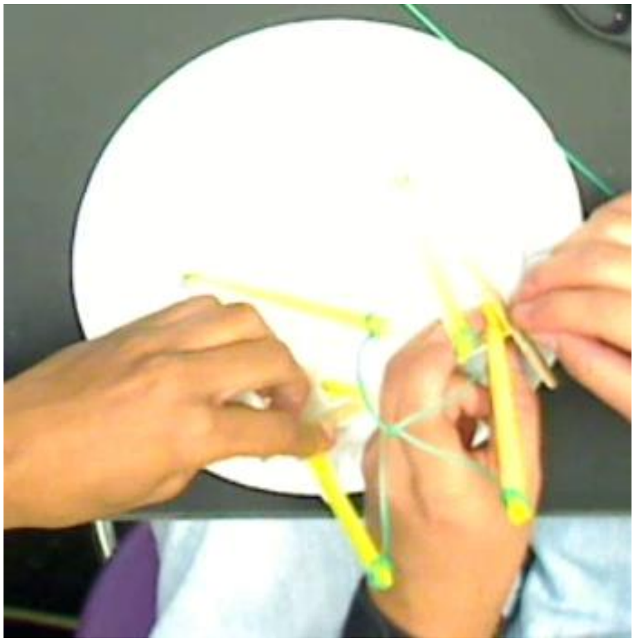


Figure 21. Top view of the final structure from an example-based reasoning pair

Nonetheless, this first pair provides a clear instance where the manipulation worked as expected. The pair of students identified an example object, from which they modeled their structure. Unfortunately for this pair, this design did not serve them well.

Principle-Based Reasoning Example. Figure 22 contains the principles identified by a pair in the principle-based reasoning condition. The items listed include: (1) balanced force holds structure together; (2) all force goes into ground; (3) no stress on joint.

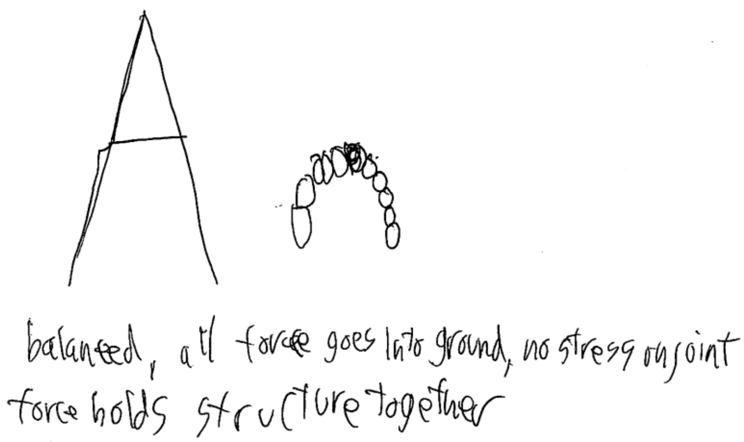


Figure 22. Mechanisms/principles identified by principle-based reasoning pair

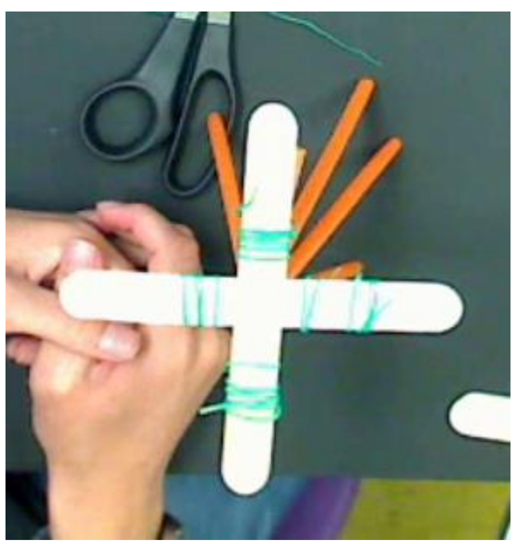


Figure 23. Top view of final structure for principle-based reasoning pair

An examination of their final structure in Figure 23, suggests that they attempted to apply these principles. Specifically they made a balanced and symmetric design. When asked about how they came up with their design, they said the following:

> Student 1: Because, it would balance more? I suppose I can try to go into detail about force, but I don't know much about that.
>
> Student 2: I mean. My guess is that... probably what we saw from that ladder picture, right?
>
> Student 1: Yeah. That's true.
>
> Student 2: So just the same concept as the ladder.

The students' description is not precise, and instead takes on the form of a general concept that they don't fully articulate. In their mechanisms from the intervention phase they identified the basic idea of balancing forces, but here admit that they don't know much about forces. In some ways these students have underestimated their own knowledge, since their intuitions about forces and balance appear to serve them well. They are able to attribute the idea as originating from the ladder, but what they extracted from the ladder was not a specific component, as was the case with the baby swing. Instead these students worked to make their structure balanced and symmetric. Thus the purpose and expectation of principle-based reasoning, is not that students will already have a well-defined understanding of engineering and science principles. Instead priming towards principle-based reasoning reasonably involves identifying principles that may be incomplete, inaccurate and imprecise. Consistent across comparisons between the example-based reasoning intervention and the principle-based reasoning intervention is that the latter examined the example structures with a different lens, namely one that was focused on deep features as opposed to surface features, and did so by creating explanations that were devoid of formulaic and of limited scientific precision.[12]

Results

The first point of analysis is the rate of success, in terms of each pair's ability to create a stable structure that could support the provided weight. In Study 2, four of the ten pairs in the principle-based condition were successful along this measure, while only one pair from the example-based condition was successful (Figure 24). I use a binomial test to investigate the hypothesis that the principle-based reasoning condition performed at a higher rate of success than expected. Specifically, if we take the probability of success on the task to be the rate observed among the example-based reasoning condition (0.10), then the probability of four of the ten pairs randomly being successful on the task is approximately 1%.[13] Hence, there is a clear indication that the principle-based reasoning condition performed better than the example-based reasoning condition.

[12] In part I have suggested that students in the principle-based reasoning condition used more principles in how they approached the task. Justifying this is taken up in more detail in Chapter 3.
[13] In this case the probability of success was used because there is no a priori value for the likelihood of student success for this task. Accordingly, the probability from the example-based reasoning condition was used as the baseline probability.

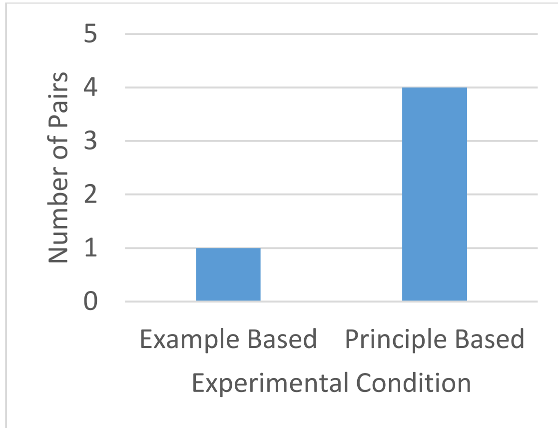


Figure 24. Number of pairs whose structures held up the weight by condition

The second measure that I examine is the height of the successful structures (Table 3). A one-tailed Wilcoxon Rank Sum test between the ranks yields a p-value of 0.013, suggesting that the rank of the principle-based reasoning participants is different from that of the example-based reasoning condition. (Note: this result also holds when including the remaining unsuccessful structures in the ranking.)

Table 3. Relative rank of structures which held up the weight by condition

Rank	Condition
1	Principle-Based
2	Principle-Based
3	Principle-Based
4	Example-Based
5	Principle-Based

From Study 3, I find analogous results. In this case, three of the five pairs in the principle-based reasoning condition were successful, whereas only one of the example-based reasoning pairs was successful (Figure 25). I again use a binomial test to compute the probability that these events happened randomly, given that the probability of success was 0.1 (based on the example-based condition in Study 2). Relative to Study 2, the effect in Study 3 is even more pronounced. The probability that the principle-based condition randomly resulted in three pairs being successful is less than 1%.

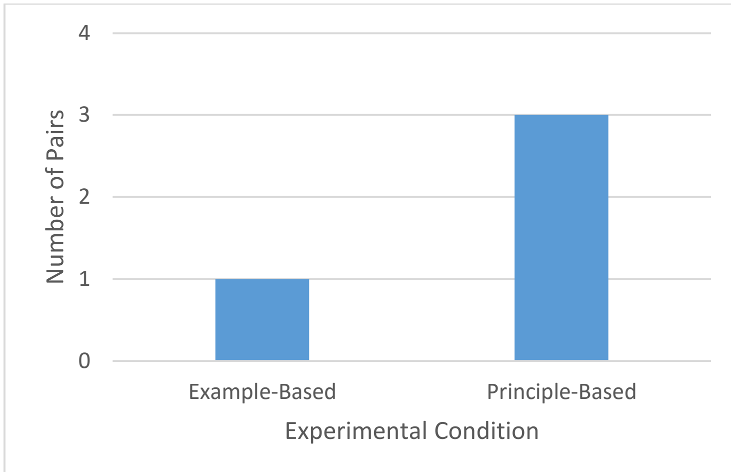


Figure 25. Number of pairs whose structures held up the weight by condition

Confirmation that principle-based reasoning is more effective was also seen in the Study 3 height ranks (Table 4). Specifically, using a Wilcoxon Rank Sum test indicates that the probability that the two conditions' ranks are equal is less 1%.

Table 4. Relative rank of structures which held up the weight by condition

Rank	Condition
1	Principle-Based
2	Principle-Based
3	Principle-Based
4	Example-Based

Discussion

The first two parts of this chapter provided initial indications that student reasoning strategies were an important area for analysis in constructionist learning environments. In this part I presented results from two independent studies that confirmed the hypothesis that strategy impacts design quality, and that principle-based reasoning is associated with better final structures than example-based reasoning. The benefit of principle-based reasoning was observed in both a classroom implementation

(N=40) and a controlled laboratory setting (N=20), with vastly different populations of students. Additionally, within both studies, the superior performance of the principle-based reasoning condition was observed across two different metrics for success, structural stability and height. Overall, this replicability suggests that the results are not random.

Finally, this analysis may have some theoretical implications in that I was able to effectively take a strategy, often times attributed as being a property of experts, and use it to empower the performance of non-experts. Despite these findings, questions still remain about why the principle-based reasoning group achieved better results than the example-based reasoning group, and whether or not the strategies impacted student learning. In Chapter 3, I address this question in great detail by examining the behaviors associated with the two reasoning strategies.

Summary

One of the primary factors motivating this chapter was the desire to chronicle how students go from a problem to a solution. Documenting the common strategies that students use provides a simple way for characterizing student learning in complex learning environments, and creating a unified trajectory for fostering student development. Accordingly, by cataloging four common strategies: principle-based reasoning, example-based reasoning, materials-based reasoning and unexplained spontaneous insight, I hope to have laid the foundation for assessing students based on a metric that is relevant for constructionist learning. Furthermore, the two coding schemes provided should serve as a manageable means for studying student strategies.

Part 2 contained theoretical evidence for (1) the existence of each reasoning strategy, and (2) the relative hierarchy among the strategies. Principle-based reasoning was likened to a more advanced reasoning strategy, based on its tendency to involve identification of deep, structural features. Example-based reasoning was described as a common strategy within both traditional and constructionist learning environments, and was presented as a way to help learners draw from their prior experiences. However, example-based reasoning often involves making analogies based on surface features alone. Materials-based reasoning involved an analogical problem solving strategy that was largely limited by the properties of the building material. Actions that require the individual to use common materials in uncommon ways prove to be quite challenging using materials-based reasoning. Finally, unexplained spontaneous insight was likened to the trained behavior of pigeons and chimpanzees who operate in a predictable manner based on their prior conditioning.

Having identified a set of strategies and a proposed theoretical hierarchy among these strategies, Part 3 included a pair of studies that compared principle-based reasoning and example-based reasoning. Across both studies, principle-based reasoning was associated with significantly higher rates of success and taller designs.

From these analyses comes the recommendation for Makerspaces and Fablabs to be more actively engaged in studying student reasoning patterns and developing minimally obtrusive pre-building activities that challenge students to draw upon their intuitions about pertinent domain specific and domain general knowledge.

This suggestion is particularly pertinent because, based on my observations, the Maker Movement follows two prevailing approaches as it relates to fostering ideation. The first approach is to place a high premium on doing, and couch ideation practices all together. This is understandable given that these environments often cater to students who have experienced repeated disappointment and ongoing challenges in traditional, teacher-centered learning spaces. Hence, any attempt to engage students in an activity that resembles school is undesirable. The other approach is to use “design thinking” and have students participate in brainstorming activities. Unfortunately, without proper direction, brainstorming bears significant similarity to the example-based reasoning experimental condition. If the goal is to advance the quality of the constructionist learning experience, both theory and practice suggest that the “maker” community use idea generation techniques that push students to more deeply consider principles, and move away from surface features. Failure to adopt strategies that improve the quality of these experiences, both in terms of cognitive development and in terms of product quality, may be a significant hindrance to “making” being viewed as a more legitimate educational endeavor. On the other hand, taking steps to document learning, through cataloguing changes in student reasoning strategies, and by fostering principle-based reasoning, in non-invasive ways, may provide a means to more closely bridge “making” and assessment.

Chapter 3. Multimodal Analysis of Making

The 21st century has seen an expansion in the set of tools available for assessing the quality of a given learning environment. A number of the traditional tools: test and quiz performance, speeches and essays; are modes of expression that have been around for centuries and remain the more privileged forms of assessment. For all of their shortcomings, these forms of assessment have the benefit of being widely accepted and easy to interpret. However, contemporary learning sciences research is increasingly concerned with additional constructs: motivation, engagement, collaboration, creativity, critical thinking, and problem solving, for example. These are constructs that tend to be much harder to quantify using traditional testing instruments and often necessitate adopting an alternative approach that more closely aligns with the design of said learning environment (Piaget, 1973; Schwartz, 1992). By virtue of the breadth of interactions students have with collaborators and various technological resources, traditional tools and metrics are probably not well suited for constructionist learning environments (Worsley and Blikstein, 2010, 2011). Instead, analyzing these environments likely requires the use of multimodal analysis.

Multimodal Analysis in Education

Multimodal analysis in education is not a new concept. On the contrary, multimodal analysis has been the primary means of analysis for decades of researchers trained in audio/video analysis, ethnography, etc. (for examples see Barron, Pea, & Engle, 2013). These researchers carefully analyze individual and group behaviors to interpret the inner workings of various learning environments. Perhaps the single individual text that best describes multimodal analysis in education is the Kress (2001). Kress examines multimodality among teachers and students in several science education classrooms. In each classroom audio/video capture, hand-written student-created artifacts, and field notes were used to study the intersection of the text modality with actions, facial expressions, diagrams and, guided noticing (Pea et al., 2004). Among the findings reported is the inability of text to accurately represent models of student learning in complex learning environments[14]. Specifically, Kress writes:

> “From our data we can demonstrate that attention to one mode alone fails to capture the meaning of a communicative event; not just that it fails to capture all the meaning, but that it fails to capture *the* meaning.” (Kress, 2001, p. 14)

Kress justifies this claim by presenting several analyses that are based on triangulating among speech, gestures and diagrams and show how a given utterance only has meaning in the context of the other events and actions that took place during that time. In leveraging Kress (2001) I would argue that if multimodal analysis is seen as a necessity for understanding student learning in a traditional science classroom, such a requirement becomes increasingly pronounced in a constructionist learning environment.

[14] Here complex means environments where students physically interact with other individuals and physical materials in the process of learning science

Accordingly, the current chapter builds upon some of Kress' findings, but also feature several important differences. At a basic level, whereas Kress used audio/video data and student artifacts, I have the advantage of having a wide set of sensory tools that can capture user behavior at high frequency and high resolution. Also, a portion of my analysis is similar to Kress' identification of characteristic multimodal actions among the population of students analyzed. His analysis identifies six different "conventionalized forms of action" which appear to have specific utility to the user. My multimodal analyses will also look for common multimodal behaviors among the population of students that participated in Study 3. However, because I have more detailed behavioral information about each participant, my analyses will span from being completely based on qualitative coding of human actions, to a largely automated analysis of speech, gesture and stress. Invoking analysis from behavior level data is another example of how my work deviates from that of Kress (2001), who is largely dismissive of behavior-based studies. Given the tools available at his time behavioral analysis may have been empirically fruitless and intractable. However, one of the things that I show in this chapter, and have shown in previous work (Worsley & Blikstein, 2013), is that multimodal behavioral analysis has merit for studying learning and performance. The ability to leverage purely behavioral data is, partially, a function of the computational tools and high resolution sensor data that allow researchers to construct user representations that are semi-semantic. A discussion of the techniques and sensors used will be briefly presented later in this chapter, and receive more detailed attention in Appendix A.

Multimodal analysis takes into considerations assessing student development as well studying the behaviors associated with different learning strategies. It is in the capacity of the second idea that I write this chapter. Namely, this chapter will motivate the use of multimodal analyses to delineate the behavioral differences between example-based reasoning and principle-based reasoning. Specifically I show three different types of multimodal analysis and demonstrate how capturing and analyzing data from non-traditional modalities has a potential to improve the field's understanding of constructionist learning, while also providing tools for more easily characterizing student development across different time scales.

Part 1: Disentangling Learning, Success and Process

Central to the current discussion of multimodal analysis is the recognition that human expression and cognition is indicated across a wide range of modalities. Along with this view is the recognition that beyond simply looking at student performance, in terms of success, there is a need to examine constructs of *learning* and *process*. Success and learning do not always coincide with one another. This is a primary criticism of the project-based learning movement. All too often the project-based learning activities privileged performance above learning (e.g. Barron et al., 1998; Bjork, 2013). In a similar vein, proponents of "making" seem to equivocate between: (1) the notion that learning simply happens by virtue of participating; and (2) the realization that the overall impact of "making" is limited. For example, Quinn & Bell (cited in Honey & Kanter, 2013) write, "learning is then instrumentally or incidentally accomplished in pursuit of the specific project goal and as participants work through the typical challenges and snags that arise in project work (p. 17)," but later go on to state that informal learning is primarily useful as an entry point to STEM. They also state that “students must engage in a ‘minds-on’ as well as a ‘hands-on’ process to achieve the conceptual growth and development that these standards will demand” (Honey & Kanter, 2013). Hence, there appears to be a combination of views concerning the efficacy of “making.” On the one hand, the act of “making” is seen as an expression and indication of learning. On the other hand, without having students mentally engaged, the desired types of learning do not occur. The latter concern is also raised by others from the constructionist community (e.g. Worsley and Blikstein, in preparation; Bennett & Monahan cited in Honey & Kanter, 2013).

Privileging performance, or outcomes, is a critique that can be made about my discussion of the different reasoning strategies from Chapter 2. In order to address this critique, and bring learning into the discussion, the next section defines learning in the context of the Study 3. As part of this discussion I provide confirmation that principle-based reasoning is more efficacious for promoting learning than example-based reasoning. The section on learning is followed by a definition of process, and a discussion of why process is important to this text. Once again, the analysis presented in terms of process will demonstrate that the principle-based reasoning condition was more effective than the example-based reasoning condition.

Learning

Depending on the target audience, learning can mean many different things. My view of learning is situated in the realm of conceptual intuitions (i.e. intuitive conceptions, naïve intuitions, etc.) (Bruner, 1960; diSessa, Gillespie, & Esterly, 2004;

diSessa, 1993; Hammer, Elby, Scherr, & Redish, 2005). These intuitions have been the primary focus of the conceptual change literature (see diSessa, Gillespie, & Esterly, 2004 for a discussion of conceptual change and intuitive conceptions) and are argued to have been brought into consideration as a result of the constructivist revolution. My objective, and expectation, in considering conceptual intuitions is not that students will necessarily leave the learning experiences with a formula or a set of engineering and science axioms. Instead, my goal is that participation helps students at a much more intuitive level (see Brown & Ryoo, 2008; Brown & Spang, 2008; Brown & Kloser, 2009 for research on students discussing science content without necessarily invoking scientific terminology). This does not exclude the possibility of developing more formal knowledge of a given domain, but this was not the intention. One consideration for remaining at the level of basic intuitions is that these intuitions are often times common across both experts and novices, and are the basis for many of "scientific" ways that human perceive different situations (Bruner, 1960; diSessa, 1993, 2008; Hammer et al., 2005; Hammer, 2004c).

In the context of this thesis, I will use "learning" to refer to a change in a student's recognition of important principles from engineering and/or science in the context of a generative task. Determination of a student's current knowledge state was based on their responses to pre- and post-tests. Specifically, Study 1 and Study 2 revealed two engineering design considerations that routinely impacted the success of students' structures. These two design considerations were slanted legs and connected legs (or cross bars).

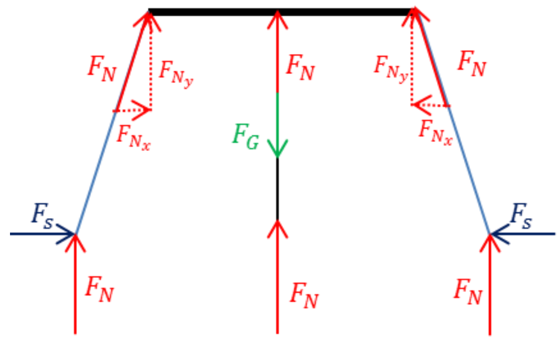


Figure 26. Free Body Diagram of Structure with Slanted Legs

Figure 26 contains a free-body diagram of a structure with slanted legs. While slanted legs are commonly used within a truss system, when using slanted legs without strong connections between the legs and the ground, or between the different legs, the structure is dependent on the force of friction between the legs and the ground to keep the structure from falling (F_s in Figure 26). Hence, the use of slanted legs was typically associated with poor performance.

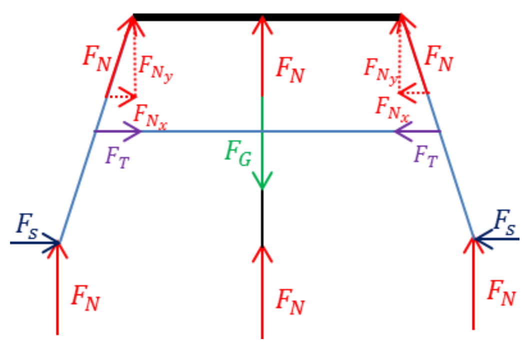


Figure 27. Free Body Diagram of Structure with Slanted Legs and Reinforcement

As noted above, one way to address the challenges of using slanted legs is to reinforce or connect the legs. By adding connections between the legs (F_T in Figure 27), the structure is no longer dependent on the frictional forces between the legs and the ground to keep the structure from falling. Accordingly, connecting the legs, in the presence or absence of slanted legs, confers stability.

Learning Coding Scheme. I developed a coding scheme based on the structural stability and instability conferred by connected legs and slanted legs, respectively. The coding scheme was used on each design from the pre-test, post-test and design sketches. In order for an item to be coded as having connected legs, there must be something that connects the legs at some place other than the top or bottom of the structure. This criterion differentiates between pairs that may have connected legs at the top or bottom of the structure by virtue of wanting to make a base, or out of necessity to make a connection between the legs and the top of the structure, and those that wished to reinforce their structure with cross-bars. Identifying the presence of slanted legs was based on the existence of a non-90-degree angle between the legs of the structure and the upper portion of the structure. When coding student pre- and post-tests, the presence of slanted legs resulted in a score of -1, and the presence of connected legs resulted in a score of +1. Accordingly, a given test or design could be coded as -1, 0 or 1 (Figure 28).

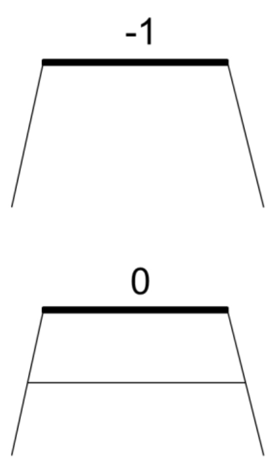


Figure 28. Example design scores for pre- and post-tests. From left to right, top to bottom, the figure contains: unconnected slanted legs (top left), unconnected parallel legs (top right), connected slanted legs (bottom left) and connected parallel legs (bottom right)

Taking the sum of the two values was adopted to stress the differences between those that recognize the challenges of relying on slanted legs, and those that capitalize on reinforcing the legs. As such, the coding scheme intentionally assigns students that use connected legs and slanted legs with the same score as students that used neither slanted legs nor connected legs. This scoring was done to account for uncertainty concerning the student's motivation for the inclusion of connected legs, or lack thereof. It may be that the student only included connected legs because they used slanted legs, but did not recognize that the two needed to be used in conjunction with another. Similarly, for students that used neither slanted legs nor connected legs, this may again be the result of their design not necessitating the connected legs.

Results[15]. Based on this coding scheme, three students received a score of -1, four received a score of +1 and the remaining students received a score of 0. Inter-rater reliability analysis yielded a Fleiss' Kappa of 0.76 based on coding from two research assistants. Based on the post-test scores, students in the principle-based condition (M = 0.4, SD = 0.48) were more likely ($t(18) = 3.46$, $p < 0.001$) to receive a higher score than their peers in the example-based reasoning condition (M =-0.4, SD = 0.48) (Figure 29).

[15] Because complete pre- and post-test data was only available for Study 3, no results are presented from Study 2.

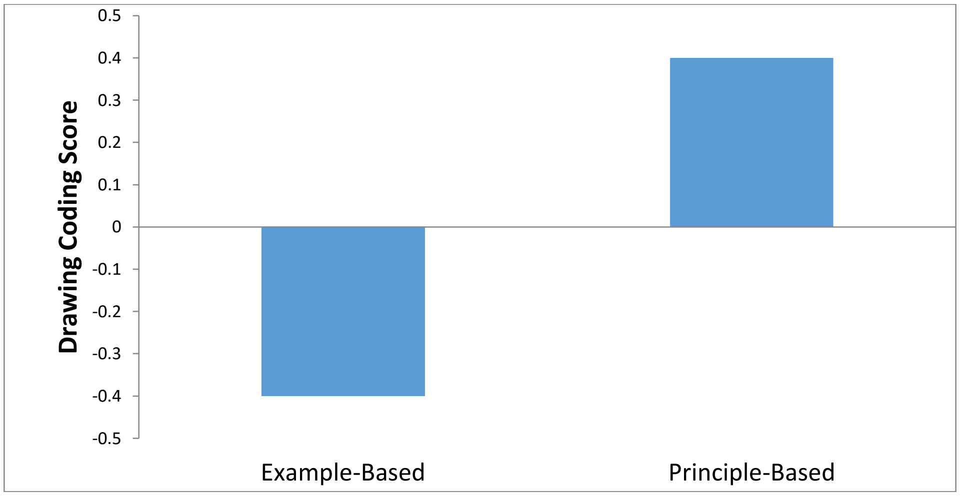


Figure 29. Average coded post-test score by condition

Substantively, this means that students in the principle-based reasoning condition were more likely to include connected legs in the absence of slanted legs, while students in the example-based reasoning condition were more likely to include slanted legs in the absence of connected legs. This is significant because there were no statistically significant differences between the two conditions immediately before the intervention, based on the pre-test ($t(18) = 1.23$, $p = 0.12$), or immediately after the intervention, based on the initial design sketch ($t(18) = 1.26$, $p = 0.11$). This result suggests that the observed difference on the post-test is not a function of the intervention alone, nor is it a function of prior knowledge, but does not appear to have relevance until *after* the building activity. As a result of doing the hands-on activity students in the example-based reasoning condition were more likely to propose a design solution that included slanted legs, without connecting the legs. To further explicate this difference I compared experimental conditions based on learning gains. A Mood Median Test, using a binomial distribution as opposed to a Chi-Square distribution indicates that students in the principle-based reasoning condition are more likely to experience an increase in score (p = 0.037) than students in the example-based reasoning condition.

These results may mean that students in the example-based reasoning condition were less likely to attribute structural failures to having slanted legs or having unconnected legs, whereas the principle-based reasoning group was more likely to make this association.

Figure 30 provides additional justification that students in the principle-based reasoning condition learned more. Namely, when comparing intra-condition scores, the example-based reasoning condition remained stagnant from pre-test, to design, to post-test, while the principle-based reasoning condition increased their score from pre-test to post-test ($t(9)=2.24$, $p = 0.026$). This, again, suggests that the principle-based reasoning condition students improved in their ability to appropriately apply engineering principles.

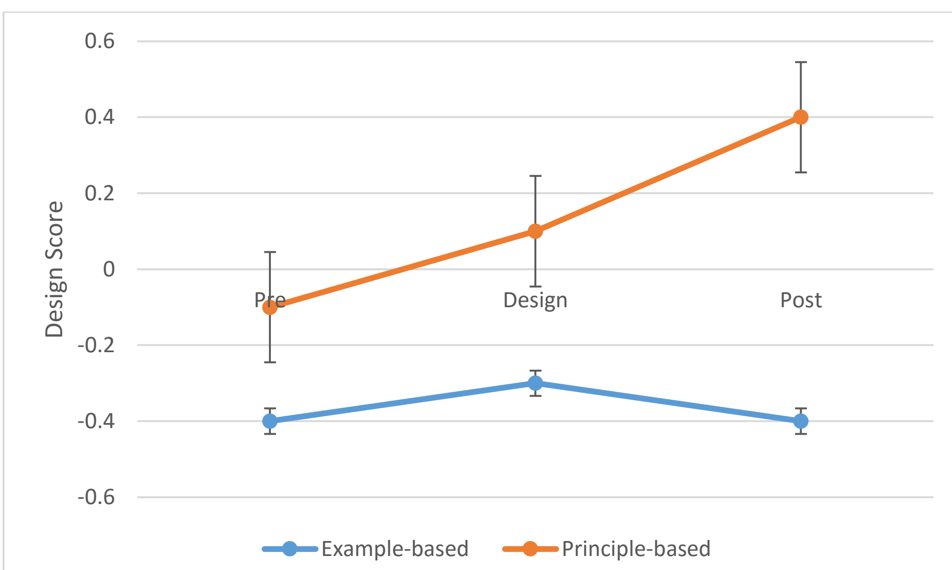


Figure 30. Pre-test, Design Sketch, and Post-test scores by condition

Having made an observation about the positive correlation between the principle-based reasoning condition and learning, I now turn to the question of whether learning and success correlate to one another. A student T-Test between student success (the structure standing with the mass on top) and student learning score shows no statistically significant result ($t(11) = 0.63$, $p = 0.27$) between the two variables. This result suggests that the metric of learning that I am using is not the same as success on the activity. Since this is the case, the forthcoming analyses look at success and learning as separate metrics.

The other perspective of learning that I consider is that while the students are participating in the engineering design task, they are engaging in a learning process, which was alluded to in the prior literature section, and will receive greater attention in the next section.

Process

The final concept that bears significance for the forthcoming analyses is the importance of process. Here process is defined as the behaviors, actions and interactions that occur while a student or group of students are completing a given task. Fundamentally, distinguishing between achievement and process has been an essential part of education research for nearly a century (Werner, 1937), and is the primary consideration of constructionist learning. For example, Turkle & Papert (1992) provides a prime instance where the researchers focus on studying learner processes, and not learner outcomes. Specifically, they write,

> Using clinical methods inspired by the Piagetian and psychoanalytic traditions, we built up case studies of children using computers in grade-school settings where they were encouraged to explore programming without preconceptions about the "right way" to go about it. We took 40 cases for which we had material both on individual personality and programming style. What we say in this chapter about gender, programming, and intellectual style is based on the analysis of these cases. But we believe that what is most important is not any statistical association between gender and programming styles, but what lies behind the styles and behind the resistance of our intellectual culture to recognize and facilitate them both. (Turkle & Papert, 1992)

The use of case studies centered on the belief that process was of primary import, and was more relevant than mere correlations between style and gender. Hence, my use of process is in line with the guiding principles and perspectives of constructionism (e.g. Harel & Papert, 1991; Kafai, 1995; Lawler & Yazdani, 1987).

Beyond the specifics of the constructionist movement, focusing on process has been espoused by several other researchers (Atman & Bursic, 1998; Atman, Cardella, Turns, & Adams, 2005; Bamberger & Schön, 1983; Lehrer & Schauble, 1998; Smith, diSessa, & Roschelle, 1994; Toulmin, 1999). Among these papers, authors are concerned with analyzing student learning with the understanding that as the student is participating in the study, they are engaging in a learning process. For example, Toulmin (1999) advocates for "knowledge as shared procedures." As such the analysis of student expertise should be situated in practices central to a domain, as opposed to solely being grounded in language or a final product. Similarly, Bamberger & Schön (1983) describe learning as a "reflective conversation with materials." The idea of a conversation encapsulates the ways that individuals interact with their surroundings, both human and non-human, to interpret and make sense of what they observe. Hence it is not enough to simply look at a structure to determine its stability. Instead the individual must engage the structure in something that is akin to a dialogue, applying stimulus to the object and getting feedback from the object. The forthcoming analysis of the inclusion of principle-based items in student structures will directly take this into consideration by stepping away from an analysis of student behavior, and instead examine how students' designs are changing over time.

Extracting process data. In Parts 2, 3 and 4, I present descriptions and results from three multimodal, process-based analyses. The purpose of these analyses is to identify the multimodal behaviors that conferred higher quality designs and improved learning in the principle-based reasoning condition, relative to example-based reasoning. All three analyses follow the same overall algorithm. This algorithm is designed to recognize process similarities between participants, and test the hypothesis that there are multimodal practices that distinguish principle-based reasoning from example-based reasoning. Within each analysis, the hypothesis is tested in two ways. First, students are clustered based on the similarity of their processes. This approach maintains the temporality of the student behaviors. I refer to this approach as computing pair-wise "process similarity metric." Second, I conduct a behavior frequency analysis which relaxes the temporal relationship. The purpose of this approach is to provide a simplified means for determining process-based differences. However, even in conducting this analysis, I look at the frequencies in aggregate, and also split each participant's process into three adjacent sections. Looking at behavior frequency at these set intervals moves closer to understanding how student processes differed without aggregating across the entire process.[16]

Figure 31 shows the general steps of the algorithm used for analysis. The algorithm builds on several previous studies (e.g. Berland et al., 2013; Blikstein et al., in press.; Piech, Sahami, Koller, Cooper, & Blikstein, 2012; Worsley & Blikstein, 2013) and makes every effort to maintain the context of each piece of data by taking temporality into consideration, whenever possible. The paragraphs to follow provide a summary of each step.

[16] With the various grain sizes of data utilized, one concern is dealing with multiple comparison bias. To address this, I used Benjamini-Hochberg posthoc analysis with an initial alpha of 0.05.

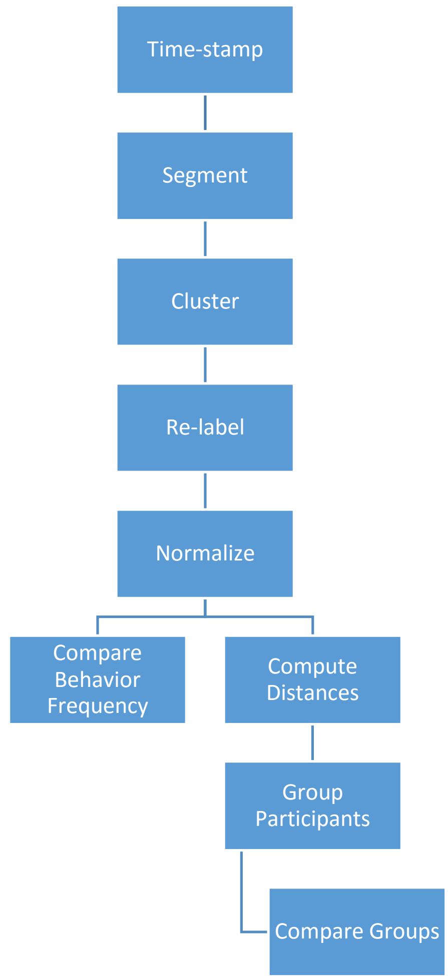


Figure 31. General algorithm used for comparing process (from top to bottom)

Time-stamp. The first step of extracting process data is to ensure that all data is properly time-stamped. This provides a means for synchronizing across the different modalities.

Segment. The time-stamped data is then segmented. Across all three analyses I segment the data every time a pair's structure is tested. Testing will be described in more detail later, but for now, the reader can interpret testing as representing an instance in which at least one person in the pair is eliciting feedback that will update the students on the current stability of their structure. Testing usually takes the form of a team member placing the weight on the structure.

Cluster[17]. The segmentation process yields several "test segments" for each student. These "test segments" are characterized by the proportion of time spent in each possible behavior. In the clustering step, similar "test segments" are grouped together.

Re-label. All "test segments" that are put into the same cluster are given the same name. Accordingly, each student's sequence of "test segments" can now be represented as a list of clusters.

Normalize. In the normalization step, each student's re-labeled sequence is lengthened so that I can more directly compare them to one another. The two forms of normalization that I use are L-1 normalization and dynamic time warping (Rabiner, Rosenberg, & Levinson, 1978). In the case of L-1 normalization, each sequence is lengthened so that all participants' sequences are of equal length. In dynamic time-warping a modification of Levenshtein distance (Levenshtein, 1966) is used to find the best match between pairs of sequences.

Compare Behavior Frequency. After L-1 normalization, the next step is to compare behavior frequency data across the three metrics of interest: success, experimental condition; and learning. The comparisons are based on Mood Median Tests along each of the individual clusters of "test segments." However, instead of the

[17] Clustering is the process of grouping together items that are similar to one another.

traditional Mood Median Test, which computes statistical significance based on a Chi-Square distribution, I use a binomial test. These two tests were used because the data did not meet the requirements for MANOVA and violated the typical requirements of a Chi-Square Test. This step represents the conclusion of one branch of the analysis tree.

Compute Distances. Following dynamic time warping, a distance is computed between each pair of participants.

Group Participants. Based on the pair-wise distances, similar participants are forced into one of two groups. In each case a student is put into the group that contains other students whose process was most similar to their own.

Compare Participants Groups. Finally, the groups are compared using a binomial test to determine the probability that individuals were randomly assigned to their specific group. Specifically, it is here that I examine the hypothesis that different groups, as partitioned by experimental condition, success on the activity, or based on post-test score, used markedly different processes from one another.

A more detailed explanation of this algorithm is provided in Appendix A.

Preliminary comparison of process via inclusion of principle-based items. Before delving into the three more complex analyses, I want to briefly present a preliminary process-based comparison of principle-based reasoning and example-based reasoning. This comparison utilizes intermediate design structures as the basis for identifying the extent to which students are making principle-based additions to their structures. This analysis tests the hypothesis that students in the principle-based reasoning condition were more likely to make a principle-based modification to their intermediate structures than students in the example-based reasoning condition. Confirmation of this hypothesis would support the claim that the principle-based reasoning condition primed students to see and use principles more easily than the example-based reasoning condition.

Intermediate designs were defined based on the segmentation procedure described above. Namely, an intermediate structure was defined as those structures that the students tested for stability[18]. Each intermediate design was compared to the previous design to identify principle-based modifications. Principle-based modifications included: (1) adding a base, (2) adding reinforcement, (3) adding triangles, (4) making strong connections and (5) adding symmetry. This set of principles expands on the two used in scoring student pre- and post-tests. The reason for using an expanded set when analyzing the intermediate structures, is because during the activity students are still in the process of learning, hence one would expect for them to experiment with a broader range of principles. However, in the pre and post-test, there is less of an expectation that students are tinkering, hence, it is appropriate to look at a constrained set of principles and focus on those principles that

[18] Intermediate structures only included system tests, and did not include structural tests. A structural test occurred when the student was only testing a single item, or component, but was not testing an actual structure. This distinction was made because structural tests were often completed on items that had not actually been built, but were only being prototyped. For more details about intermediate structures see Worsley & Blikstein (2013)

proved to be particularly useful or problematic based on previous studies. In conducting this analysis, I looked at the rate that principle-based items were added over the course of the activity by taking the total number of principle-based items divided by the total number of intermediate structures.

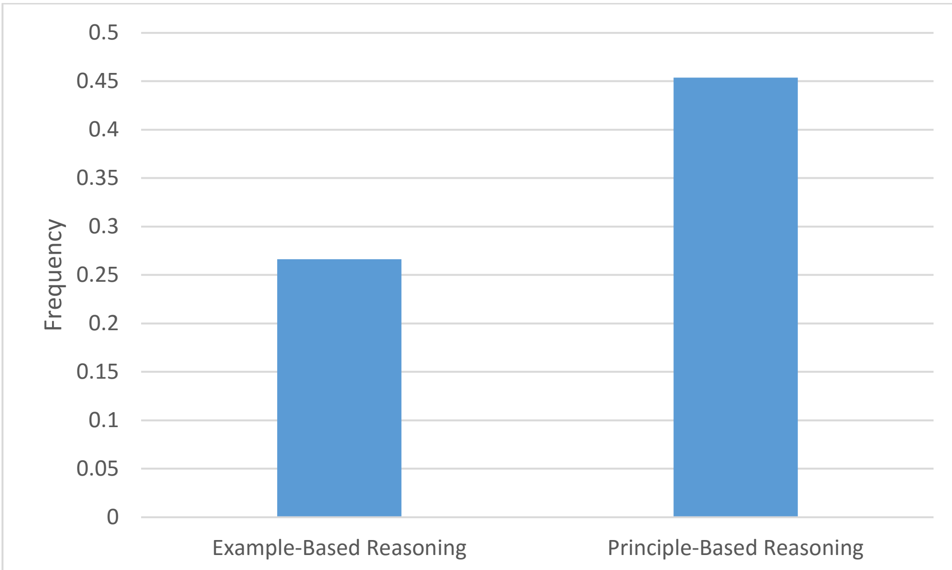


Figure 32. Frequency of Principle-Relevant Additions by Condition

When looking at the rate of principle-item inclusion in intermediate structures (Figure 32), the principle-based reasoning group ($M = 0.45$, $SD = 0.078$) has significantly more ($t(6)=2.02$, $p = 0.044$) than their peers in the example-based reasoning condition ($M = 0.26$, $SD = 0.16$). Pairs in the principle-based reasoning condition were almost twice as likely to add a principle-based item between intermediate structures. This result could stem from differences in the total number of principle-based items added, the total number of intermediate structures, or a combination of these variables. Based on the data, the most plausible explanation is that students in the principle-based reasoning condition are less likely to have consecutive structures in which no principle-based items were added. Hence, it is not that the principle-based reasoning condition used a larger number of principle-based items, nor is it the case that they had a smaller number of intermediate structures. Instead, students in the example-based reasoning condition simply had more intermediate structures that represented making minor changes, than their peers in the principle-based reasoning condition. This finding provides additional support for the claim that students who participated in the principle-based reasoning condition were more likely to take a principled approach to solving the engineering design challenge.

As an addition point of relevance for studying process, I also found that increased principle-based item inclusion was associated with increased learning scores. Namely, pairs of students with a higher frequency of principle-based item inclusion, received higher post-test scores (Figure 33). In particular, principle-based item inclusion, significantly predicted a pair of students' combined post-test score (β=5.5, t(8)=9.75, $p < 0.001$). Principle-based item inclusion also explained a large portion of the variance in learning scores (R^2=0.93, F(1,8)=95.16, $p<0.001$).

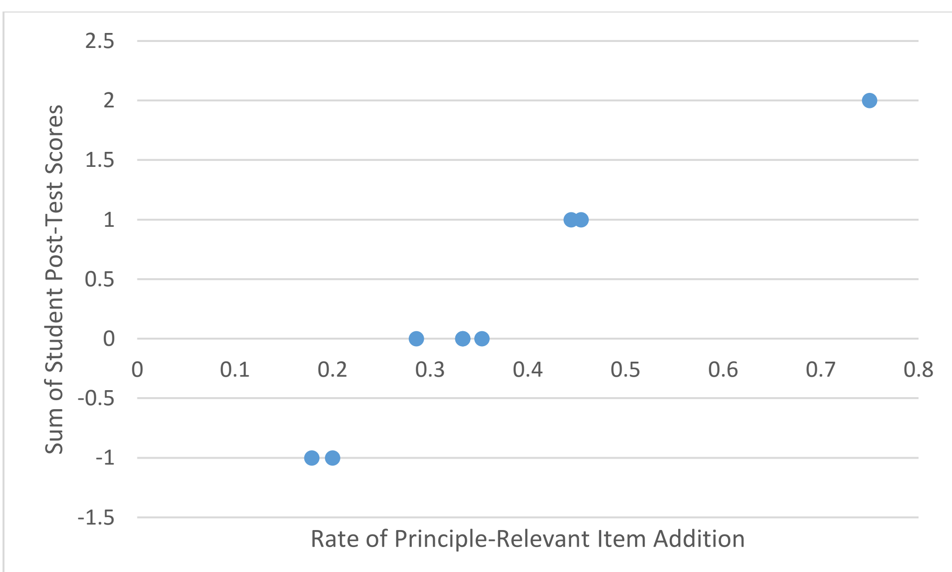


Figure 33. Combined post-test scores by the rate of principle-revelant item addition

Summary

During Part 1 of this chapter, I have looked to expand the scope of analysis by moving beyond success as the primary metric. The metrics of interest now include success, learning and process. I described learning as being related to the student's ability to generate ideas that solve a related open-ended problem. Process was invoked to highlight that the students were engaging in a dynamic activity. Throughout the task they are changing and learning. Accordingly, considering metrics that look beyond the end product are of considerable importance and relevance. In the Learning and Process sections I presented results that support the hypothesis that for this task principle-based reasoning is more efficacious than example-based reasoning. In particular I showed that students in the principle-based reasoning condition showed improvements from pre-test to post-test and made more frequent principle-based modifications to their intermediate structures than their peers in the example-based reasoning condition. These advantages are still maintained when controlling for prior ability, and were only observed after both the intervention and the building activity, suggesting that the principle-based reasoning intervention better prepared students to learn from the building activity. Furthermore, addition of principle-based items to intermediate structures strongly correlated with student learning, providing additional support for the analysis of student processes.

By adding learning and process as metrics of interest, I now have a total of four variables that I am analyzing: reasoning strategy, success, learning and process.

As one may have inferred from the previous chapter, my primary concern is in better understanding the nature of different reasoning strategies. However, I wish to show that these reasoning strategies have salience for various audiences, i.e. researchers, practitioners, parents and policy makers alike. For example, since the "making" community places a premium on the development of publicly shared artifacts, creating projects that are successful is important. At the same time, in order for "making" to have relevance to education, one has to demonstrate that students are learning. Finally, prior work in engineering education and constructionism has placed an emphasis on process. Hence, in validating that reasoning strategies matter, it seems appropriate to consider success, learning and process. That said, because the reasoning strategies are most clearly defined by the process students follow, comparing student processes is of greater significance to my argument that the metrics of success and learning.

Part 1 also included an initial argument for utilizing multimodal analyses. Given that multimodal analysis is traditionally presented as being an alternative to the text modality, one could consider the two analyses presented thus far as multimodal. In the case of assessing learning, I leveraged students' drawings; while in the case of assessing intermediate designs, I coded snapshots of student generated structures; neither of these constitute traditional text-based modalities.

In Parts 2, 3 and 4, I present analyses that more deeply employ the tools of computation, and that involve a more diverse set of modalities. In each part I will discuss the implementation of the general algorithm as applied to hand-coded data, multimodal sensor data, and a combination of hand-coded and multimodal sensor, data. Each section follows the same organizational structure and begins with a description and motivation for the particular data used. This is followed by a presentation of the common multimodal behaviors observed within the data set. In each of these presentations I include the four prototypical behaviors that emerged by starting with the most prevalent and ending with the least prevalent. Additionally, in describing each behavior I highlight the actions that characterize that cluster relative to the average values for the entire set of “test segments.” After presenting the common multimodal behaviors, I then proceed to show the extent to which the process similarity metric captures pertinent differences between the two experimental conditions. In the event that no correlations are found, I also consider how well the process similarity correlates with success[19] and learning[20]. Finally, after showing that the process similarity metric has relevance, I compare cluster frequency usage between groups of students with different experimental conditions, different success rates and different post-test scores. Preliminary implications of the finding are included at the conclusion of each section, and discussed more broadly in Chapter 4.

Part 2: Qualitative Analysis of Learning, Success and Strategies

One of the common strategies used for video data is to produce annotations of student behaviors. In many respects, producing timestamps any time a student begins a new action is tantamount to transcribing where the modality of interest is user behaviors. Prior work on multimodal analysis has leveraged this technique and shown

[19] Success is based on whether or not the pair’s structure was able to support to requisite amount of weight.
[20] Learning is based on student post-test scores.

it to be important for characterizing and understanding student learning (Barron et al., 2013; Kress, 2001). In similar fashion, in Part 2 I describe an analysis in which I time-stamped the video data, for every instance of six prototypical actions, or Object Manipulation Classes, as outlined in

Table 5. These actions are based on Worsley & Blikstein (2013, in press) which showed that this coding paradigm is a useful way for studying hands-on learning, and that the coding paradigm bears similarity to prior work in engineering education (Atman et al., 1999).

Table 5. Object Manipulation Classes

Class	Codes
PLAN	Prototyping ideas or inspecting the materials
EVALUATE	Testing a mechanism or testing the system
MODIFY	Making changes to an existing design
NOTHING	Not actively engaging in the activity
REVERT	Undoing one of more parts of a previous design
REALIZE	Putting pieces together as to make the structure

Common Behavior Analysis

The segmentation that follows hand-coding results in approximately two hundred unique “test segments.” For this analysis each “test segment” is defined based on the proportion of time spent in each of the five Object Manipulation Classes (REALIZE, PLAN, MODIFY, REVERT, NOTHING)[21]. Clustering those “test segments” resulted in four common behaviors, or clusters. Each cluster can be characterized by the relative proportion of time spent in each of the five activities. As

[21] EVALUATE does not appear because it was used to determine where to segment the data.

an overview, Figure 34 shows the distribution of the four common "test segment" types. The labels assigned to each region of the pie chart will make more sense based on the data presented in Figure 35 and in the following paragraphs.

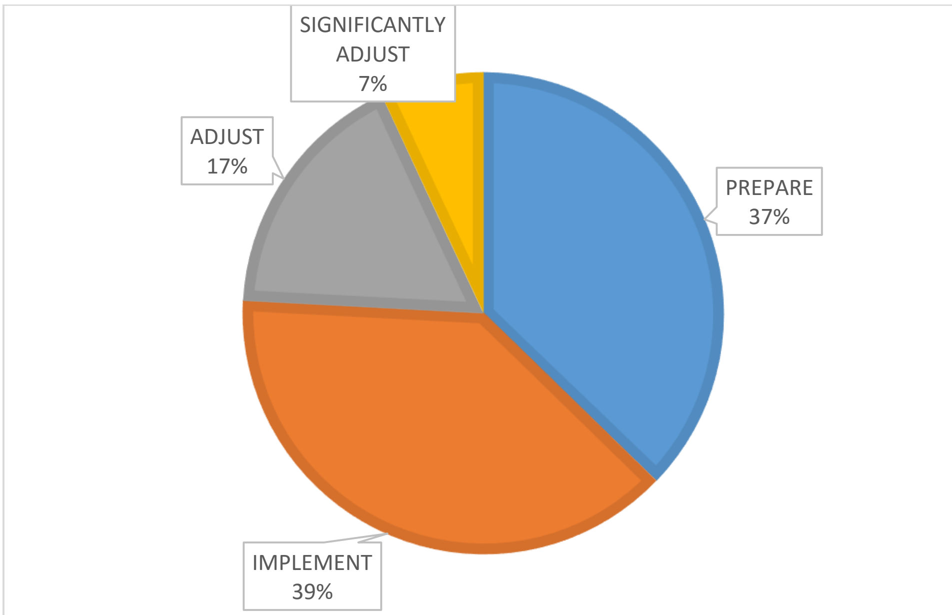


Figure 34. Relative frequency of common behaviors

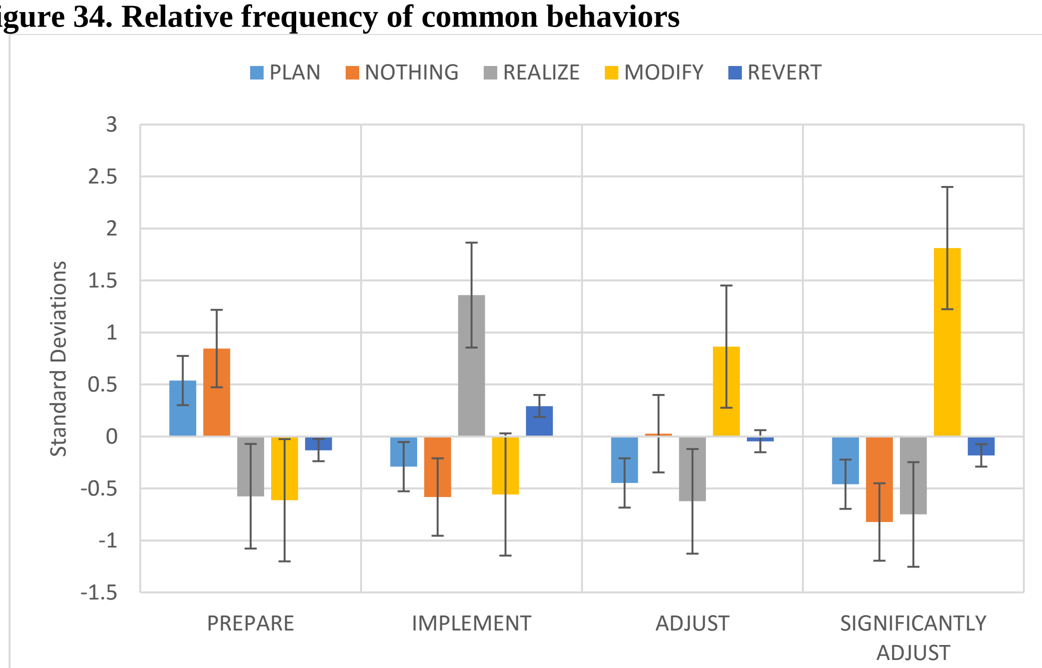


Figure 35. Characteristics of common behaviors for Part 2

The most frequently occurring cluster, IMPLEMENT, is characterized by significantly above average proportions of REALIZE and REVERT (Figure 36). The proportion of NOTHING is below average, as is PLAN. Accordingly this cluster seems to represent project or idea implementation in the absence of planning and/or modifying.

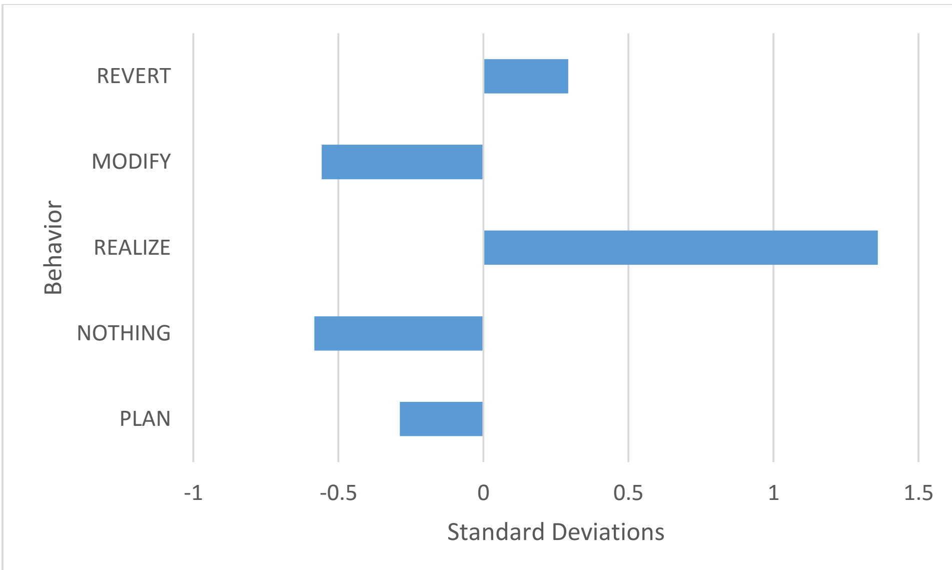


Figure 36. Relative Occurrences of Object Manipulation Classes for IMPLEMENT

Because the primary actions for this cluster of "test segments" involves either adding to an existing structure (REALIZE), or undoing an existing structure (REVERT), I call this cluster IMPLEMENT. That nearly two-fifths of the "test segments" are characterized by implementing an idea, is in line with the fact that this task is focused on hands-on manipulation of materials.

The second most frequently occurring cluster, PREPARE, is typified by above average PLAN behavior and above average NOTHING behavior (Figure 37).

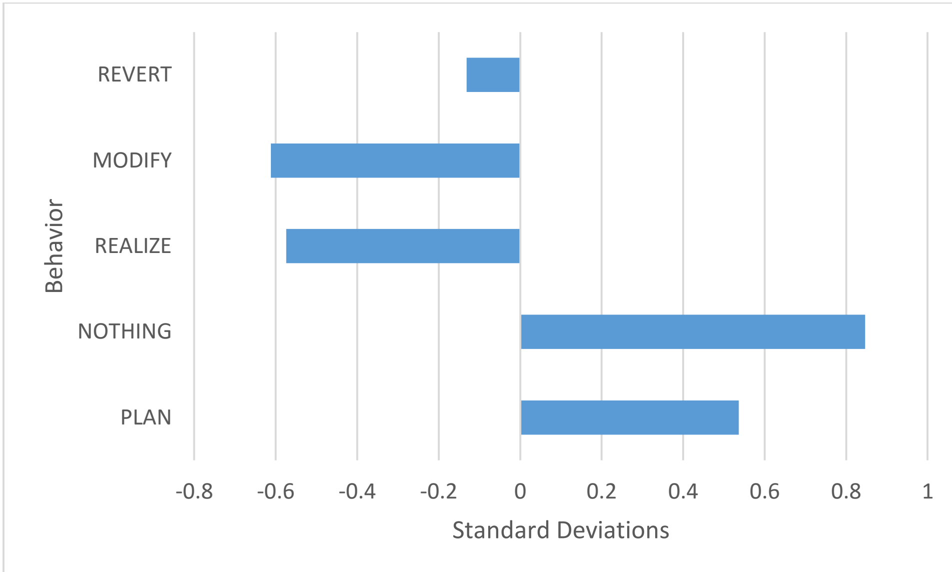


Figure 37. Relative Occurrences of Object Manipulation Classes for PREPARE

At the same time, this cluster also represents below average MODIFY, and REALIZE and appears to be roughly average for REVERT. I call this cluster PREPARE, as the students seem to principally be concerned with actions that are either explicitly or implicitly indicative of preparing to actually build. The fact that a large proportion of segments is spent doing PREPARE indicates that even though the focus of the activity is geared towards "making," many students are engaging in reflective processes that help them think about how best to complete the task. Furthermore, since several of the "test segments" are PREPARE segments, students are likely using PREPARE throughout the process.

The third cluster, ADJUST, is characterized by above average MODIFY, and below average PLAN and REALIZE (Figure 38). The behavior's average MODIFY value is approximately one standard deviation above the mean value for the entire population of "test segments."

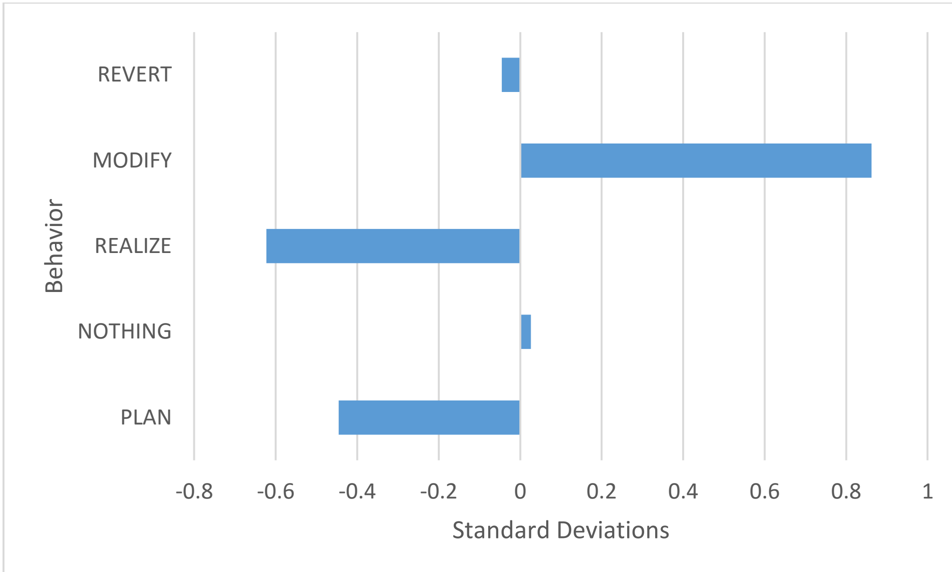


Figure 38. Relative Occurrences of Object Manipulation Classes for ADJUST

This indicates that when using this "test segment" students are spending a significant proportion of their time adjusting their structure, but may also occasionally spend a portion of the "test segment" doing nothing, or undoing. This, again seems reasonable. Based on my observations during data collection and video annotation, making adjustments to a structure was a fairly common activity. As can be seen from Figure 34, ADJUST is nearly 20% of all "test segments" across all users.

Like the third cluster, the fourth is also characterized by an above average proportion of MODIFY actions (Figure 39).

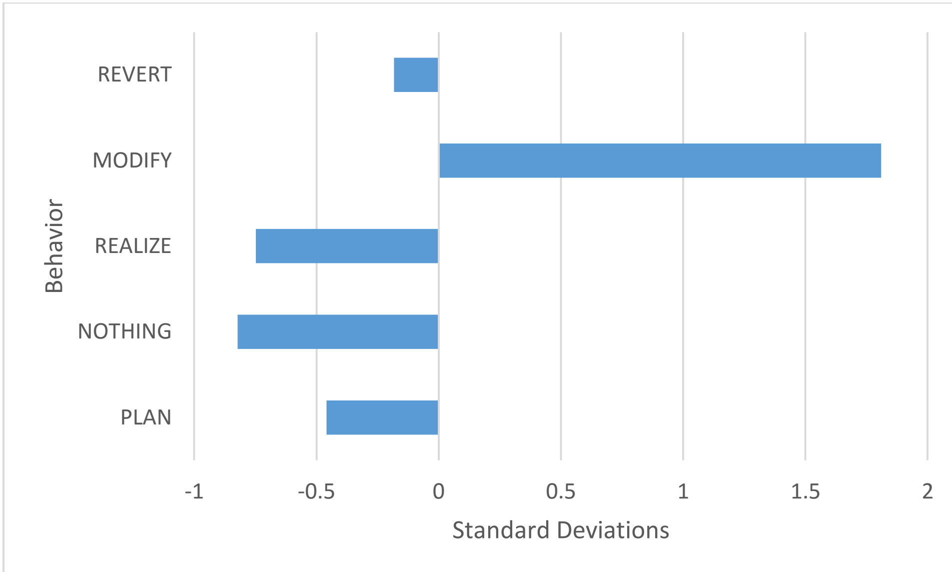


Figure 39. Relative Occurrences of Object Manipulation Classes for SIGNIFICANTLY ADJUST

However, whereas the ADJUST cluster involved MODIFY values that were one standard deviation above the mean, SIGNIFICANTLY ADJUST has MODIFY values that are closer to two standard deviations above the mean. To compensate for this increase in the proportion of time spent modifying, the proportion of time spent in REALIZE, PLAN and NOTHING are all well below average. In this case it appears as though these "test segments" are typified by students *only* making adjustments to their structures. Again, based on personal observation, this seems like an accurate characterization of several "test segments" as some students tried to make their structure work without a clear sense of how to do so. Because the focus almost exclusively resides in MODIFY, I call this cluster SIGNIFICANTLY ADJUST. Ten percent of the "test segments" were grouped into this cluster.

Process Similarity Comparison

Recall that the process similarity comparison groups students based on the pair-wise similarity of their processes. This particular metric maintains the order that each student completes each action, at the “test segment” level. Figure 40 shows the results of grouping students based on their process similarity, with a focus on comparing the number of student from each condition assigned to a given group.

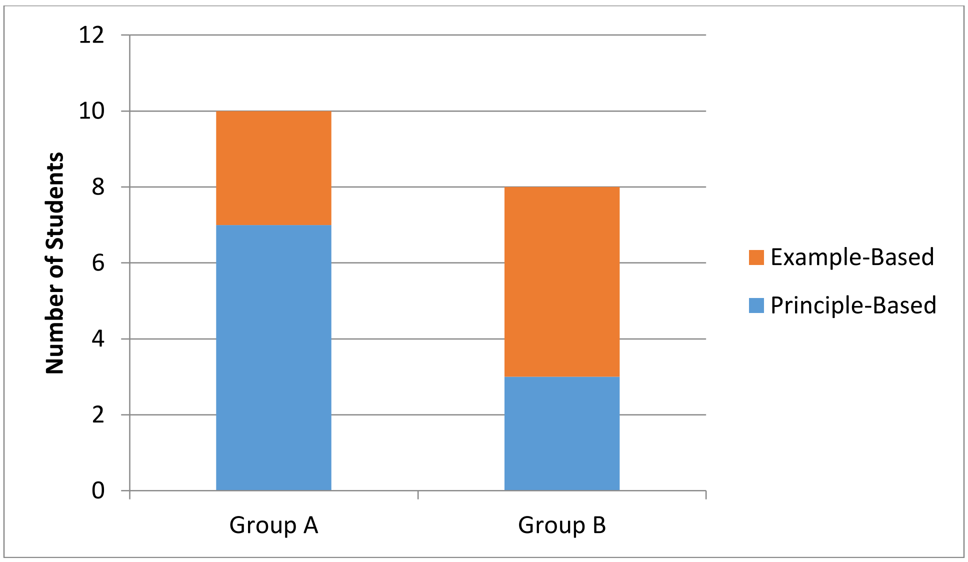


Figure 40. Composition of groups based on experimental condition as derived from process similarity for Part 2

Seven students from the principle-based condition were assigned to Group A, while the remaining three were assigned to Group B. For the example-based condition, three students were assigned to Group A, while the remaining five were assigned to Group B. According to a binomial test, there is approximately a 7% chance of this happening at random. While 7% is still a relatively small value, I am typically looking for results that have less than a 5% chance of representing spurious findings. Thus, instead of looking only at experimental condition, I also examine how Group A and Group B differ in terms of success (Figure 41).

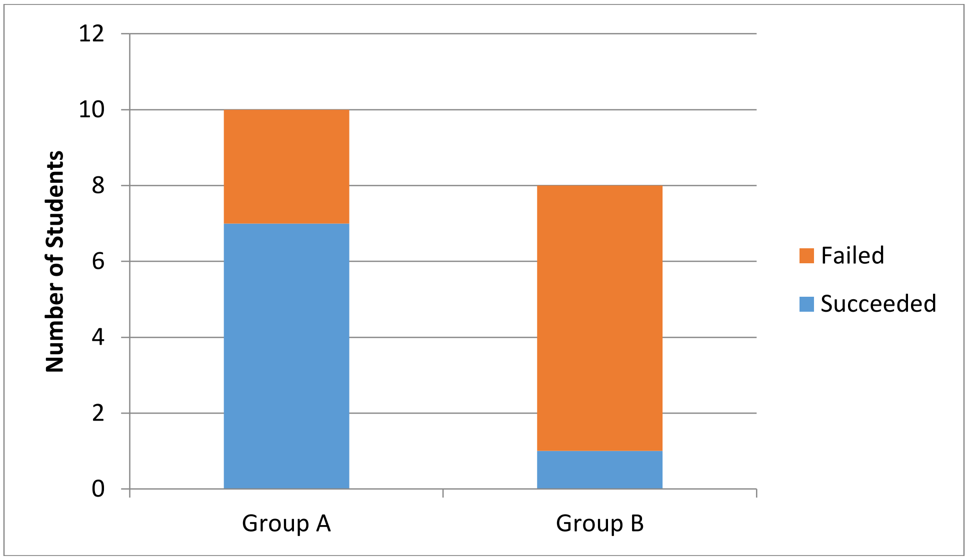


Figure 41. Composition of groups based on success as derived from process similarity for Part 2

Comparing Group A and Group B based on success rates produces a much clearer distinction. Seven of the ten students assigned to Group A succeeded on the activity, whereas only one of the eight students in Group B succeeded on the activity. The likelihood of this happening at random is less than 2%, suggesting that there were substantive process based differences between successful and unsuccessful students when considering their actions. To explore these differences more deeply, I now proceed to compare cluster frequency usage.

Comparing common behavior usage takes on two forms. At the most general level, it involves aggregating cluster frequency across the entirety of each student's process. However, in order to provide a more fine-grain comparison, I also look at cluster frequency usage in the first, second, and third portions of each student's process.

Coarse-grain cluster usage comparison

Figure 42 shows the median cluster frequency usage for the example-based and principle-based conditions. The most pronounced differences between the example-based condition and the principle-based condition is in the IMPLEMENT cluster. The principle-based reasoning condition makes significantly more (p = 0.011) use of the IMPLEMENT cluster than their peers in the example-based reasoning condition. This is the only dimension for which there are statistically significant differences between the two conditions.

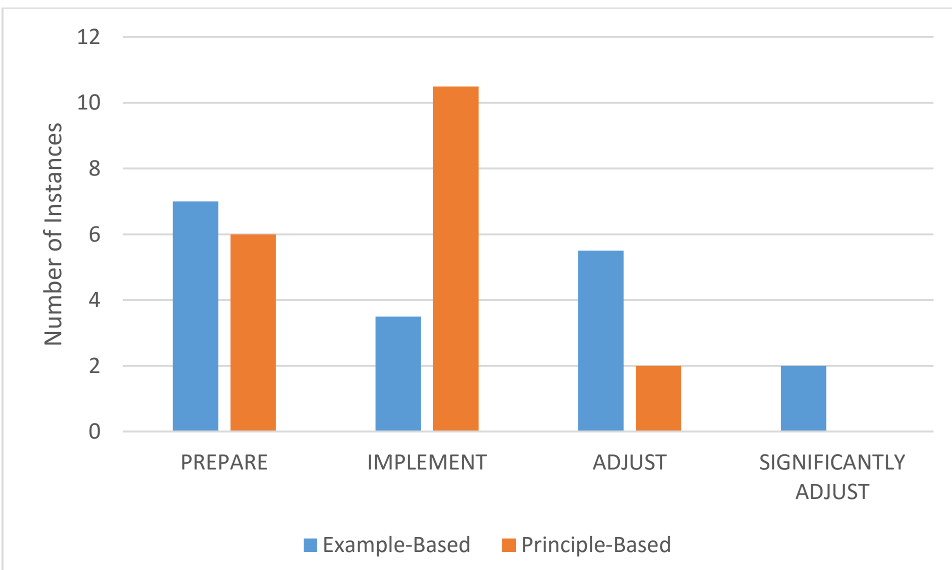


Figure 42. Median common behavior usage by condition for Part 2

Figure 43 contains the same analysis, but now with success as the dependent variable. The most pronounced differences appear to be in the IMPLEMENT and ADJUST categories. However, tests of statistical significance reveal that

IMPLEMENT is the only cluster for which successful and unsuccessful students significantly differ (p = 0.011). It is important to note, however, that while usage of SIGNIFICANTLY ADJUST was not significantly different, no successful students used that particular action.

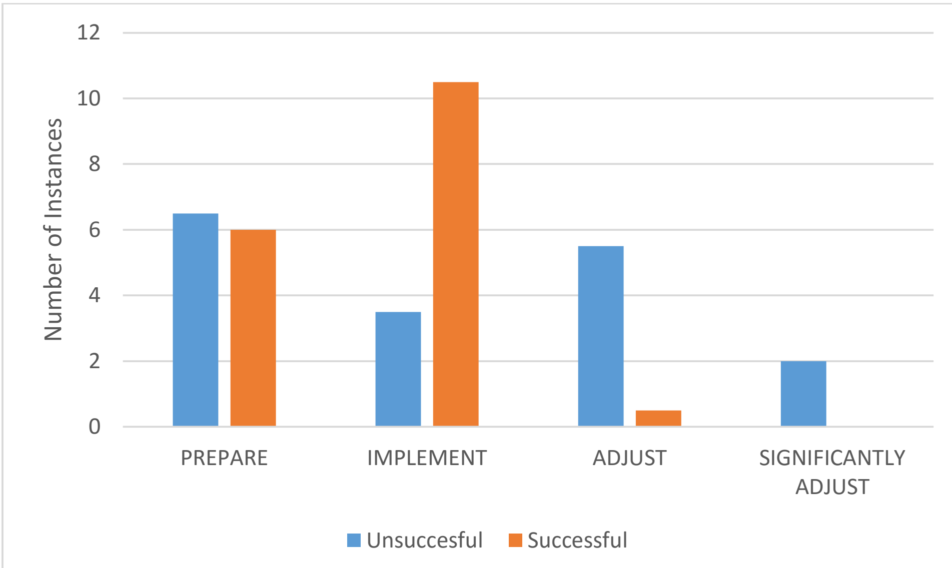


Figure 43. Median common behavior usage by success for Part 2

Finally, for the analysis of learning, I find that students who learned more spent relatively more time in IMPLEMENT (Figure 44). When I compare students with positive learning scores, with those receiving negative learning scores, I find a statistically significant difference (p = 0.0018). Again there is a trend that students who spend more “test segments” in IMPLEMENT spend fewer “test segments” in ADJUST and SIGNIFICANTLY ADJUST, but still spend approximately the same number of “test segments” in PREPARE.

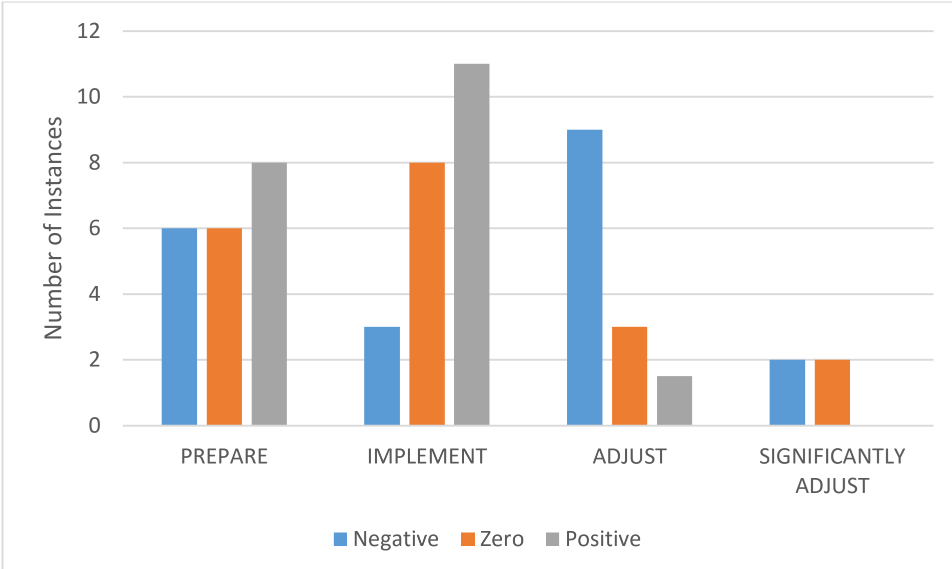


Figure 44. Median common behavior usage by learning score for Part 2

The coarse-grain analysis supports the hypothesis that students significantly differed in their processes when comparing experimental condition, success and learning. However, the coarse-grain analysis provides little in the way of describing where those differences are occurring and whether or not there is any causality in what is observed. To address this, the following section features a fine-grain analysis of cluster frequency usage that splits each student's process into three equally-sized parts.

Fine-grain cluster usage analysis

A fine-grained analysis across each part indicates that there are no significant differences between the two conditions. While there are places that have noticeable differences, those results are dropped after post-hoc analysis with Benjamini-Hochberg.

The fact that this particular algorithm appears to primarily be distinguishing successful students from unsuccessful students is reiterated through a fine-grained analysis of cluster usage. Specifically, there is a statistically significant difference ($p = 0.0003$) between successful and unsuccessful students in the amount that they use IMPLEMENT during the first third of the activity. Successful students were more likely to use IMPLEMENT, whereas the unsuccessful students were more likely to be in a ADJUST or PREPARE. Apart from IMPLEMENT usage in the first third, there were no differences between successful and unsuccessful students.

Much like the case of condition, the fine-grained analysis did not identify any statistically significant differences between students who received positive post-test scores, and those who received negative post-test scores.

Discussion

In this section I have presented results that confirm the hypothesis that student processes differed along several dimensions for the principle- and example-based experimental groups. I began by discussing the four common “test segment” types. These were termed PREPARE, IMPLEMENT, ADJUST and SIGNIFICANTLY ADJUST. I then moved on to show that the process similarity comparison yielded weakly significant results when comparing between experimental conditions. However, when looking at success rate, the process similarity metric did substantially better. I then proceeded to analyze how the cluster usage frequency data could be used to describe the differences observed in the process similarity metric. The coarse-grain analysis found that increased usage of IMPLEMENT was correlated with success, learning and the principle-based reasoning experimental condition. In interpreting this information it is important to recall that segmentation was based on when students tested and not based on the total amount of time spent on the task. Hence any attempt to argue that students spent more time in a given activity is not the appropriate inference to be made. Instead the results should be thought of in terms of the proportion of a students' test segments that were spent in a given activity, recalling that these can be of variable length.

Moving to the more fine-grain analysis provided additional insight into how successful and unsuccessful students differ in how they start the activity. Namely, successful students were likely to spend more of the first-third in IMPLEMENT, than unsuccessful students. That said, even though the coarse-grain analysis consistently reported that usage of IMPLEMENT was important, the analysis, on the whole still leaves many questions about why the two experimental conditions significantly differed from one another.

Part 3: Multimodal Analysis of Learning, Success and Strategies

In Part 2 of this chapter I used hand-annotated data to pinpoint differences in how students enacted the engineering design process. In Part 3, I transition into using automated multimodal sensor data. This multimodal data includes audio, hand/wrist movement and electro-dermal activation. Whereas the analysis in Part 2 included the semantics of each user's actions, the analysis in Part 3 will take a purely behavioral approach, but leverages multiple data streams in order to better capture the context in which each piece of data is recorded. This has commonly been a justification for undergoing multimodal analysis. Furthermore, prior research has studied how student posture and audio can be used as indicators for inferring student epistemological frames (Elby & Hammer, 2010; Hutchison & Hammer, 2009; Russ, Lee, & Sherin, 2012; Scherr & Hammer, 2009). Accordingly, in this analysis I will examine student behavior at a similar level of granularity and identify the amount of audio, hand/wrist movement and electro-dermal activation that students generate at different points in time.

Data Collection

The decrease in the cost and complexity of advanced sensing technology has greatly reduced the difficulties of capturing high frequency sensor data. For example, the Xbox Kinect sensor allows researchers to capture audio, video and gesture data with relative ease. The Kinect sensor was among the devices that I used for this multimodal analysis. In addition to the Kinect sensor I also used a high resolution web camera and a pair of Affectiva Q Sensors.

Audio data. Audio data was captured from the Kinect sensor and an overhead web camera. These two data channels were merged and processed to compute the presence or absence of audio at each time-step.

Hand/wrist movement data. The Kinect sensor also provided bodily displacement data for each participant. From this I was able to extract wrist and hands displacement at each time step. Previous work using Xbox Kinect gesture data has focused on the technology as an input device that students can use to control or interact with a computer-based system. In line with this approach, many researchers have been interested in detecting basic arm, hand and wrist gestures. Nonetheless, there are additional opportunities to use this data to study two-hand coordination, and joint gestural attention between participants. In this study, however, I examine hand/wrist movement at a more behavioral level.

Electro-dermal activation data. Data from a Q-sensor was used to determine average electro-dermal activation at each time step. Electro-dermal activation is associated with arousal, galvanic skin response and stress. Devices within this class provide relatively sensitive measurements of the amount of sweat being effused from the skin. They also include three-axis accelerometer data, as well as skin temperature data.

Appendix A goes into more detail about the specifics of multimodal data extraction from these three data sources.

Common Behavior Analysis

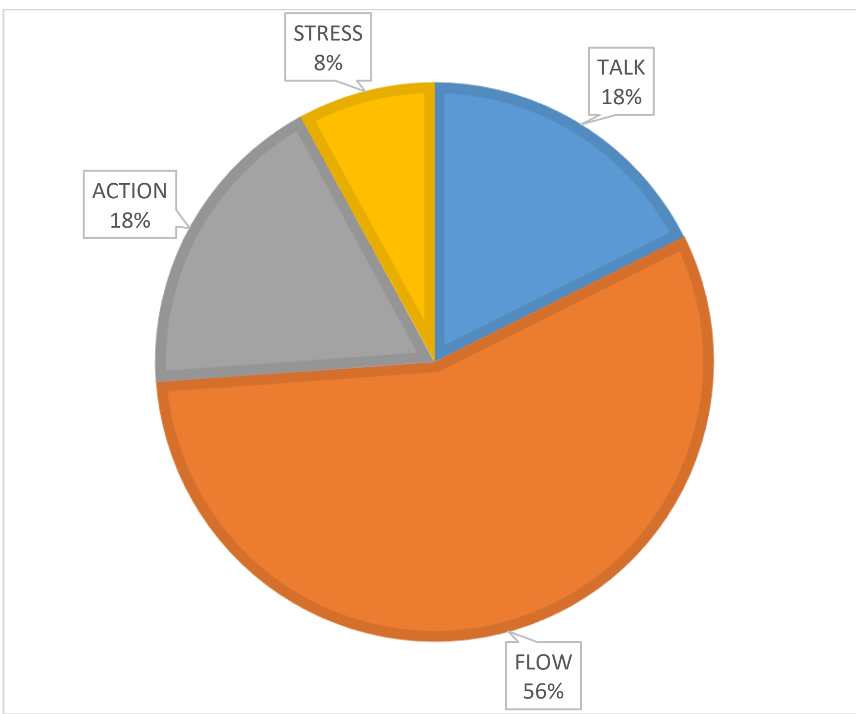


Figure 45. Relative frequency of common behaviors for Part 3

Figure 45 shows the relative frequency of each of the characteristic "test segments." Again, the labels will be more understandable following the discussion of each common behavior and their accompanying graphical representations which are summarized in Figure 46.

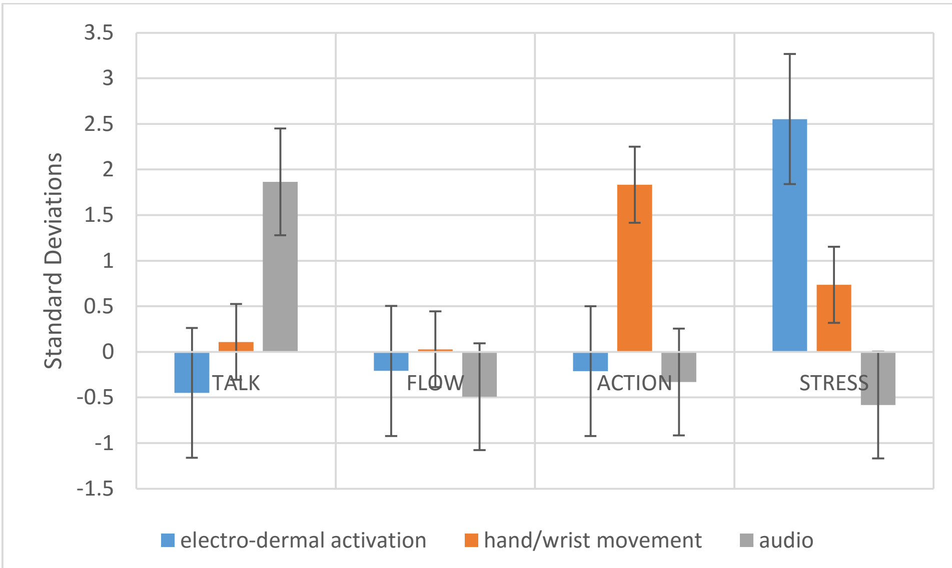


Figure 46. Characteristics of common behaviors for Part 3

The most common segment, which I call FLOW, is characterized by near or below average behavior across all three variables: audio, hand/wrist movement and electro-dermal activation (Figure 47). When examined in comparison to the other common behaviors (Figure 45), FLOW is the most balanced, with no individual modality significantly dominating the others. This cluster represents roughly 60% of all "test segments."

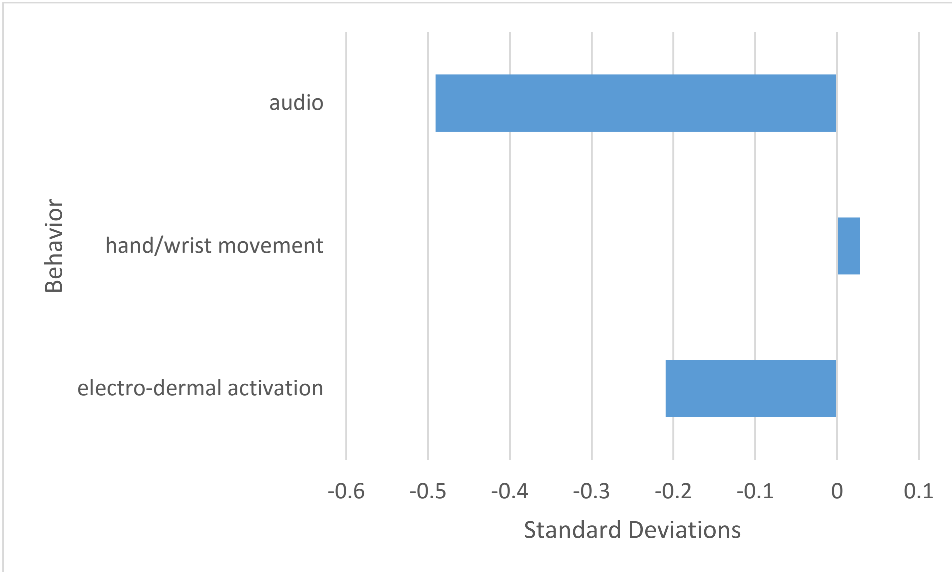


Figure 47. Relative Occurrences of Multimodal Behaviors for FLOW

As one can likely deduce, calling this cluster FLOW is a reference to Csikszentmihalyi (1992). As I compare the usage across conditions, rate of success and quality of learning, the argument for calling this category FLOW will become clearer. For now, suffice it so say that this cluster represents the vast majority of all "test segments", and that it is typified by average stress, average movement and little speech, providing initial indication that this state could be in line with Csikszentmihalyi's flow. One can picture students in FLOW concentrating on the task by carefully manipulating the materials without the need for extended discussion, movement or arousal.

The second most frequently occurring "test segment" is one that I have entitled ACTION. This behavior primarily consists of students who are currently engaging in above average hand/wrist movement (Figure 48).

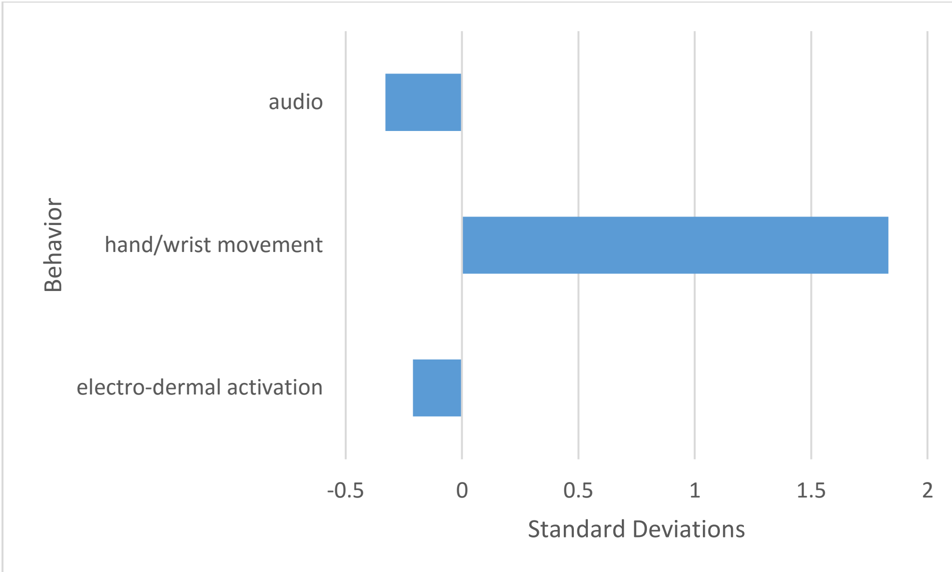


Figure 48. Relative Occurrences of Multimodal Behaviors for ACTION

What's more, though, is that this occurs in the absence of high electro-dermal activation, which is normally correlated with body movement. An additional point of interest is the lack of audio associated with this behavior. Students are focused on building and refraining from extensive discussion with one another. Accordingly, one might conjecture that the students are findings other means through which to communicate with one another.

After ACTION, the most frequently occurring state is TALK. This particular cluster represents approximately 18% of all "test segments."

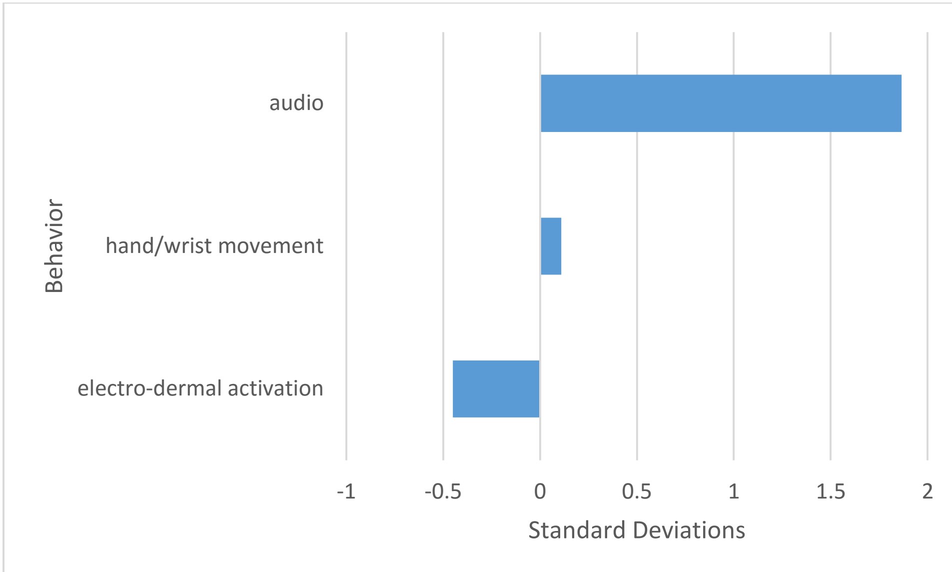


Figure 49. Relative Occurrences of Multimodal Behaviors for TALK

Based on Figure 49 the amount of audio in this cluster is approximately two standard deviations above the mean. Hand/wrist data is just above the mean, and electro-dermal activation is nearly half a standard deviation below average. Again, this is analogous to ACTION in that students appear to only engage one of the multimodal behaviors at a given time.

The final cluster is one that I call STRESS. This behavior is characterized by extremely large values of electro-dermal activation, as well as above average hand/wrist movement (Figure 50).

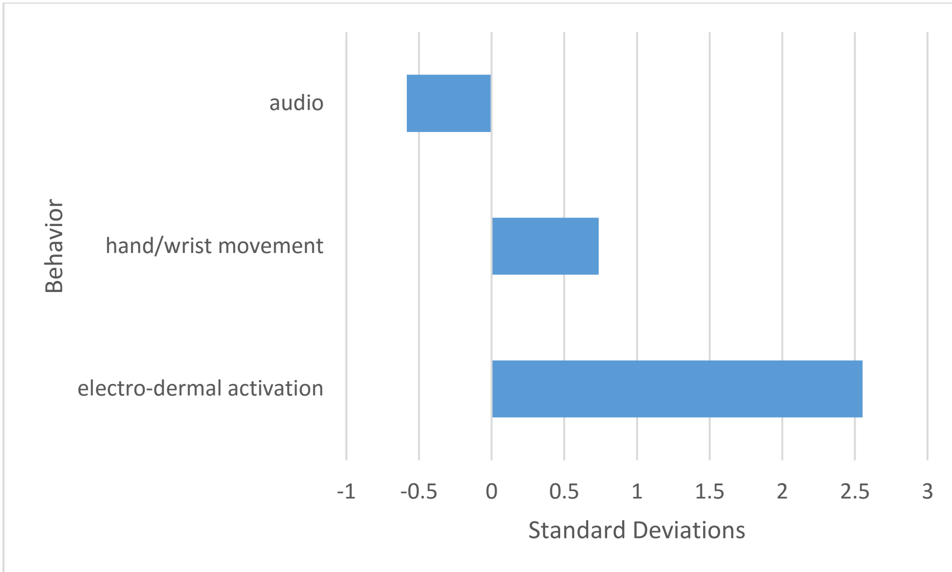


Figure 50. Relative Occurrences of Multimodal Behaviors for STRESS

As mentioned before, one would expect for electro-dermal activation and hand/wrist movement to correlate with one another. Hence we can anticipate that the electro-dermal activation values may be slightly inflated, but are still, more than likely, well above average. On the other hand, that the students are stressed, may be causing them to work more frantically, which would result in an increase in hand/wrist movement. STRESS accounts for approximately 10% of students "test segments."

Process Similarity Comparison

Before discussing the specifics of the process differences I first present results from grouping students based on the similarity of their processes. Process similarity was based on the two participant clusters created from the pair-wise comparison of student sequences. Seven of the eight students assigned to Group A are from the principle-based condition. The inverse pattern is observed for Group B, with seven of the eight individuals in that group coming from the example-based reasoning condition (Figure 51). The likelihood of this happening randomly is less than 0.002, suggesting that the two conditions did, in fact, utilize markedly different processes.[22]

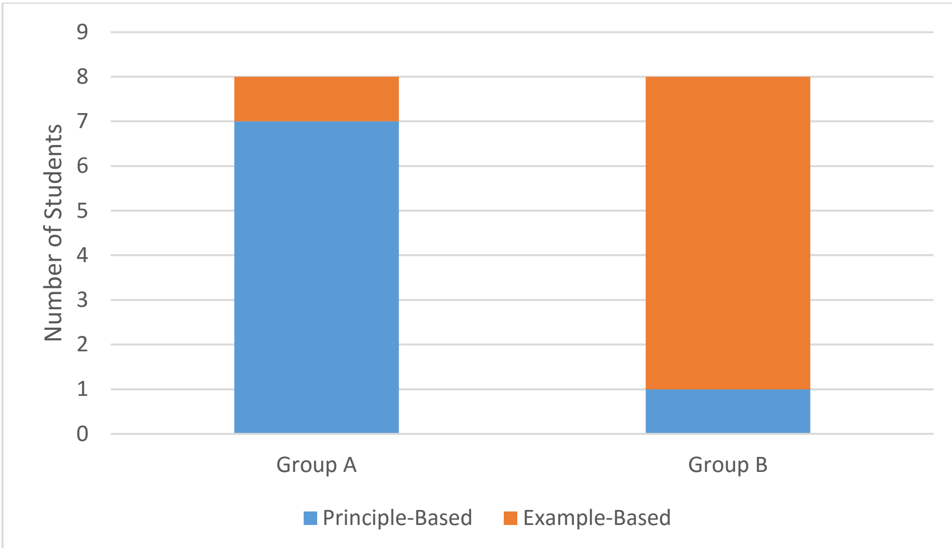


Figure 51. Composition of groups based on experimental condition as derived from process similarity for Part 3

Having observed that there are salient differences between the processes that the two conditions use, as determined through multimodal data, I now consider the nature of those differences. As in Part 2, I examine cluster usage at coarse- and fine-grains.

[22] One question is if these results are being inflated by the dyadic nature of the task. For example, two individuals that work together are likely to mirror each other's behavior. When I do an analysis to determine how frequently a given student's process is most similar to that of their partner, I find that this is only the case for two of the eight pairs, and if I remove the partner from consideration, that student is still likely to be most similar to another individual from the same experimental condition.

Coarse-Grain Cluster Usage Analysis

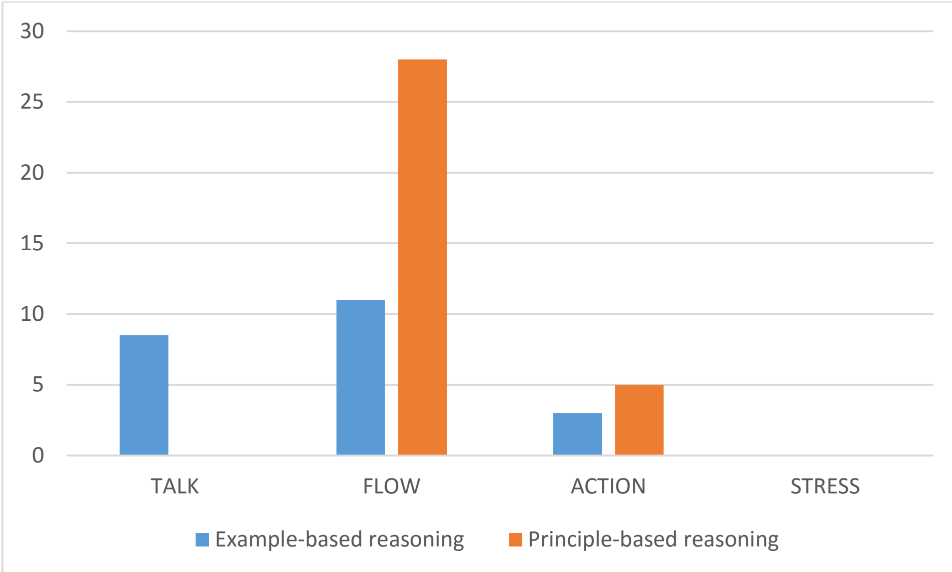


Figure 52. Median common behavior usage by condition for Part 3

Figure 52 shows the median normalized frequency of cluster usage by experimental condition. From the figure it is apparent that the median value for FLOW is quite different between the two conditions. A test of statistical significance confirms that students in the principle-based reasoning condition were more frequently ($p = .007$) in FLOW than their peers in the example-based reasoning condition. No other statistically significant differences emerged between the two groups when comparing their cluster frequency.

In terms of success there are no statistically significant differences in terms of cluster, or behavior, usage (Figure 53).

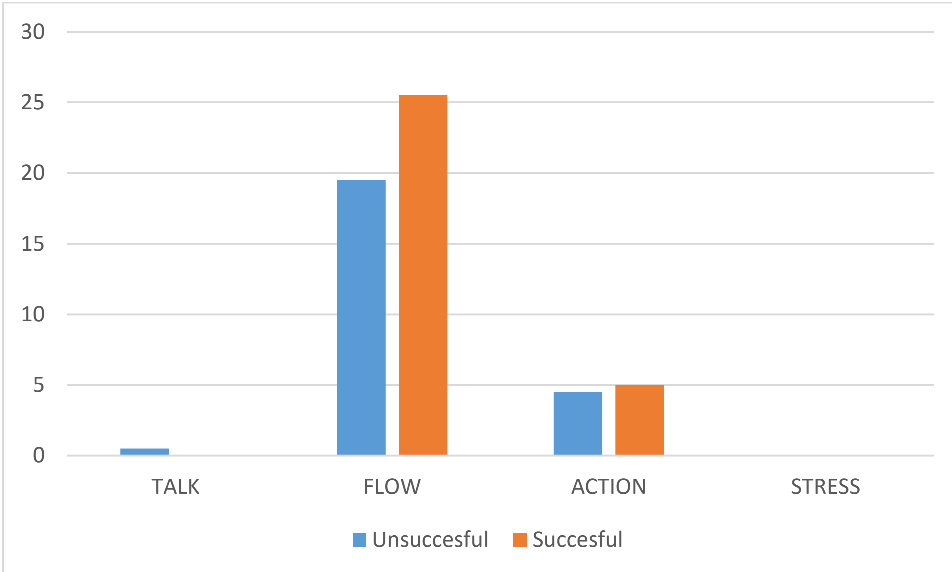


Figure 53. Median common behavior usage by success for Part 3

Much like the case of success, the data does not reveal any significant differences between students who experienced positive learning, and those who experienced negative learning (Figure 54).

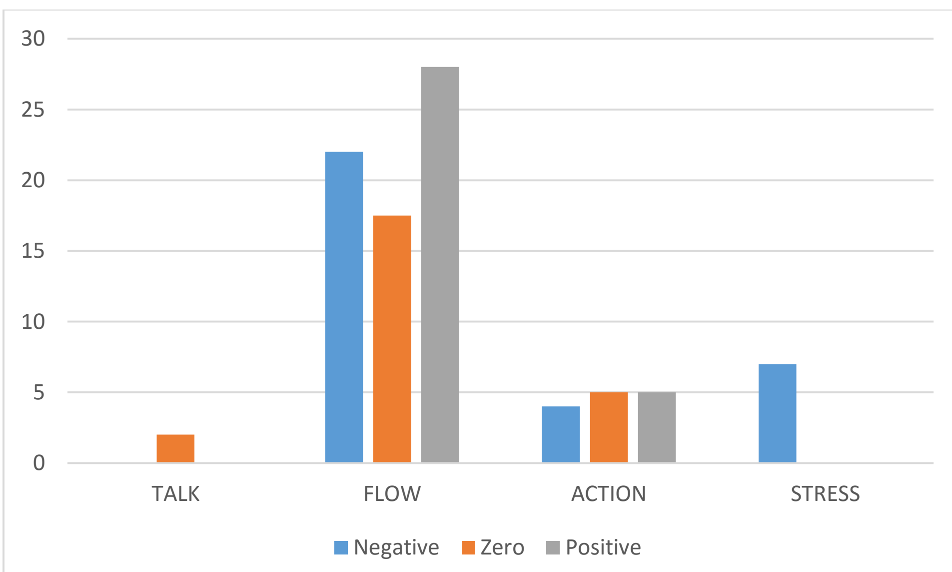


Figure 54. Median common behavior usage by learning score for Part 3

Fine-Grain Cluster Usage Analysis

A fine-grained analysis indicates that there are significant differences between the two conditions during all three portions of the activity. Students in the principle-based reasoning condition are more likely to be in FLOW for the first ($p < 0.001$ (1.9×10^{-5})), second ($p = 0.007$) and third ($p = 0.007$) thirds, than their peers in the example-based reasoning condition. This suggests that the two conditions differed at all three stages, but that the greatest divergence occurred during the first portion of the activity. In particular, many of the students from the example-based reasoning condition primarily spend the first third of the activity in ACTION. In a later section, I discuss the implications of this in more detail.

The fine-grained analysis did not identify any significant differences between successful and unsuccessful students in terms of how frequently they used the different clusters, or common behaviors.

The fine-grained analysis did indicate that students that learned more through the activity were more likely be in FLOW during the first third of the activity. Specifically, the students that received positive post-test scores were much more likely ($p < 0.001$ (0.0004)) to use FLOW than students who received a negative post-test score. This trends continues among the students that received a score of zero, but not at a statistically significant level.

Discussion

In deciphering the differences between the principle-based reasoning condition and the example-based reasoning condition, the current multimodal analysis offers a significant improvement above the analysis from Part 2. Specifically, this particular analysis included statistically significant differences in the process similarity metric between the experimental conditions. Furthermore, both the coarse-grain and fine-grained analyses offered some additional insight into identifying the elements of each condition's process that differed. Specifically, coarse and fine-grained analyses showed that students in the principle-based reasoning condition made more extensive use of FLOW than students in the example-based reasoning condition. That difference emerged during all three portions of the activity and was most pronounced during the first third. This is telling because it indicates that the differences were not merely the result of students being more or less successful on the activity. In fact, this analysis did not reveal any differences between successful and unsuccessful students at any grain size. Instead, the only other difference was observed from the learning metric. Once again, FLOW was positively correlated with student learning.

Realizing that the analyses from Part 2 and Part 3 provided different benefits, in Part 4, I will examine the merits of combining the two approaches with the hope of identifying differences along all three metrics: experimental condition, success and learning.

Part 4: Combined Analysis

Part 2 leveraged semantic-level descriptions of each student's actions as the means for exploring the hypothesis that processes significantly differed. This analysis concluded that the use of IMPLEMENT was positively correlated with success, learning and the principle-based reasoning condition. However, grouping students based on their processes only yielded significant results in terms of success, and didn't have a strong correlation with experimental condition. In Part 3, I presented an analysis that involved automatically-derived, behavioral data. This analysis effectively distinguished between students from the different experimental conditions, based on their differential usage of FLOW. FLOW was also important for predicting student learning. Additionally the process similarity metric confirmed these results by creating two groups that almost perfectly align with the two experimental conditions. However, the analysis did not provide much in the way of determining the behaviors associated with success. Having garnered different benefits from each analysis, one can’t help but wonder if combining approaches would provide the quintessential multimodal analysis. Thus, in Part 4 I combine data from Part 2, with data from Part 3 in an effort to push the limits of multimodal analysis, and explore the possibility of generating results that highlight differences in condition, success and learning.

The data for this analysis includes an eight dimensional vector (PLAN, REALIZE, MODIFY, REVERT, NOTHING, audio, hand/wrist movement and electro-dermal activation). As before, I begin by presenting the most common clusters of behavior among the population of research participants.

Common Behavior Analysis

In describing the clusters, I used names that make reference to the cluster names in Part 2 and Part 3, where appropriate. Figure 55 and Figure 56 contain the relative frequency and characteristics of each behavior, respectively. As before, the following paragraphs will be used to describe each of the common multimodal behaviors.

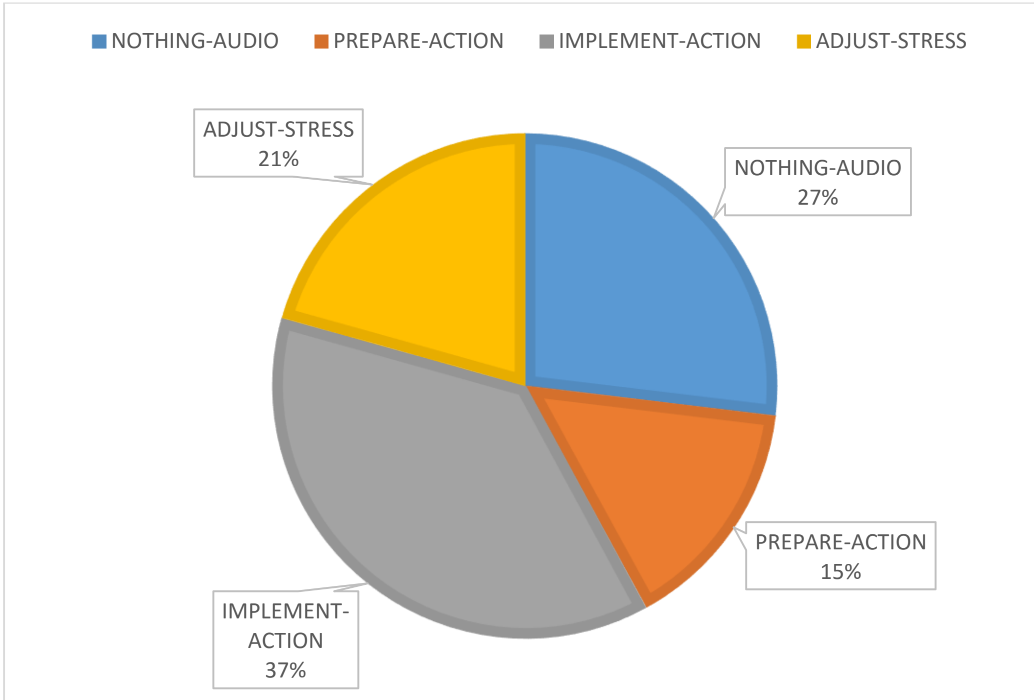


Figure 55. Relative frequency of common behaviors for Part 4

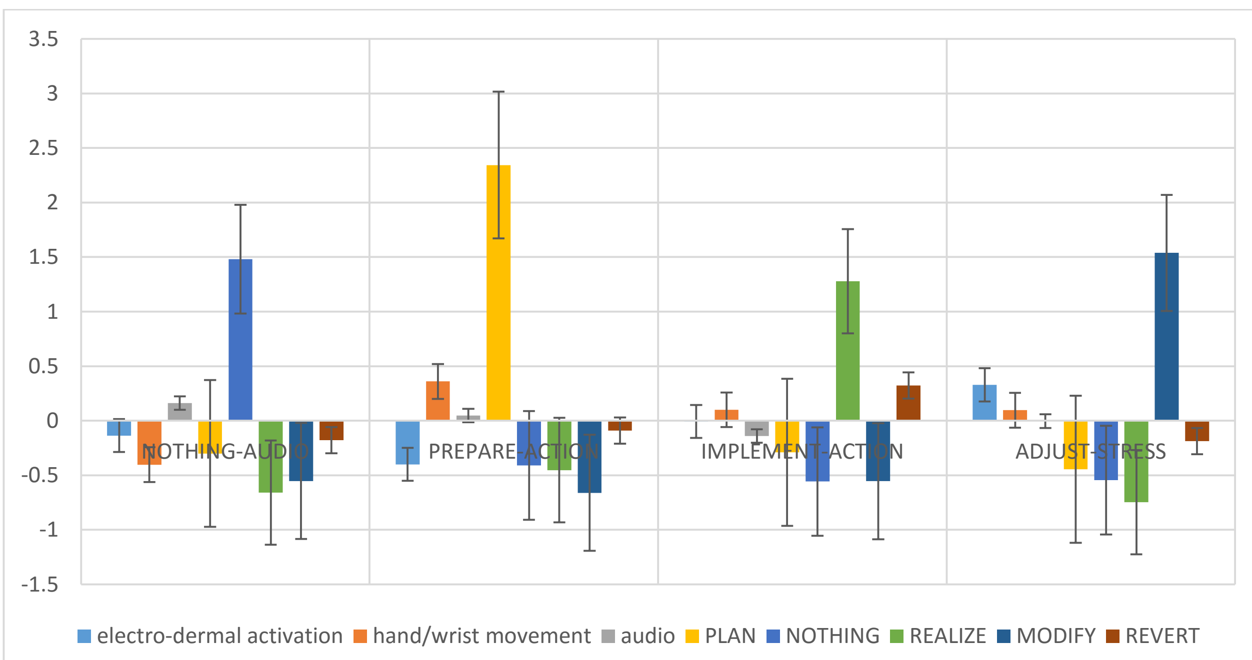


Figure 56. Characteristics of common behaviors for Part 4

The first cluster that I describe is the IMPLEMENT-ACTION cluster (Figure 57). This cluster represents 37% of all "test segments" and is typified by significant

REALIZE, slightly above average hand/wrist movement and slightly above average REVERT.

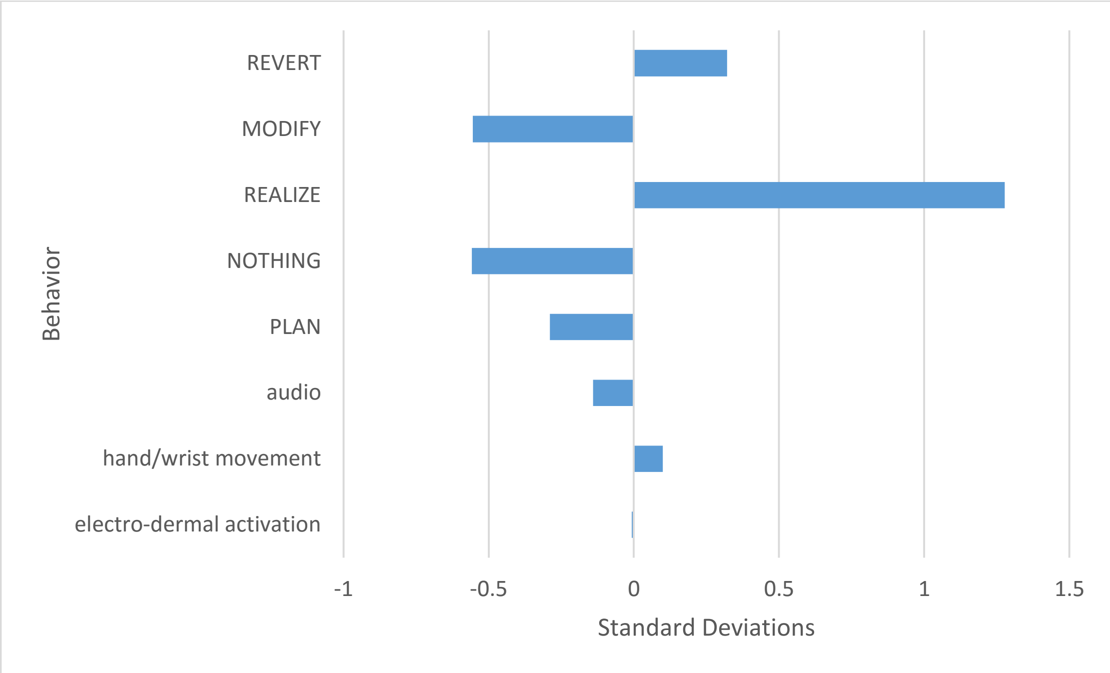


Figure 57. Relative Occurrences of Multimodal and Object Manipulation Classes for IMPLEMENT-ACTION

Values for PLAN, electro-dermal activation and audio are also approximately average, while values for NOTHING and MODIFY are generally below average. I interpret this data as being associated with students actively working towards physically completing their structure. Above average REALIZE very clearly keys the reader into this fact. The additional modalities measured then provide insights into the other behaviors associated with project implementation, namely, hand/wrist movement.

The second more frequent cluster that I describe is the PREPARE-ACTION (Figure 58) cluster which constitutes 17% of all "test segments." Much like the PREPARE cluster in Part 2, this cluster is characterized by above average PLAN.

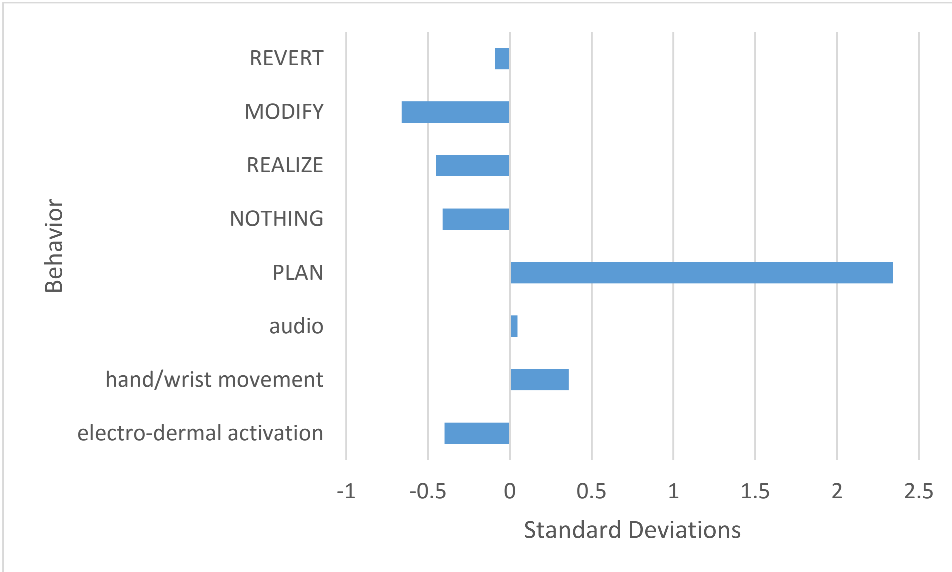


Figure 58. Relative Occurrences of Multimodal and Object Manipulation Classes for PREPARE-ACTION

However, it differs in that NOTHING is not associated with this particular behavior. Furthermore, there is greater hand/wrist movement associated with PREPARE-ACTION than for IMPLEMENT-ACTION. This runs contrary to the initial assumption that ideating about one's project does not require extensive hand/wrist movement. In this case, the students use more body gestures while planning than while engaging in IMPLEMENT-ACTION. All of the other values are well below average for this common behavior. In particular, this cluster is associated with the lowest average value for MODIFY. Because one would not expect for a students to make changes to their structure while in the planning phase this result seems reasonable.

The third cluster is NOTHING-AUDIO (Figure 59). As the name implies, it is associated with students appearing to do nothing and occasionally engaging in dialogue.

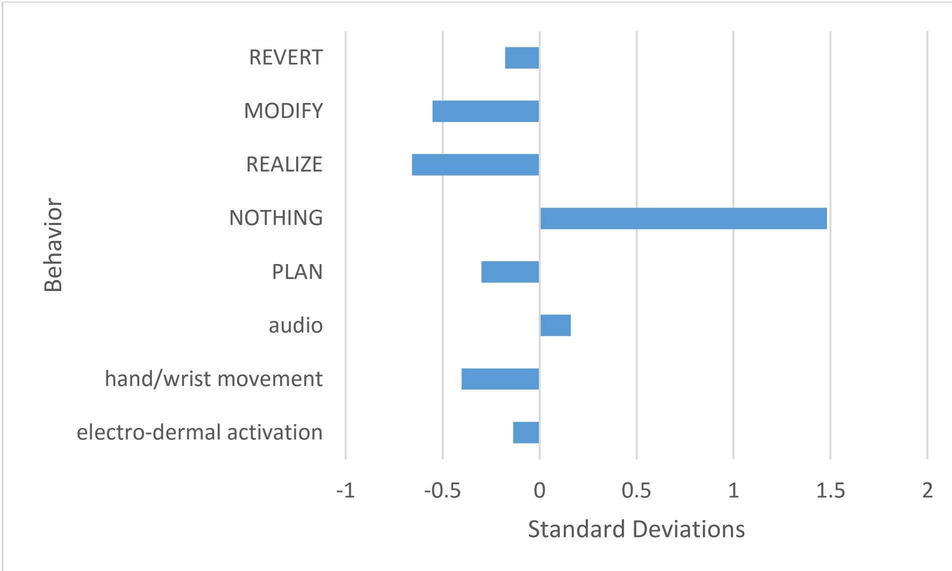


Figure 59. Relative Occurrences of Multimodal and Object Manipulation Classes for NOTHING-AUDIO

This behavior has the highest average audio value and the lowest hand/wrist movement value. As such, combining the data streams provides a new way to think about the multimodal behaviors associated with planning and appearing to do nothing. It also offers validation that the coding of the Object Manipulation Classes was consistent. Specifically, NOTHING is associated with well below average values of hand/wrist movement.

The final behavior is ADJUST-STRESS (Figure 60). This behavior accounts for 22% of all “test segments.”

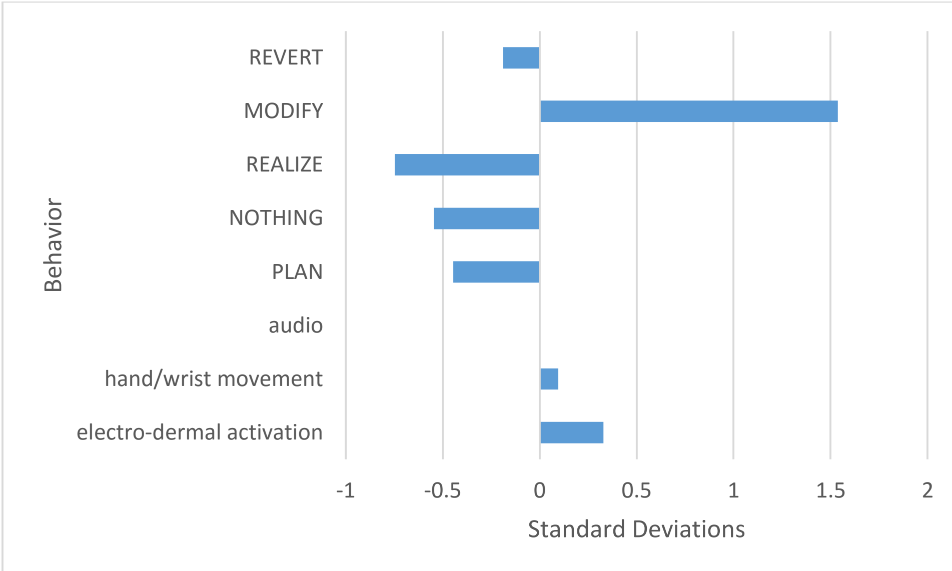


Figure 60. Relative Occurrences of Multimodal and Object Manipulation Classes for ADJUST-STRESS

Recall that the previous analyses contained two behaviors associated with adjusting (ADJUST and SIGNIFICANTLY ADJUST) and one associated with high electro-dermal activation (STRESS). In combining the two analyses, one of the common behaviors that emerges lies at the intersection of the previously identified behaviors (from Part 2 and Part 3). Amidst this common behavior, the student is unlikely to participate in PLAN, NOTHING or REALIZE, and is instead focused on modifying their design.

Process Similarity Comparison

Despite the increased contextualization afforded through the combination of multimodal sensor data and the hand-coded data, the results of a process similarity analysis are less than stellar. Group A consists of five students from the example-based reasoning condition, and four from the principle-based reasoning condition. The remaining seven students are placed in Group B (Figure 61).

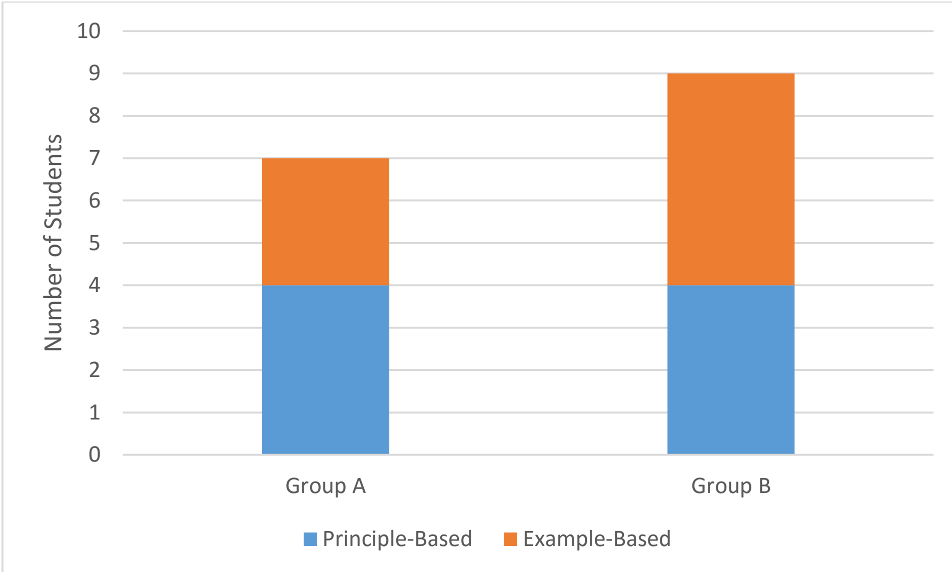


Figure 61. Composition of groups based on experimental condition as derived from process similarity for Part 4

The lack of differentiation, in terms of the process similarity metric, is also observed when comparing successful students with unsuccessful students (Figure 62). Similarly, the process similarity metric results do not align to student learning scores. From this standpoint, combining the data streams did not produce the desired result. By all accounts this provides a good indication that embarking upon this line of analysis may be less fruitful than the two individual analyses in Part 2 and Part 3. Even so, as an exercise in completeness, I still use the upcoming sections to investigate coarse- and fine-grain differences in common behavior usage.

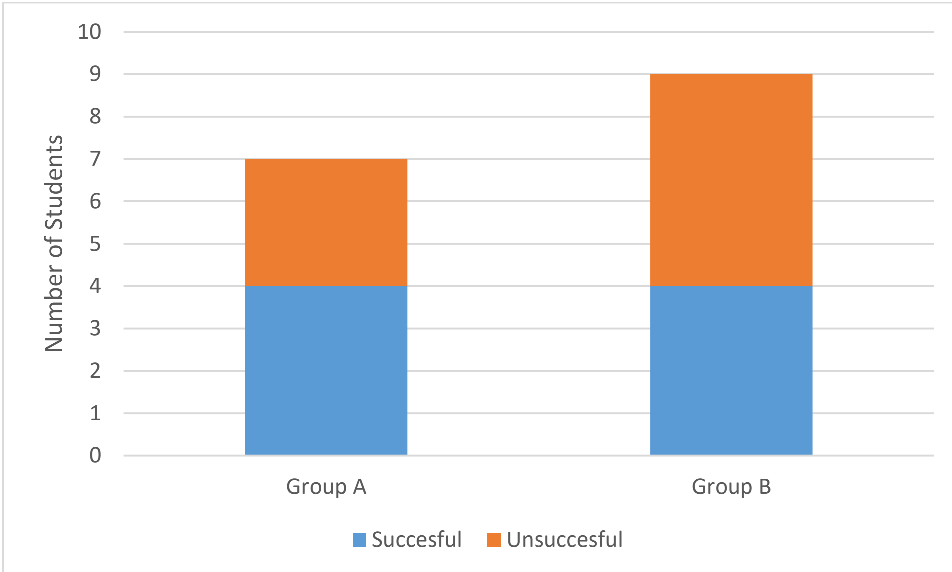


Figure 62. Composition of groups based on experimental success as derived from process similarity for Part 4

Coarse-grain Cluster Usage Analysis

A coarse-grain analysis of cluster usage between conditions suggests that there are no differences between the two conditions (Figure 63). The median value for ADJUST-AUDIO and PREPARE-AUDIO appear to be markedly different between conditions, but these differences aren't statistically significant. This lack of results is somewhat surprising given the consistent results observed in Part 2 where the IMPLEMENT state was associated with success, learning and the principle-based reasoning condition. Here IMPLEMENT-ACTION occurs at equal rates between conditions. One potential implication of this finding is that students enact IMPLEMENT differently within each condition. Since adding hand/wrist movement distorted the IMPLEMENT based differences between conditions, one can infer that students in the principle-based reasoning condition likely completed more IMPLEMENT in the context of smaller scale hand/wrist movements – this result follows from the previous observation that students in the principle-based reasoning condition spent more time in FLOW which was characterized by lower wrist/hand movement. In this sense, IMPLEMENT may not be enacted in the same way across conditions.

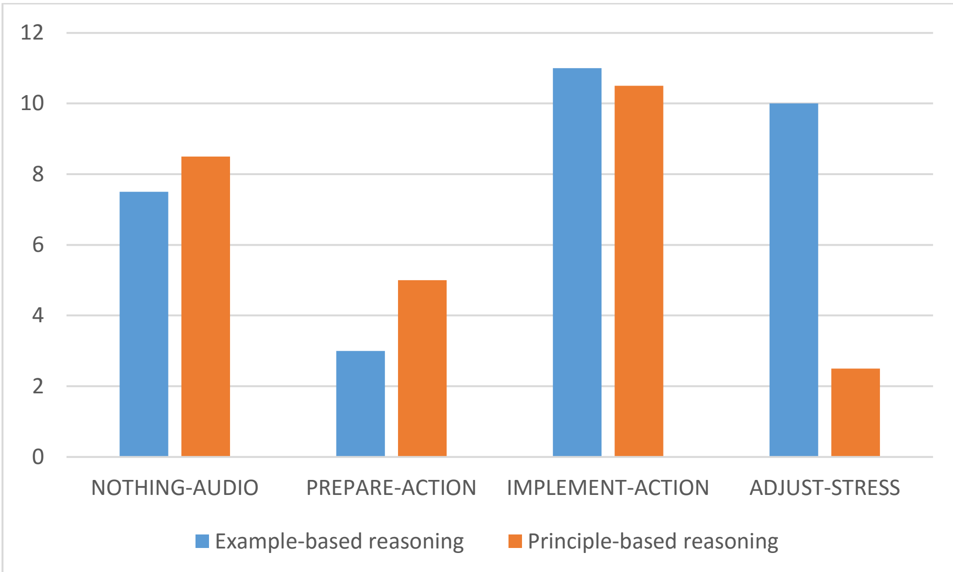


Figure 63. Median common behavior usage by condition for Part 4

Considering success, there are statistically significant results for usage of ADJUST-STRESS (Figure 64). Successful students are less likely ($p = 0.007$) to use ADJUST-STRESS than unsuccessful students. This finding differs from that observed in Part 2 for the ADJUST behavior. In Part 2, ADJUST was not associated with any significant differences. In this case it appears as though adding the additional context of multimodal behavioral data creates a more complete picture of what students are experiencing. For success, there were no statistically significant differences along the other three common behavior types.

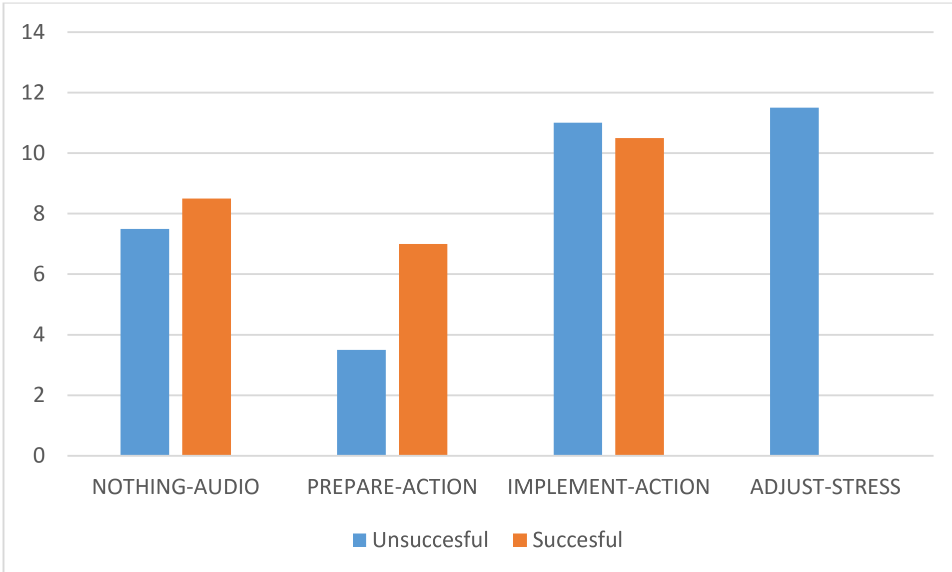


Figure 64. Median common behavior usage by success for Part 4

Finally, like success, learning also negatively correlates with ADJUST-STRESS (Figure 65). Students that received positive post-test scores are less likely ($p = 0.005$) to use ADJUST-STRESS than students that received negative scores on the post-test.

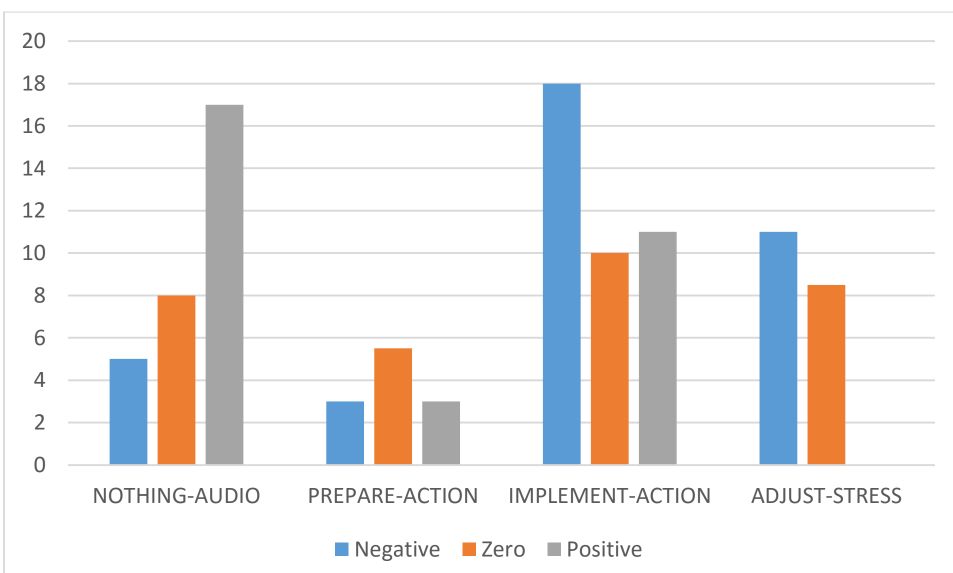


Figure 65. Median common behavior usage by learning score for Part 4

Fine-grain Cluster Usage Analysis

Consistent with the coarse grain analysis, the fine-grain analysis does not reveal any statistically significant differences between the two conditions. This is somewhat expected given that Part 2 did not garner differences in fine-grain cluster usage, nor did the coarse-grain analysis. Furthermore, the common behaviors from the combined data did not contain an analog to FLOW which characterized principle-based reasoning in Part 3.

Comparing fine-grain cluster usage between successful and unsuccessful students expands the set of dimensions on which students differed. Specifically, successful students were more likely ($p = 0.007$) to use IMPLEMENT-ACTION and less likely ($p = 0.007$) to use ADJUST-STRESS during the middle and final thirds, respectively.

Finally, students with positive post-test scores were less likely ($p = 0.007$) to use ADJUST-STRESS in the final third of the activity than students with negative post-test scores.

Discussion

This analysis was intended to combine the benefits of the hand-coded analysis, in terms of cluster usage, with the increased precision, and process-oriented distinctions associated with the multimodal sensor data. However this approach failed the harness the benefits of each of the previous analyses, and, instead, provided useful insights on a different dimension. That this analysis did not produce the desired outcome was evident from the results of the process similarity analysis. These results failed to align with success, learning or experimental condition. Despite these shortcomings, this combined analysis did reveal a methodology for predicting student success and learning. In particular, coarse- and fine-grained cluster usage analyses found that ADJUST-STRESS was negatively correlated with success and learning. While one would theoretically anticipate that students who have not succeeded or learned, would experience stress as their time begins to expire, and that they would resort to trying extensive adjusting in order to cope with the pending failure of their structure; having a computational tool to detect this has practical utility for supporting constructionist learning environments.

Nonetheless, the real benefit of this analysis comes, in part, through the ability to more closely understand the nuances of multimodal behaviors. For example, one such finding is drawn from taking the results of Part 4 in relation to those of Part 2. In Part 2, IMPLEMENT was associated with principle-based reasoning, positive learning gains, and success. However, when IMPLEMENT was put in the context of hand/wrist movement (i.e. IMPLEMENT-ACTION), the differences identified in Part 2 vanished. It therefore followed that the gestural enactment of IMPLEMENT may have differed by experimental condition, success and post-test score. One interpretation of this difference is that the two experimental conditions differentially impacted the rates of epistemic and pragmatic actions (Kirsh & Maglio, 1994). Epistemic actions are described as modifications to a system or an environment that help uncover information that may be hard to compute mentally. These epistemic actions are in contrast to pragmatic actions which are necessarily focused on physically moving the participant closer to their goal state. The combined multimodal analysis introduced the possibility that students are using building actions in different ways.

In addition to offering important insights into the nuances of multimodal behaviors, the fact that this analysis did not combine the affordances of the two previous analyses is informative. More specifically, this unexpected result highlights important aspects and considerations about the nature and complexities of conducting multimodal learning analytics research that would be sorely overlooked had I excluded the combined analysis. These considerations will be discussed in greater detail during Chapter 4.

Summary

This chapter began with an appeal to think more broadly about how to describe when a given learning strategy is effective. As a starting point I expanded on the results from the previous chapter, which focused on structural success, to also include learning and process as important considerations. For both learning and process there are clear indications that principle-based reasoning is more effective than example-based reasoning. Thus, having seen the benefit of principle-based reasoning across metrics, I raised questions about what is mediating these differences. Particularly I was interested in determining the practices associated with principle-based reasoning, and suggested that leveraging multimodal data could provide a means for conducting such an analysis.

Based on this assumption I proposed a general algorithm that allowed me to (1) identify common multimodal behaviors, (2) conduct pairwise process comparisons that maintained the temporal elements of the data and (3) get a glimpse of the different behaviors used by different groups (in terms of experimental condition, success and learning). Using this general algorithm I conducted three analyses. The first analysis utilized human generated time-stamps of individual actions based on a coding scheme from prior work (Worsley & Blikstein 2013, in press). From this analysis, I learned that students that spent more of their "test segments" implementing their ideas were more successful, learned more, and were more likely to be from the principle-based reasoning condition. In this light, one could argue that this supports the current "maker" practice of encouraging students to tinker at the expense of thinking. But I would suggest that perhaps the direction of causality is more in line with the idea that the principle-based reasoning condition *enabled* students to spend more time in implementation and less time in adjusting, and that this mediated student success and student learning. Put differently, without the initial focus on principles during the intervention, students would not have been able to engage in sustained implementation. However, this level of causality cannot be determined based on the current analysis. Nonetheless, even without being able to establish causality, this analysis revealed a student practice that bore significance across all three metrics of interest. At the very least, then, this form of analysis can be used to for producing predictions. One shortcoming, though, was that the pair-wise process comparison was more closely aligned to success than to my experimental conditions.

Given that a primary focus of this analysis was to uncover the practices that distinguish principle-based reasoning from example-based reasoning, I conducted a second analysis based entirely on data from various multimodal sensors. These sensors included audio, hand/wrist movement and electro-dermal activation. This analysis produced much higher prediction accuracy for distinguishing between the two experimental conditions based both on the cluster frequency usage and when conducting the pair-wise process similarity comparisons. Specifically, students who made more extensive use of a behavior pattern that I called FLOW, were more likely to be from the principle-based reasoning condition. FLOW was also associated with better performance on the post-test. However, this analysis was unable to identify distinctions among students whose structures were of different levels of stability. This lack of correlation with success is significant, because it means that students in the principle-based reasoning condition remain in FLOW even though they may not have been successful. Hence, one cannot make the argument that deviation from FLOW was merely the result of students experiencing challenges with their structures. Nor can one make the argument that it was only more knowledgeable students that were likely to remain in FLOW, as the correlation between learning and FLOW only emerged during the final third of the activity. In summary, then, the two sets of data seemed to offer complementary benefits. As such, Part 4 of this chapter combined the data sets. This combination produced clusters that were far less predictive than the two previous analyses. The analysis did, however, provide common associations that exist between the human-coded data and the multimodal behaviors, and in this way served as additional validation for the human coding. Additionally the analysis uncovered important nuances of multimodal behaviors. Furthermore, the shortcomings of the process similarity results from the final analysis motivate my later discussion of overarching concerns and considerations that one must account for when conducting multimodal learning analytics research.

Looking across analyses, there are clear instances where each provided some novel insights. In this sense, the overall algorithm appears to have relevance for studying learning, success and experimental condition; but honing in on these correlations requires different modes of analysis.

As a whole this chapter has shown that success, learning and process are not equivalent, though they may occasionally overlap. Thus, when thinking about measuring the effectiveness of a given learning environment it is important to be clear about which metrics one hopes to optimize. At the same time, this chapter has provided additional evidence that experimental condition can have an impact on learning, success and process. Because of this, one has to be cognizant about how to develop learning and reasoning approaches that allow the environment to realize the desired outcomes.

Finally, the three analyses in Parts 2, 3, and 4 provided evidence that multimodal analysis can provide a means for studying effective practices. Furthermore, they contribute to the argument that conducting research in constructionist environments likely necessitates adopting non-traditional modes of assessment. Reverting to traditional, uni-modal, outcome-based assessments will belie the goals of constructionist learning. For example, the results from Part 3 made this clear by showing that students spent most of their time in FLOW, a behavior that was characterized by a combination of multimodal sensor behaviors. As a result, there is a need to embark on data analysis techniques that go beyond the current strategies used for studying and assessing “making” and, espouse approaches that provide a broader perspective on learning, and that take a much more multimodal perspective. At the same time however, Part 4 demonstrated that simply concatenating different forms of data does not guarantee a successful analysis. This should not be taken to suggest that multimodal learning analytics does not have utility for advancing the field. On the contrary, all three computational analyses make it evident that multimodal analysis has considerable merit. One simply needs to use considerable care in ensuring alignment among methodology, data fusion and hypotheses.

Chapter 4. Discussion, Implications and Limitations

This dissertation began by describing the current excitement surrounding "making" and presented a pair of concerns about the Maker Movement's apparent disconnect from the founders of constructionism and prior research in the learning sciences. One example of this disconnect is the shallow, or non-existent, level of assessment used within the "making" context. This point was not raised to suggest that "making" needs to use traditional tests and quizzes. Instead, the concern is more closely based on the observation that the current assessment strategies in "making" pay very little attention to measuring or identifying student learning and dichotomize play and abstract thinking. With that back drop, I proceeded to make two suggestions that could be used to change the practices and culture of "making" for the better. The first involved engaging students in more useful post-activity reflection by having them indicate where their ideas came from. The second involved taking a broad perspective on measuring and identifying student development. Taking a broad perspective entails paying attention to more than just success and student learning, but also includes considering the process that students employ. Additionally, taking a broad perspective means looking across the various modalities that students use while engaging in "making."

Justification for the suggestions was based on a set of three studies and was presented across chapters 2 and 3. In Chapter 2, I used interviews to identify four strategies that help students complete open-ended engineering design tasks. These four strategies included: unexplained spontaneous insight, materials-based reasoning, example-based reasoning and principle-based reasoning. Analyzing the characteristics of each strategy, provided an initial indication that the different strategies were associated with different levels of efficacy, and/or different levels of expertise. Drawing from prior literature provided confirmation that the four strategies exist in a relative hierarchy in complexity, and also confirmed that the strategies have clear connections to previous research from other domains. Finally, Chapter 2 concluded with a pair of studies that experimentally tested the hypothesis that principle-based reasoning is more efficacious than example-based reasoning. Through these studies it was made apparent that principle-based reasoning was associated with higher quality designs than example-based reasoning. This result was consistent across multiple studies and across multiple measures of design quality. Given that these strategies are associated with differential success and appear to match the expected hierarchy proposed in the previous parts of Chapter 2, I made the suggestion that tracking student strategies is a useful, yet easy, way for characterizing student development within constructionist learning environments. As an offshoot of these findings, I also suggested that Makerspaces adopt idea generation strategies that go beyond simple open-ended brainstorms, but instead push students to more deeply draw upon pertinent prior knowledge and intuitions (Hammer, 2000).

Chapter 3 builds on the findings of Chapter 2 by expanding Study 3 to also consider *learning* and *process*. The appeal to learning was based on a concern that "making" does not necessarily result in understanding, or learning. Accordingly I conducted an analysis of student learning between the example-based and principle-based reasoning conditions, and showed that students in the principle-based condition learned more from the building activity than their peers in the example-based reasoning condition. Chapter 3 also included four separate analyses of student process. The first was based on the frequency of principle-item usage in intermediate designs. This analysis indicated that students in the principle-based reasoning condition were significantly more likely to add principle-based items in between successive tests, than students in the example-based reasoning condition. Following this analysis, I used techniques from Multimodal Learning Analytics to correlate student processes with success, experimental condition and learning. For each analysis I examined (1) a process similarity measure that maintained the temporal elements of each student's sequence of "test segments;" (2) relative usage of common behaviors on aggregate; and (3) relative usage of common behaviors at a fine-grain level.

The first of these analyses involved hand-coded behavior data and revealed the importance of students spending an ample number of "test segments" actually building (IMPLEMENT) their project. IMPLEMENT was seen to have relevance for experimental condition, success and learning, with more IMPLEMENT "test segments" associated with better success and learning. However the analysis was only weakly successful at predicting experimental condition using the process similarity metric.

The second analysis utilized audio, hand/wrist and electro-dermal activation data. This analysis very accurately grouped students into their experimental conditions using only their process similarity. The coarse- and fine-grain analyses suggested that the process differences were based on differential usage of FLOW, with students in the principle-based reasoning condition, more likely to use FLOW than their peers in the example-based reasoning condition.

Finally, I combined the hand-coded data with the multimodal data and found that the combined model was quite ineffective at grouping students based on experimental conditions. Nonetheless, combining analyses helped pinpoint the multimodal behaviors associated with different Object Manipulation Classes (i.e. PLAN, REALIZE, EVALUATE, MODIFY, REVERT, NOTHING) and highlighted a behavior that was highly predictive of student failure on the activity and negative post-test scores, ADJUST-STRESS. It also showed how the combined analysis provided a means for discovering nuances in the ways that students utilize IMPLEMENT. In summary then, across all of the analyses it was consistently the case that the process data provided meaningful insights into the study of student learning in constructionist learning environments. The next several sections describe possible interpretations and limitations of these findings, if we accept and generalize these results as presented.

The Practices of Making

One of the key findings and contributions from this dissertation is a set of overarching strategies – i.e. unexplained spontaneous insight, materials-based reasoning, example-based reasoning and principle-based reasoning - that students use for approaching engineering design tasks. These four strategies provide insights into the ways that students generate ideas to solve complex problems. Elements of each strategy bear resemblance to prior research on expertise and complex thinking (Chi et al., 1981; Cross & Cross, 1998; Epstein, 1999; Gentner & Holyoak, 1997; Maier, 1931; Moss et al., 2006; Nokes et al., 2010), but have seldom been used in the context of engineering design and hands-on learning. As I previously suggested, using these strategies as a way track and characterize student learning in constructionist environments has the potential to be tractable and of significant value given the differential outcomes and practices associated with each strategy. In chapter 2, I presented several reasons for why the strategies could have differential outcomes. Among these was the notion that principle-based reasoning pushed students to consider deep features, as opposed to surface features, which was one of the practices of experts. Here, however, I will suggest that another possible interpretation is that the reasoning strategies are associated with different epistemological frames (Russ et al., 2012; Scherr & Hammer, 2009).

Epistemological Framing

Epistemological framing is concerned with the role that a student's perception of the nature of knowledge and the task, has on how they approach the task. Scherr and Hammer (2009) discuss four epistemic frames that are typified by a combination of speech, gaze, posture and engagement (Figure 66). While their frames do not have direct applicability to the current analysis, the notion that multimodal indicators can be used to determine a student's current frame is of relevance to my work.

TABLE 1
Specific Behaviors Found in Each Cluster and Expectations Likely to be Associated with Those Behaviors

Blue: Worksheet Frame		*Green: Discussion Frame*	
Behavior	*Expectation*	*Behavior*	*Expectation*
Hands quiet, face neutral	Minimal interaction, individual activity	Prolific gesturing	Peers are watching and want to understand
Body leans forward, eyes on paper	Attention belongs on the worksheet	Animated tone, face	Intellectual and/or emotional engagement
Brief glances at peers	"Check-ins" expected	Sit up straight, eye contact	Attention belongs on peers
Muttering	Peers not attending to details of speech	Clear utterances	Peer interest in details of speech
Red: TA Frame		*Yellow: Joking Frame*	
Behavior	*Expectation*	*Behavior*	*Expectation*
Sit up straight; eye contact with TA	Attention belongs on TA	Giggle, smile, self-touch, fidget, unsettled gaze	Embarrassment, perceived vulnerability
Reduced gestures	Rehashing thinking		

Figure 66. Epistemic Frames from Scherr and Hammer (2009)

Russ, Lee and Sherin (2012) build on Scherr and Hammer (2009) by identifying three specific epistemological frames that emerged during cognitive clinical interviews. These three frames include the expert frame, the oral examination frame and the inquiry frame. In the inquiry frame, students perceive the task as an opportunity to construct new knowledge. They spend considerable time carefully thinking about how to respond to a given question or task. In the oral examination frame, students perceive the activity as a test that requires them to immediately and articulately produce the correct answer. Finally, in the expert frame, students perceive themselves as being highly knowledgeable, and in the role of sharing their knowledge with someone else.

While these frames, again, are not a direct mapping onto the four reasoning strategies, each form of reasoning can be seen as primarily adopting a different epistemological frame. More specifically, even though each of the reasoning strategies may give rise to any combination of the epistemological frames from Russ, Lee & Sherin (2012), each reasoning strategy appears to be associated with a single dominant epistemological frame. For example, materials-based reasoning can be interpreted as most closely aligning to the oral examination frame. Recall that in materials-based reasoning the student is looking for the material to cue them into what to do. Moreover, students are often times looking for the “trick.” Previously I described a student who significantly struggled with the activity until realizing that he could use the roll of tape as a primary support. This student serves as is a prime case of an individual that perceives the task as requiring him to recognize the “trick.” The following exchange reiterates this by highlighting that the materials provided did not “fit” the task.

> Interviewer: You mentioned earlier that the materials don’t fit. What would have been better? What would have been ideal materials to build with?
>
> Student: I mean that what I meant by that was like, I wanted something to have a fatter, what is it? Like a beginning, more like…

The idea that the materials “don’t fit the task” suggested that this student was looking for the right answer. Hence, even though he may have also employed other epistemological frames, there is a distinctive similarity between how this student approached the task, and the oral examination frame.

Moving up the reasoning strategy hierarchy, example-based reasoning may be most similar to the inquiry frame. In example-based reasoning, students engage in a search to identify salient examples from their prior experiences that can be useful in tackling the current challenge. They recognize that they don't immediately know how to complete the task, but also realize that there are several ways to go about completing the activity. Hence, their perception is that they simply need to build from their prior knowledge in order to succeed, which is characteristic of the inquiry frame. The analyses from Chapter 3 provide some additional justification for drawing this parallel. For example, in Chapter 3 Part 2 principle-based reasoning was associated with students spending more "test segments" in IMPLEMENT, than students in the example-based reasoning condition. One could argue that the example-based reasoning students were adopting more of an inquiry approach and had to use more of the "test segments" exploring different options, and that this was evidenced in their less frequent usage of the IMPLEMENT "test segment."

Finally, principle-based reasoning appears to most closely parallel the expert frame, though I again emphasize that this is not necessarily a direct mapping. Within principle-based reasoning, students perceive that they have the knowledge and prior experience to successfully complete the activity, and assume that they can piece their way through the task. Hence, it's not about finding the right answer, or that they don't have sufficient prior knowledge. It's more so a matter of confidently applying principles until their structure is successful. As such, principle-based reasoning probably includes some elements of the inquiry frame also. The analysis from Chapter 3 Part 3 offers quantitative support for the association between principle-based reasoning and the expert frame. Specifically, students in the principle-based reasoning condition were more likely to remain in FLOW than their peers in the example-based reasoning condition. One way to interpret this is as students in principle-based reasoning condition being able to stay in a single behavior (FLOW), and not having to deviate much to behaviors typified by significant action, audio or stress. However, this difference was not at all related to student success, meaning that one cannot make the argument that students deviated from FLOW solely because their structures were not working properly. On the contrary, even though students in the principle-based reasoning condition were not universally successful in completing the task, the principle-based reasoning condition students remained in FLOW during the majority of the activity. Put differently, even in the face of challenges, students in the principle-based reasoning condition remained in FLOW. On the other hand, students in the example-based reasoning condition, transitioned to other states throughout the activity. One could interpret this as the example-based reasoning condition students having to engage in real-time inquiry as they encountered difficulties.

That said, the interpretation that the two reasoning strategies were solely a function of epistemological framing comes into question when considering the results obtained by studying students' intermediate structures. Recall that students in the principle-based reasoning condition were more likely, than individuals in the example-based reasoning condition, to add principle-based elements to their structures between subsequent tests. Based on this result, it seems more plausible that the observed differences are a function of both framing and the recognition and application of deeper design features. However, the current study cannot clearly differentiate between epistemological framing and improved recognition and application of STEM principles. Instead an additional study would need to be conducted that controlled for framing and inclusion of principles separately.

Epistemic Forms and Epistemic Games

Collins & Ferguson's (1993) epistemic forms and epistemic games provide another interpretation of the differential results of example-based reasoning and principle-based reasoning. In providing a definition for each term, Collins & Ferguson wrote, "epistemic forms are target structures that guide inquiry. Epistemic games are general purpose strategies for analyzing phenomena in order to fill out a particular epistemic form." Epistemic games range in complexity and are commonly used in everyday activities. For example, among the most basic epistemic games is a "list game." Collins & Ferguson refer to list games as the simplest structural analysis game, but still of significant utility in the early stages of the inquiry process as they provide value in allowing people to decompose an entity into its subcomponents. In addition to structural analysis games, Collins & Ferguson describe a set of functional analysis games. In these games the "goal is to determine the causal or functional structures that relate elements in a system" (Collins & Ferguson, 1993). From this perspective, the comparison of example-based reasoning and principle-based reasoning can be interpreted as comparing the efficacy and impact of two different epistemic games on hands-on engineering design tasks. Example-based reasoning could be interpreted as an instance of a *structural analysis epistemic game*, namely a form of list making, whereas principle-based reasoning could be viewed as a *functional analysis epistemic game*, specifically a form-and-function game. As previously suggested, the structural analysis games often represent an important place to start an inquiry, while the functional analysis epistemic games provides a deeper level of inquiry, and may be an appropriate next step after completing a structural analysis epistemic game. Importantly, the two different types of epistemic games, result in different epistemic forms. Accordingly, the results from the various analyses may imply that the epistemic forms derived from a form-and-function game, provide a relatively more effective form for students to use when completing this type of activity, than the epistemic form associated with a list game. What is uncertain, however, is the nature of interaction among epistemic forms and games, and epistemological framing. These studies have started to shed some light on this interaction by suggesting that functional analysis epistemic games are more closely associated with adopting an expert frame, whereas structural analysis games foster an inquiry epistemological frame. However, to further disentangle the relationship among these ideas it would be prudent to develop a study that has this interaction as the main point of comparison.

The Nuances of Common Behaviors

In addition to identifying common behaviors at an overarching level, in terms of the reasoning strategies, I also provided evidence for common multimodal behaviors on shorter time scales. Specifically, all three of the multimodal analyses in Chapter 3 (Parts 2, 3 and 4) showed that even though students are all unique in their prior experiences and knowledge, there are a number of shared practices in terms of the combinations of actions that co-occur. From the hand-coded data, these common behaviors included PREPARE, IMPLEMENT, ADJUST and SIGNIFICANTLY ADJUST. Based on observation of student actions, and from prior work, the PREPARE, IMPLEMENT and ADJUST, all seemed intuitive, as each of these involved actions that one would expect to see together. For example, when students are in PREPARE, it makes sense for them to demonstrate their planning through perceptible (PLAN) and imperceptible (NOTHING) actions. In the same way, when working towards constructing one's structure, one would expect to see large combinations of building (REALIZE) and undoing (REVERT). However, the presence of two different forms of modifying, namely ADJUST and SIGNIFICANTLY ADJUST, is meaningful because it emphasizes that a given instance of an action is not necessarily the same across contexts. A "test segment" that is characterized by MODIFY, and nothing else, is functionally different than MODIFY in the presence of other actions. Specifically, there were instances where a student would adjust a small piece in order to prepare for the next step in the building process, versus instances where a student would spend the entirety of the "test segment" adjusting their design. The use of MODIFY, by itself, then, signals a different objective than MODIFY in conjunction with REALIZE, for example. Thus, from a methodological perspective it is encouraging that tools from learning analytics can help tease apart these differences. Furthermore, it makes salient the importance of capturing the context of a given action. Had I merely looked at the raw actions (PLAN, REALIZE, REVERT, MODIFY, EVALUATE, NOTHING) in isolation of one another I would have overlooked this important difference in how MODIFY is used.

The common behaviors from Chapter 3 Part 3 also highlighted the importance of context. The most frequently occurring “test segment,” FLOW, involved below or near average values for audio, hand/wrist movement and electro-dermal activation. Had I only looked at one modality, or examined each modality in isolation I would have also overlooked the importance of FLOW.

Process Matters

An additional implication for the practice of “making” is that the *process matters*. This certainly comes as no surprise to proponents of “making,” but what may come as a surprise is the *extent* to which process matters, and how easily process can be influenced. The interventions that I developed lasted 3 minutes or less. Despite being relatively short in duration, the effects on student process, success and learning were quite pronounced. Identical final products can be constructing using drastically different strategies. From this observation comes the recommendation to pay close attention to process.

As an additional point related to judging students based on success, I want to highlight that it is important to recognize positive changes in students processes, event amidst projects that may seem to be unsuccessful. Among the findings from Chapter 3, Part 3 was that students in the principle-based reasoning condition used FLOW more than their peers in the example-based reasoning condition, even after controlling for success. Those students who were unsuccessful but still utilized a valuable practice should be recognized, despite their final product not meeting the target objective.

A further consideration related to process is the importance of researchers and practitioners, who have the available resources, to make a larger investment in capturing and more thoroughly analyzing student learning processes. Many Makerspaces are using journals, blogs and videos as platforms to store process data. However, most of these materials are just for documentation and rarely get analyzed by researchers. Alternatively the artifacts are gathered in order to recruit new students and new sponsors. Documentation and recruiting certainly have value in the short-term, but as I've shown, studying student artifacts can reveal the emergence of student knowledge development. In the case of this research, examining student intermediate structures provided clear evidence that students were actively applying engineering principles.

At the same time, the nature of studying student reasoning strategies should take a more central role in the enactment of "making" environments. Practitioners and researchers should be aware of how students are framing an activity before it begins, and should also consider ways to ensure that the framing is appropriate for the goals of a given learning experience. In fact, I would push this argument further, to suggest that Makerspaces should be more adamant in changing the culture and practices in such a way that students are more likely to adopt an expert, or principle-based, frame, instead of the prevalent practice of relying on materials- or example-based reasoning strategies where students are encouraged to build from ready-made solutions, or utilize shallow ideation practices.

Limitations

With regard to the findings and interpretations garnered from the enclosed analyses, one concern is that students were limited by time. While time constraints are a reality in most learning environments, by placing this constraint on students, I have, arguably, privileged efficiency, and, in so doing, privileged success and principle-based reasoning. Furthermore, if given more time, students in the example-based reasoning condition would probably have been more successful on the task. Even if this were the case, this would not change the findings that the processes that students followed were different. All too often the "maker" culture makes the functionality of the project, or success, the primary goal. However, the process that students follow to be successful, or unsuccessful, is of importance. If students are using strategies that are not engaging their cognitive faculties, but are instead depending an exhaustive trial and error or step-by-step instructions, their long-term development is being stunted. Additionally, based on the coarse- and fine- grain analyses, there were clear instances where the factors that correlated with experimental condition and learning had no bearing on student success. Hence, being overly concerned about time limitations invalidating these findings is of little relevance to the overall argument that process matters.

Another limitation relates to the common behaviors. For this analysis I used the X-means algorithm to find commonalities among the student "test segments." X-means was used because it (1) automatically determines the number of clusters, (2) provides discrete assignment for each data point; and (3) has been shown to work well in previous analyses (e.g. Blikstein et al., in press; Worsley & Blikstein, in press, 2013). However, there remain several other approaches for grouping similar student actions or identifying the most salient factors within a dataset. For example, singular value decomposition and principle component analysis are other approaches that can be used to identify common behaviors. Nonetheless it has been shown that K-means clustering, which is used in X-means clustering, and principle component analysis provide similar outcomes, with the main difference being that K-means clustering provides a discrete assignment of each data point to a given cluster whereas principle component analysis involves the probability of each data point being assigned to a given cluster (Ding & He, 2004). Instead the limitation with the common behavior analysis has more to do with the fact that the clusters that I identified are probably not universal. Hence, a comparative analysis in a different context and with different students would likely generate a different set of common behaviors. This is especially true given that X-means clustering groups the data based on the characteristics of the data. This is both a benefit and a drawback. If the "test segments" are all very similar, than the algorithm will effectively help identify the dimensions over which the data is most different. However, if the population of "test segments" is extremely heterogeneous, the resultant clusters may contain groups of "test segments" that are quite dissimilar to one another. Hence one must carefully and critically examine the results derived from clustering and computational analysis. Nonetheless, I would still expect for the general approach of clustering, or dimensionality reduction to yield a set of common behaviors that would provide valuable insights about learning in that particular context.

Reflections on Cognition

Analogical Problem Solving and Encoding Strength

From a cognitive perspective, the current studies can be seen as adding to the analogical problem solving discussion on encoding. Within this body of literature researchers carefully consider the elements that facilitate the ease and depth of example retrieval, and how well the details of the example are applied to a new problem (Carbonell, 1982; Francis, Fernandez, & Bjork, 2010; Gentner et al., 2003; Gentner, 2004; Kurtz & Loewenstein, 2007). Specifically, the two experimental conditions that I studied were designed to represent different levels of example encoding. The example-based reasoning condition involved students identifying structures from their home, community or school that could be used to design their structure. The structures that they generated would therefore be considered deeply encoded examples that the students would have had extensive familiarity with. On the other hand, the principle-based reasoning condition involved shallowly encoded examples, as the students were deriving principles from three two-dimensional images. Furthermore, two of the three objects presented were items that students in the sample population would never have handled or gained extensive familiarity with (i.e. the igloo and the bridge).

Building from this work on encoding, my results mirror prior findings, in that an example need not be deeply encoded in order to successfully be used as an analog

(Kurtz & Loewenstein, 2007). Instead, the mere task of looking for deep features from within an analog appears to be of greater benefit than surface-level inspiration from deeply encoded structures.

Object Closeness

The discussion of encoding strength also has relevance in reconciling the findings of this paper with the ideas expressed in the seminal paper “Epistemological Pluralism and Revaluation of the Concrete” (Turkle & Papert, 1992). In Chapter 2, I wrote that concrete is the opposite of abstract, only in so much as concrete ideas are based on entities that the student can relate to. Hence, from Turkle and Papert’s perspective, example-based reasoning is as effective for learning and success as principle-based reasoning, because example-based reasoning affords “object closeness.” As such, one could interpret my results as being “anti-constructionist,” but I would like to suggest just the opposite. Principle-based reasoning is not concerned with students developing a more abstract view of the world. Instead it is pushing students to get “closer” to the objects being examined, by having them more deeply consider how that object moves and behaves. The difference between the example-based reasoning condition and principle-based reasoning condition is that in example-based reasoning the task involves generating objects that one can presumably become “closer” to, but without the requirement that the students deeply consider the nature of the objects; whereas in the principle-based reasoning condition, students are tasked with becoming “close” and deeply examining three examples that may have been, heretofore, foreign to them. Despite potentially being unfamiliar objects, because the activity asks for three things about the structures that confer strength, students have no choice but to try to embody the objects they are examining. In this way, then, principle-based reasoning enabled students to engage in a level of "closeness to the object" that they probably would have not done without being charged with finding three mechanisms or principles.

Thus, "closeness to the object" provides a way to recast the hierarchical relationship of the reasoning strategies. Previously I appealed to a traditional argument that centered on the increasing complexity and applicability that is achieved as one moves from unexplained spontaneous insight, to materials-based reasoning, to example-based reasoning, and finally to principle-based reasoning. Unexplained spontaneous insight involves little to no embodiment or identification with the object, or objects, in use. As one embarks on materials-based reasoning and example-based reasoning, the participant is beginning to become more aware of the object, and is moving closer to understanding and relating to it. Finally, principle-based reasoning involves the students developing a deeper intimacy with the object where they must embrace the most salient and central features of the object. Hence, suggesting that the reasoning strategies represent a hierarchy need not depend on conventional arguments of scientific or mechanistic reasoning.

While on the topic of the relationship of the reasoning strategies, one could also view each reasoning strategy as representing different levels of the students' mastery over the materials and the design process. In materials-based reasoning the student is at the mercy of the materials, and is principally looking at how the materials dictate the design. In example-based reasoning and principle-based reasoning, the student exercises greater control over the design, and is forcing the materials to fit the design, as opposed to having the design fit the materials. This difference was made earlier, but is one worth noting again since it relates to another argument in justifying that student reasoning strategies may be associated with different levels of efficacy.

Non-Experts Using Expert Strategies

These studies could also be used to suggest that non-experts can effectively make use of certain expert strategies. Despite the fact that expertise is typically defined as requiring both expert knowledge and expert organization of that knowledge (e.g. Nokes, Schunn, & Chi, 2010), the students in this study demonstrated that they could use ill-defined deep features in order to increase their rate of success and rate of learning. As a result of this, one could also conjecture that using principle-based reasoning is not what defines an expert, but is instead a practice that experts use and which supports their classification as experts. As such, even though it is unreasonable to expect non-experts to always behave like experts, there may be a number of practices associated with expertise that can be fruitful in helping novices learn and perform more effectively. Furthermore, I argue that some of these practices, since they might appear in fragmented or untraditional forms during students' work, are difficult to detect with traditional methods, and that my methodological contributions in this dissertation point to promising ways to conduct this type of research and find these "markers of expertise" (Worsley & Blikstein, 2011).

Bridging the Bands of Cognition

The final area of cognition for which this dissertation bears significance is in relation to Newell's "bands of cognition." Newell (1994) describes time scales across which human actions can be interpreted as biological, cognitive, rational and social. Specifically, the work suggests that from time scales of tenths of a second all the way up to months, each band relates to a different set of theories. Anderson (2002) builds on this framework by considering the extent to which human actions that occur within a very small time scale, and one of the lower bands (Biological Band or Cognitive Band) influence human actions on the larger time scales. Anderson proposes three theses: the Decomposition Thesis, the Relevance Thesis and the Modeling Thesis. The Decomposition Thesis claims that the events that occur at larger time scales, can be decomposed into actions on shorter time scales. The Relevance Thesis relates to the claim that the "microstructure of cognition is relevant for educational issues." In practical terms, this means that there are short time scale actions that are important for studying and diagnosing learning development. Finally the Modeling Thesis is concerned with the ability for cognitive modeling to help explain how to use the fine-grained information to improve instruction.

The analyses presented in this dissertation add to this discussion of the "bands of cognition" by examining three different time scales, and that all attempt to make a bridge between learning that is evidenced on the scale of hours, by analyzing data that is derived on the scales of milliseconds and seconds. For example, Chapter 3 Part 2 was based on second by second action codes that were subsequently used for modeling student behavior. Modeling this behavior occurred by using both aggregate measures and process based measures. From the aggregate measures, there were clear differences based on success, learning and experimental condition that emerged on the time scale of “test segments.” Most notably, as students began to deviate away from IMPLEMENT, they were less successful and learned less from the activity. They were also less likely to succeed and less likely to learn from the example-based reasoning condition. In this way, there was a clear instance of how this analysis was able to bridge actions at a relatively short time scale, with human actions, learning and performance at a larger time scale. In this same analysis when I modeled student learning in a way that preserved temporality, the result that emerged bore the greatest correlation with student success. This, again, suggests that the fine-grain information provided a meaningful way to bridge across various time scales.

The analysis in Chapter 3 Part 3 looked to push this further by examining the human actions that occurred on an even lower time scale than the previous analysis. Audio, hand/wrist and electro-dermal activation data were combined in order to describe user behaviors. Both electro-dermal activation and hand/wrist movement data were collected on the order of a tenth of a second. Hence, this data falls in the Biological Band (Newell, 1994). The analysis, again aggregated these data points to the “test segment” levels, and showed that there were statistically significant differences between the behaviors and practices of students from the two experimental conditions. This was most apparent from the analysis that involved pair-wise sequence similarity comparisons. However, this difference was also confirmed in the fine-grained common behavior usage analysis, with students from the principle-based reasoning condition being far more likely to stay in FLOW during all "test segments." This, again provides a very clear opportunity for diagnosis of student experimental condition, as well as learning. Accordingly, this particular analysis showed how moving to even more fine-grained data, as compared to that of the preceding analysis, continues to have demonstrated efficacy in bridging between human actions, outcomes and processes across various time scales that are important for learning.

Through the lens of "bands of cognition," the combined analysis from Chapter 3 Part 4, was an effort to combine data that was derived from two different time scales. The hand-coded data, which was taken at 1-second increments, but ultimately represented actions that were more on the order of 10s of seconds, and the multimodal sensor data, which was derived and recorded on the order of tenths of a second. This dissimilarity in the time scale used for the different types of data may be one of the reasons that this analysis yielded results that bear little significance for distinguishing between experimental conditions. The analysis did provide information that can be used for diagnosing student learning within this context, namely that students who spend considerable time in ADJUST-STRESS are likely to perform poorly, and that there are nuances in how students use IMPLEMENT. Nonetheless, recognition that combining data streams from different time scales may not garner the desired outcome is an important piece of information to take into consideration for future multimodal learning analytics research.

However, even this interpretation of the process similarity results from the combined multimodal analysis from Chapter 3 Part 4 is dissatisfying. Suggesting that analyses which merge data from different time scales are unlikely to succeed is problematic. In many respects the multimodal sensor data used for the analysis from Chapter 3 Part 3 was based on data from different time scales, or levels. Audio and hand/wrist movement represent seemingly conscious modalities. Even if one wanted to argue that human actions and speech are only conscious when examined in aggregate, there is a clear difference between these two data streams, and the lack of control an individual has over their electro-dermal activation. Students can certainly be trained to exercise some control over their physiological response to stress, but this level of control is drastically different than the level of control that students hold over speech and gesture. In this way the multimodal sensor analysis also involved multiple levels. So instead of suggesting that combining data from different orders of magnitude is problematic, it may be more appropriate to state that researchers must be increasingly cautious as they begin to merge data from various time scales. The analyses in this dissertation have made it clear that data synthesis from different time scales may be useful, but as one attempts to bridge across multiple time scales in the same analysis, this task may become more complicated. For example, by combining the human-coded actions with the multimodal sensor data, Part 4 was mostly providing a context for examining the interactions between these two forms of data. Specifically, it was enabling us to examine the multimodal behaviors associated with the different Object Manipulation Classes. While examining this interaction is an important area of analysis, it was not that initial impetus for the analysis. Hence, embarking on multimodal learning analytics research should not be considered as the mere accumulation and aggregation of data streams. Research that lacks a strong theoretical framework may easily result in missed or uninteresting findings.

Limitations

Reflecting on Newell's "bands of cognition" also brings to the forefront question about how deliberate or conscious students were of their actions. For example, the analysis of intermediate structures found that students in the principle-based condition were more likely to add principled-items between subsequent tests. However, it is unclear whether the students were actually conscious of these principles, or their overall strategy, or if their actions were subconscious. The results from the post-test, the primary metric used to measure learning, suggest that students were conscious enough to make selective use of principled-items and that this was not merely a function of success. However, to truly measure this requires a more precise instrument that forces students to make their reasoning more visible and more explicit.

There are also limitations in terms of the previous discussion of the strength of encoding. Despite following an experimental design similar to that of Kurtz & Loewenstein (2007) one critical difference was that I used pairs of students as opposed to individual students. This is important because even if one student was building from a deeply encoded example, the other partner was probably more familiar with the bridge, ladder and igloo, than she was with the example from their partner's life. For example, if we consider the student who identifies a specific chair from his home as the motivating example, and a partner who has no knowledge of that specific chair, the second person in the pair is at a loss, and is left imagining what the chair looks like. Because of this, it is likely that, at best, only one person in each pair was working from a familiar example, while the partner theorizes about the nature and structure of their partner's example structure. At the same time, one could argue that students in the principle-based reasoning condition were still working from deeply encoded analogs, but that these analogs were at a different level of organization. For example, students in the principle-based reasoning condition occasionally identified triangles as being important for conferring strength to the bridge and the ladder. Each individual may have envisioned a slightly different triangle, but the fact remains that each probably had strong familiarity with triangles. Hence, triangles could have been deeply encoded based on each student's prior experiences with them. As such, one could argue that this study involves shallow encoding and deep encoding at the example level, but does not preclude the practice of deep encoding at the principle level. Accordingly, a future study that more closely controlled for encoding strength would help solidify the findings from this study. One way to start to address this would be to interview students about the different mental representations that they used during the activity, and find out if both partners in a pair shared a strong representation of their example object, or the principles that they identified.

Methodological Implications

The Role of Qualitative Analysis in Learning Analytics

There are also important methodological implications and contributions to be gleaned from this research. First, it makes the case for semi-automated analyses. Qualitative analysis is often received with hesitation among the learning analytics and data mining communities. However, this research demonstrates how the qualitative data can provide a powerful lens for analyzing complex data. Specifically, the qualitative codes provided the most consistent results among the three multimodal analyses of Chapter 3. Additionally, the qualitative data was crucial to segmenting the multimodal data streams in a way that aligned with pedagogy. Namely, in the case of studying “making,” I have shown that using “test segments” provides an effective means for deconstructing a complex sequence of building actions into chunks that have significant meaning. As an offshoot of this, then, to completely forego the use of qualitative data may not be appropriate for analyzing these complex learning environments, as the current set of automated tools may not afford meaningful data segmentation strategies. Alternatively, one can view this work as providing justification and guidance for the development of new data extraction tools that attempt to automate the hand-coding performed in Part 2 of Chapter 3.

Leveraging Multimodal Data

Another implication concerns the utility and reality of multimodal analysis. The multimodal analyses in Chapter 3 Part 3 provided clear benefits, though the results differed from those from Chapter 3 Part 2. Using multimodal sensor data was far more effective at distinguishing between experimental conditions than the analysis of qualitative data. This, in conjunction with the importance of capturing and documenting the context of various actions, suggests that multimodal sensor data can play an important role in furthering the field's understanding of effective learning strategies in complex learning environments by more holistically capturing common behaviors and the nuances of multimodal expressions of knowledge.

Algorithm Generalizability

An additional implication is that the general algorithm presented in Chapter 3 appears to have benefit across data streams. Beginning by identifying common behaviors, and then recasting and comparing each student's sequence of data points based on those common behaviors has demonstrated payoff across several analyses (Worsley & Blikstein, 2013, in press). This suggests that the algorithm, itself, may have utility beyond the present studies. Solidifying the general utility of the algorithm would require applying it to other content areas, and in other contexts. For example, the domain of computer programming might be prime a candidate for this approach, as there are discrete moments in which students test and compile their code. In fact, prior work in this area has shown some initial promise in this regard (Blikstein et al., in press).

Limitations

As before there are a handful of limitations that the reader should recognize in order to interpret these findings. One such limitation is the level of semantic understanding that can be inferred from the multimodal behavioral data. At the moment, the multimodal data offers little in considering the content of student utterances, or the nature of student gestures, for example. In future work, I plan to examine the ways that semantic information can be extracted and interpreted from the multimodal sensor data.

An additional limitation has to do with the combination of individual and dyadic data. In most of the analyses, excluding the comparison of intermediate structures (Chapter 3 Part 1) I treated all data as individual data. For the hand-coded actions, hand/wrists movement, electro-dermal activation, design documents and pre- and post-tests; this seems appropriate since there is a different set of data points being generated for each student. In reality, even this data is subject to dyadic bias with the actions of one student impacting the actions of their partner, which is why I control for this when comparing pair-wise sequences. Audio data, on the other hand is treated identically for every pair of participants. Ideally, techniques can be used for looking at each pair of students as a unit, as opposed to looking at each student individually. However, this becomes complicated when the measure of learning is determined individually. Technologically, using individual audio streams for each participant would solve the problem of identical audio channels. However, pedagogically the potential for dyadic bias still remains as the two individuals continue to engage, or fail to engage one another. Accordingly, focusing on the individual and mixing data from individuals and dyads is one of the limitations of this research.

Practical Implications

Coding Schemes for Studying Making

From a practical perspective, this dissertation provides a pair of preliminary coding schemes that can be used for studying changes in students at various time scales. With the coding schemes I also provided examples that characterized the different reasoning strategies. These resources can be used by others to forge ongoing research in the study of constructionist learning. Along with the coding schemes, the specific task is an additional tool that others can use within their learning environments as a way to engage students in a hands-on building activity that may provide a context for studying student reasoning strategies. However, I want to reiterate that an expressed strategy or behavior on a single activity should not be the means for classifying the student as a novice, intermediate, expert, etc. For one, as I mentioned in Chapter 2, the use of unexplained-spontaneous insight, materials-based reasoning and example-based reasoning is not exclusively for non-experts. Experts also use these strategies, especially when encountering new situations. Instead, the idea of tracking reasoning strategies is as an alternative to traditional outcome based metrics that may overlook the rich process that students utilized. In this way looking at strategies is designed to validate, and not discredit, student development. One would hardly consider disregarding the ideas of an acclaimed expert because they were unable to think back to the origins of their idea.

Furthermore, even if we accept the notion that the students' behaviors and comments are indicative of their framing of the task, one has to be careful not to generalize that framing without examining how that student frames different tasks on other topics and in other contexts. The theory of “manifold cognitive resources” (Hammer, 2004c) provides an important perspective on this point by explaining that students have a variety of resources at their disposal, but often times only activate those resources in certain contexts. For example certain elements of a task, or an experience, are likely to trigger the activation of different resources and practices. Accordingly, the role of an instructor is to help students draw upon the student’s diverse cognitive faculties. As such, even in describing the students in this dissertation, I want to emphasize that there is a difference between a student using a principle-based reasoning strategy, and that student being classified as a principle-based “reasoner.” The coding schemes provided in this text are intended to categorize student utterances and behaviors, and not to categorize students as a certain type of person.

Finally, epistemological frames, reasoning strategies and multimodal behaviors are dynamic and can be enacted at different levels. As was observed in this study, students may begin with a given strategy and then transition to another strategy. Depending on the timescale over which students strategies are being modeled, one may come to different conclusions about that student’s development.

Using Reasoning Strategies to Inform Instruction

Beyond this, tracking reasoning strategies provides a different glimpse into how students are changing and gives the teacher, facilitator or parent a basis from which to engage the student. As an example of this, imagine a teacher who is preparing to talk with a student whose project did not satisfy the student’s goals. This conversation could be markedly different based on the knowledge of the reasoning strategy that the student employed. For example, a student who based their idea on an example structure, but overlooked an important piece of that structure would be likely to differentially benefit from a different kind of discussion than a student who based their design on a misunderstanding, or misconception about the properties of one of the provided materials. Being aware of this type of information provides a better context for interacting with students.

Multimodal Analysis for Targeted Sampling

Finally, while not explicitly discussed in this dissertation, the tools described for studying student behavior can conceivably be used for doing selective data search. Specifically, the semi-automated multimodal analyses can be used to direct the researcher or practitioner towards salient pieces of data that merit human inspection. For example, the researcher or practitioner could algorithmically examine the segments where students transition from one common behavior to another, in order to identify elements of the learning environment, dialogue, current project configuration, etc., that appear to be contributing to the change in behavior. Taking a step back, one could look across entire class reasoning strategy trajectories and identify elements of a given project, that are associated with a given strategy, or that seem to help students transition from one strategy to another. In this regard, leveraging multimodal learning analytics need not be limited to automated studies of “makers” but can easily be extended to a broad set of learning environments and a variety of research practices.

Limitations

Practically speaking, one limitation is the fidelity of student generated responses. I have argued for practitioners to solicit explanations from students around the origins of their ideas, but this seems to assume that students will produce authentic responses to teachers’ questions. In the absence of authentic responses teachers must rely on process data and the analysis of student artifacts which becomes increasingly complicated and time consuming.

A further limitation is knowing the appropriate time span over which one should expect to see changes in the patterns of student development, as well as a useful sampling frequency for querying students about their reasoning strategies. At first glance, Kafai (1995) provides some insights on the appropriate time scale, in that she monitored students for several months. However, the examples mentioned in Chapter 2 already demonstrated that even over short periods of time one can document meaningful transitions in student reasoning strategies. In conjunction this would suggest that while this is a limitation there are likely to be meaningful insights that can be drawn from nearly any time scale and sampling frequency.

Another limitation that I will mention here is the limits of what can be usefully derived from the coding schemes. For example, in the study where I applied the two coding schemes, I found that the majority of the students were using elements of principle-based reasoning. Thus, one limitation is that I don’t identify additional distinctions that can be drawn once students have transitioned to primarily using principle-based reasoning. This is another area that I would like to explore in future research, but would currently posit that a fruitful line of research could include an examination of how students are utilizing the reasoning strategies in non-academic contexts. So instead of expecting for increasingly refined forms of principle-based reasoning in the academic setting, it may be more important, and more meaningful, to study how the strategies are being used in significantly different contexts.

The fact that materials-based reasoning did not receive greater attention in my empirical studies is a limitation in the interpretation of the relative efficacy of the different reasoning strategies. In using example-based reasoning and principle-based reasoning, I touched on the reasoning strategies that I believed were most prevalent among the populations of students that I worked with and that appeared to have the greatest relevance to “making.” However, it may be that materials-based reasoning has more in common with principle-based reasoning than I am currently aware of. It’s also possible that materials-based reasoning provides a natural entry point for certain populations of students. Accordingly, in future work I intend to include materials-based reasoning as an experimental condition in my research studies.

The two final limitations that I discuss relate to the results being a representation of average behavior, and to my focusing on the building phase. In particular, my discussion of the two experimental conditions used throughout this dissertation focused on group-wide commonalities and trends. However, important observations that still warrant some discussion, also emerged among subsets of each experimental condition. For example, in Study 3, the two pairs of students from the principle-based reasoning condition that did not successfully build a stable structure came up with designs that were far less viable than the designs from the example-based reasoning experimental condition. As an example of this, one pair of students attempted to make a structure that would balance a paper plate on just two straw legs. Hence, one should not assume that principle-based reasoning will always be more effective than the other reasoning strategies, as this is certainly not the case, and has also been corroborated by other research (e.g. Werner, 1937). Furthermore, within the two principle-based reasoning pairs that failed on the task, there were several instances of dismissive and self-deprecating statements that occurred primarily during the intervention phase. In some cases this was instantiated as one partner discrediting or ignoring their partner's ideas. In other cases the student would dismiss their own ideas as being ill-informed. Accordingly, in future work I intend to more closely studying the different stages of the experiment and devote greater attention to studying the behaviors of sub-groups of students. Furthermore, there is an opportunity to more deeply engage the collaboration literature (Barron, 2003; Roschelle & Teasley, 1995; Schwartz, 1995) as part of understanding and comparing the affordances of different reasoning strategies when enacted among dyads of students.

Synopsis

In concluding this section on the implications and limitations of this work, I want to return to the two suggestions that I used as the central foci for Chapter 2 and Chapter 3. In Chapter 2 I suggested that students' development be tracked by periodically querying them about the origins of their designs and ideas. I presented a pair of coding schemes based on insights drawn from a study of a diverse population of students; as well as theoretical and empirical examples of why reasoning strategies matter. In Chapter 3 I suggested that a multimodal perspective be taken when analyzing student development. To support this claim, I presented several multimodal analyses of students' building processes. In many respects these two suggestions occupy vastly different ends of a complexity spectrum. At one end I am discussing the use of student generated content about the origins of their ideas, while at the other end I am leveraging high frequency, multimodal data and computational analysis. Understanding the relationship between these two suggestions is fundamental to recognizing the larger intent of this research.

The overarching argument that I have made is that student reasoning strategies play an important part in their learning experience. Using multimodal analysis provided a means for making this salient. However, I am not advocating that all constructionist learning environments be equipped with multimodal sensors that can measure audio, video, gesture, stress, etc. Instead, I view these data streams as being useful for validating the importance of student reasoning strategies. But having now realized that these strategies are important, conducting multimodal data analysis in every classroom is not needed. I do, however, still encourage practitioners to take a broader perspective on what and how students are learning in Makerspaces, and to strongly consider the ways that instrumentation can better help contextualize student behavior in a way that makes the practitioner more effective in their work. In this case, then, doing work that resides at opposite ends of the spectrum is intended to provide practical and easy to use advice that has been substantiated through a rigorous research agenda.

Conclusion

> "Despite a screaming need for it, tinkering is still stuck in a charming-but-unnecessary peripheral corner in society's collective mind." (Gabrielson, 2013)

In his chapter about the history of tinkering, Curt Gabrielson offers the above lament that tinkering is not viewed as a legitimate learning activity. Gabrielson is not alone; several others have expressed a similar sentiment (e.g. Honey & Kanter, 2012; Resnick & Rosenbaum, 2013). Given the way that the current advocates of the Maker Movement have drawn a contrast between play, and thinking and assessment, this does not come as a surprise. Reconciling "making" with demonstrations of abstract thinking and the ability to quantifiably assess student development, will need to be of central importance as the movement continues to grow. On the whole, though, this is not a new concern (Blikstein, 2013; Honey & Kanter, 2012). What is new, however, is the ways that I have proposed for documenting student development by capturing their reasoning strategies and by taking a multimodal, process-based perspective. These recommendations have not been made to make light of the powerful ways that constructionism can foster agency, engagement and motivation, but instead have been made to suggest that there is an additional form of empowerment that can be derived from students being challenged to more deeply draw on their prior experiences, and developing reasoning strategies that have applicability across contexts. Failure to do this is likely to create a culture of students that can "make," but who have been stunted in their propensity to think deeply about "making". Hence my goal is to promote "making with understanding."

Appendix A. Algorithm Details

In Chapter 3 I presented three analyses that provided justification for using multimodal analysis in identifying the behavioral practices associated with principle-based reasoning, success and learning. In this chapter I provide a more detailed explanation and justification for the algorithm used with the intent that readers will have sufficient understanding to replicate the current analysis. This description proceeds by presenting a detailed account of each step of the algorithm. Figure 67 provides the general overview of the algorithm presented in Chapter 3.

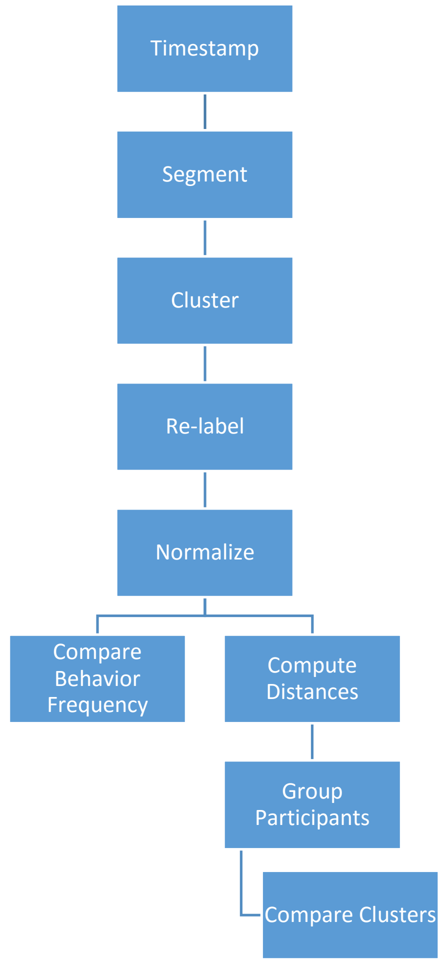


Figure 67. General steps of the data analysis process (moving from top to bottom)

Data Extraction

Data extraction occurred independently across each of the different modalities. In particular, each data stream had been labeled with the local date and time of its occurrence. In the case of the electro-dermal activation data, a data synchronization step was either manually completed, or completed through the Q-sensor software. Apart from the electro-dermal activation sensors, all data was collected on the same computer. Having synchronized data sources allowed me to merge the data as needed for the different analyses.

The other important piece of data annotation included marking the beginning and end of each phase of the experiment. Start and stop times for each activity were recorded based on the images from the Kinect sensor and the video data. Through these annotations I could group data based on the activity that it is associated with. For example, I know which data corresponds with the intervention phase, the design sketch phase, and the building activity phase, for each participant.[23]

Hand-coded data. Hand coding of human actions occurred at approximately 1-second intervals. A snapshot of each pair's behavior was generated at 1-second intervals, and labeled based on the Object Manipulation Class that it corresponded to. The snapshots were generated from a custom application that takes pictures using the Xbox Kinect sensor. In many cases, determination of the manipulation class could only be determined several seconds after the action was completed. For example, as a student is preparing to put two pieces close to one another, it is not clear as to whether this action will be a REALIZE action or a PLAN action, until seeing what the student ultimately does with those two items. In the case that the items are affixed to one another, the action, beginning from the point that the students gets a hold of the materials, would be classified as a REALIZE action. However, if the items were only placed near each other to physically prototype an idea, the entire action sequence would be classified as PLAN.

[23] For later studies, I created a custom tool for doing these annotations in real time, and would strongly recommend that anyone intending to complete a similar analysis, also develop a data annotation tool that can be used in real-time, as opposed to relying on post-hoc annotation

This approach for coding provides an interpretative lens to each action, when compared to an approach in which a given participant's behavior must be described at the time of observation. In previous work I demonstrated that this form of coding is effective for studying students in hands-on learning activities (Worsley and Blikstein 2013, in press)

Audio data. Audio data was derived from a combination of audio channels from an overheard web camera, and audio from the Xbox Kinect sensor. A custom piece of software was developed based on the Carnegie Mellon University (CMU) Sphinx Speech Recognition Toolkit. Specifically, the source code was modified to leverage the program's voice activity detection feature. Voice activity detection is an automated means for determining when voice-based audio in being generated. Several speech recognition software solutions contain some variant of voice activity detection. The custom software provided voice detection start and stop times for all of the audio channels. Audio was considered to be present if either of the audio sources detected a voice, within a given second of time. Thus the final format of this data is a binary representation. Every second of the activity is labeled with a zero or one, for the absence or presence of audio at that time stamp. Because the audio channel captured sound from both participants this piece of data is the same for each person in a pair.

Hand/wrist movement. Hand/wrist movement data was also generated from the Xbox Kinect sensor. Once again a custom built application was used to store three dimensional data for twelve upper body joints. The application uses native features available from the Kinect for Windows SDK, specifically, the ability to conduct skeletal tracking in the seated position. The custom application stores the data at 10 Hz. From the file generated, I utilize only the left and right wrist, hand and elbow data points. For each successive pair of data points I compute the angular displacement for the vectors that connect: left wrist and left hand; left wrist and left elbow, right wrist and right hand; right wrist and right elbow. The eventual angular displacement that is recorded is an average of the four angular displacements. Using angle as the means for comparison reduces biases introduced by participants having different sized limbs. Accordingly, for each tenth of a second in time I have stored the total angular hand/wrist displacement.

Electro-dermal Activation. Electro-dermal activation (also referred to galvanic skin response and/or skin conductance) readings were captured at 8 Hz. Processing of electro-dermal activation involved controlling for individual differences in variance, as well as individual differences in stress response. In practice, this was achieved by collecting baseline data as students completed the task of counting down by 7. I will refer to this as the "math" stress test. As additional baseline data, students also completed a Stroop test, and had their electro-dermal activation recorded during non-task oriented activities. As before, each data point was time-stamped with the local date and time. Each data point was then transformed into an index value by subtracting the mean from the "math" stress test, and then dividing by the standard deviation of the "math" stress test data for that student. As validation that this approach reduced individual bias, when I compared electro-dermal activation index values across the different activities, there were no statistically significant differences between experimental conditions for the baseline data, the Stroop test, or the math test.

However, across the intervention, design phase and the building activity differences were statistically significant.

Segmentation

The hand-coded data included every time one or more persons in a dyad started or completed an EVALUATE action. The starting and ending times for EVALUATE actions were the basis for segmenting the data streams. To be clear, an individual student's sequence would undergo segmentation if she or her partner initiated an EVALUATE behavior. This was done to reflect the fact that either though a given partner may not have initiated a certain structural test, they are likely to gain some insights from observing the outcome of the test, and may adjust their future actions as a result of those insights. Based on video observation this was in fact the case. Students tended to watch even as their partners initiated a test.

As I alluded to in Chapter 3, for the hand-coded data, segmentation resulted in a five dimensional vector, with each column corresponding to one of the Object Manipulation Classes (REALIZE, PLAN, REVERT, MODIFY, NOTHING). Specifically, the value for a given column was the proportion of time that Object Manipulation Class was used during that "test segment." Since these value represent a proportion, in the case of the hand-coded data, the values are not linearly independent as they must all sum to one.

For the multimodal sensor data, segmentation always resulted in a single value for each data stream. For the audio data the value is the proportion of the "test segment" during which voice activity was detected. For the hand/wrist movement data the value is the average total angular displacement during that “test segment.” Finally, the electro-dermal activation value is the average index value during that particular “test segment.”

As a whole, the segmentation process serves to smooth some of the noise in the data. Instead of having to take into account each of the spikes and troughs that may emerge from any of the data streams, segmentation allows me to look more for trends. Noise reduction is also a consideration for the next step.

Clustering

After the segmentation process, there are hundreds of unique “test segments.” Some of these will be very similar to one another, only differing by an infinitesimal amount, while others vary quite extensively from one another. The goal of clustering is to identify natural groupings among the various “test segments” and ultimately provide a common set of states, or behaviors, by which to compare individual user sequences. However, before proceeding with clustering, I first do data standardization. Namely, I adjust each value, such that all of the data in a given column has a mean of zero and a standard deviation of one. This process eliminates bias in clustering, by ensuring that each column contributes equally to the distance metric, which is this case was Euclidean distance. After standardizing the data, I used X-Means clustering to group the data points into a set of clusters that place each “test segment” with the other “test segments” that it is most similar to. Once each “test segment” has been grouped with similar “test segments,” each cluster, or group, can be described based on the average values of all of its members. These values provide the basis for determining common behavioral practices across the three analyses in Chapter 3.

Re-label

With each “test segment” assigned to a single cluster, it is simple to reconstruct a given user’s sequence of “test segments” in terms of the clusters. This is the purpose of relabeling.

Normalization

Two separate normalization processes are used depending on the eventual analysis. The first, and simplest is L-1 normalization, which scales each sequence to the length of the longest student sequence. Conducting L-1 normalization maintains the order and relative proportion of each cluster type, but increases the total number of “test segments” for all students who had fewer than the maximum number of “test segments.” The other form of normalization is dynamic time warping, which computes the minimum distance between every pair of students, by using a variant of the Levenshtein distance.

As with L-1 normalization, dynamic time warping preserves the order of the data. However, unlike L-1 normalization, dynamic time warping may change the proportion of total time spent using each cluster type. Because of this, I used the L-1 normalization when examining the frequency of cluster usage, but broke the cluster usage frequency into a beginning, middle and final segment to maintain some of the temporal elements of students’ sequences. For answering questions around the cycles of iterations, and similarity in point-by-point process data, the dynamic time warping algorithm seemed more appropriate, as it preserves the order of the process and is able to capture the distance between each pair of participants.

Group Participants

After completing dynamic time warping, pairwise distances are computed. Those pairwise distances are used to construct an n-by-n matrix. As before, this matrix is standardized before conducting K-Means clustering with k=2, as there are two experimental conditions.

Cluster Comparisons

From the two clusters of participants, I compute the probability that the cluster assignments would be randomly derived, as determined through a binomial test.

Appendix B. Example Intermediate Structures

1. 
2. 
3. 
4. 
5. 
6. 
7. 

Appendix C. Additional Definitions of Making

Dr. Eli Silk, one of the keynotes at the Design, Make, Play workshop, defines each term as well as their interplay.

The design-make-play triad leads to "a collection of well-designed activities, targeting a full spectrum of STEM understandings and skills, engaging to the full range of learners."

Design is a process with a goal—to create a working solution to a problem, governed by specifications and constraints. The process involves generating multiple solutions, building and testing models, analyzing solutions, and improving on those solutions.

Making is about building and figuring out how to build, driven by personal interest and concerns. It involves getting things to work, with exploratory tinkering and manipulating objects beyond their typical use. Open source and cheap and common materials encourage hacking and adapting. The process of making is not just a means to an end, but has its own value, and involves adapting and customizing and sharing with others so they can adapt it to their own uses.

Play is voluntary and flexible, active and sometimes about make-believe. It has no extrinsic goal except to have fun. The process is minimally constrained so it is easy to pursue new directions, and there is no external judgment so the player can do things that might otherwise be unreasonable.

Honey and Kanter (2013) provide a similar definition of "making:

> Make- to build or adapt objects by hand, for the simple personal pleasure of figuring out how things work. Long before the rules of science were written down people engaged with scientific disciplines by making things, things to help us do what we need to do, or things that are just fun.

Tim Ingold (2013) has the following to say about making

> We are accustomed to think of making as a *project*. This is to start with an idea in mind, of what we want to achieve, and with a supply of raw material needed to achieve it. And it is to finish at the moment when the material has taken on the intended form….
>
> I want to think or making, instead as a process of *growth*. This is to place the maker from the outset as a participant in amongst a world of active materials. These materials ware what he has to work with, and in the process of making he 'joins forces' with them, bringing them together or splitting them apart, synthesizing and distilling, in anticipation of what might emerge.

Resnick & Rosenbaum (2013) contribute to this discussion by describing tinkering.

> We see tinkering as a valid and valuable style of working, characterized by a playful, exploratory, iterative style of engaging with a problem or project. When people are tinkering, they are constantly trying out ideas, making adjustments and refinements, then experimenting with new possibilities, over and over and over.

Brahms and Werner (2013 in Honey & Kanter, 2013) define:

"[M]ake: to mess around at the crossroads and fringes of disciplines such as technology, engineering, art and math."

Appendix D. Classroom Study Worksheets

Example-based Reasoning Intervention

Research shows that in order to come up with the best idea, it's often helpful to come up with three viable options before settling in on a final design. To help you leverage this tool, we will show you three sample structures that may be helpful in informing the design of your structure.



Additionally, before you start working on the actual design of your structure, please identify three real world objects from your home, school, community, or everyday life that might be useful for your design. Please use the space below to come up with a picture or a written explanation of this structure.

1.

2.

3.

Use the space below to draw a picture of what you think your final structure will look like: (full blank sheet provided)

Principle-based Reasoning Intervention

Research shows that in order to come up with the best idea, it's often helpful to come up with three viable options before settling in on a final design. To help you leverage this tool, we will show you three sample structures that may be helpful in informing the design of your structure.



Additionally, before you start working on the actual design of your structure, please identify three things about the above structures that give them stability and note that in the spaces below. You only need to generate three mechanisms total. Please use the space below to come up with a picture or a written explanation of this structure.

1.

2.

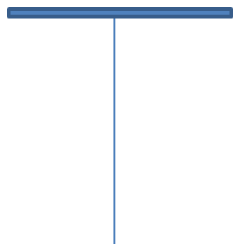
3.

Use the space below to draw a picture of what you think your final structure will look like: (full blank sheet provided)

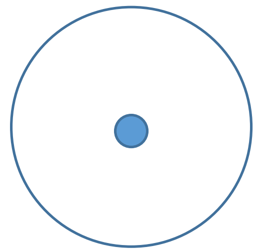
Post-test Worksheet

Starting with the following structure, please come up with as many of the easiest ways that you would make this structure more stable. When thinking of the easiest ways, you can also interpret this as the least expensive ways.

Side View:



Bottom View:



1.		
2.		
3.		
4.		
5.		
6.		
7.		
8.		
9.		
10.		

1. How did you come up with your design?

2. Did something in particular motivate the design?

3. Looking back what, if anything would you do differently?

4. Any additional thoughts?

References

Ahmed, S., & Wallace, K. M. (2003). Understanding the differences between how novice and experienced designers approach design tasks, *14*, 1–11. doi:10.1007/s00163-002-0023-z

Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology*, *103*(1), 1–18. doi:10.1037/a0021017

Anderson, J. (2002). Spanning seven orders of magnitude: a challenge for cognitive modeling. *Cognitive Science*, *26*(1), 85–112. doi:10.1207/s15516709cog2601_3

Anderson, J., & Greeno, J. (1981). Acquisition of problem-solving skill.

Apedoe, X. S., & Schunn, C. D. (2012). Strategies for success: uncovering what makes students successful in design and learning. *Instructional Science*, *41*(4), 773–791. doi:10.1007/s11251-012-9251-4

Atman, C., & Bursic, K. (1998). Verbal protocol analysis as a method to document engineering student design processes. *Journal of Engineering Education*, (April), 121–132.

Atman, C., Cardella, M., Turns, J., & Adams, R. (2005). Comparing freshman and senior engineering design processes: an in-depth follow-up study. *Design Studies*.

Atman, C., Chimka, J., Bursic, K., & Nachtmann, H. (1999). A comparison of freshman and senior engineering design processes. *Design Studies*.

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

Barron, B. (2003). When smart groups fail. *The Journal of the Learning Sciences*, (January 2014), 37–41. doi:10.1207/S15327809JLS1203

Barron, B., Pea, R., & Engle, R. (2013). Advancing understanding of collaborative learning with data derived from video records. *International Handbook of Collaborative Learning*, 1–40.

Barron, B., Schwartz, D., Vye, N., Moore, A., Petrosino, A., Zech, L., & Bransford, J. (1998). Doing With Understanding: Lessons From Research on Problem and Project-Based Learning. *Journal of the Learning Sciences*, *7*(3), 271–311. doi:10.1207/s15327809jls0703&4_2

Berland, M., Baker, R. S., & Blikstein, P. (in press). Educational Data Mining and Learning Analytics: Applications to Constructionist Research. *Technology, Knowledge and Learning*, 1–16.

Berland, M., Martin, T., Benton, T., Petrick Smith, C., & Davis, D. (2013). Using Learning Analytics to Understand the Learning Pathways of Novice Programmers. *Journal of the Learning Sciences*, *22*(4), 564–599. doi:10.1080/10508406.2013.836655

Bjork, R. (2013). Desirable difficulties perspective on learning. *Encyclopedia of the Mind. Thousand Oaks: Sage*.

Blikstein, P. (2011). Using learning analytics to assess students' behavior in open-ended programming tasks. In *Proceedings of the 1st international conference on learning analytics and knowledge* (pp. 110–116).

Blikstein, P. (2013). Digital fabrication and "making"in education: The democratization of invention. *FabLabs: Of Machines, Makers and Inventors*, 1–21.

Blikstein, P., & Krannich, D. (2013). The makers' movement and FabLabs in education: experiences, technologies, and research. In *Proceedings of the 12th International Conference on Interaction Design and Children* (pp. 613–616).

Blikstein, P., Worsley, M., Piech, C., Sahami, M., Cooper, S., & Koller, D. (in press). Programming pluralism: Using learning analytics to detect patterns in the learning of computer programming. *Journal of Learning Sciences*.

Bolger, M. S., Kobiela, M., Weinberg, P. J., & Lehrer, R. (2012). Children's Mechanistic Reasoning. *Cognition and Instruction*, *30*(2), 170–206. doi:10.1080/07370008.2012.661815

Brennan, K., & Resnick, M. (2012). New frameworks for studying and assessing the development of computational thinking. *Proceedings of the 2012 Annual Meeting of the …*, 1–25.

Brown, B. a., & Ryoo, K. (2008). Teaching science as a language: A "content-first" approach to science teaching. *Journal of Research in Science Teaching*, *45*(5), 529–553. doi:10.1002/tea.20255

Brown, B. A., & Spang, E. (2008). Double talk: Synthesizing everyday and science language in the classroom. *Science Education*, *92*(4), 708–732. doi:10.1002/sce.20251

Brown, B., & Kloser, M. (2009). A view of the tip of the iceberg: revisiting conceptual continuities and their implications for science learning. *Cultural Studies of Science Education*, 921–928.

Bruner, J. S. (1960). *The Process of Education*.

Bruner, J. S., Goodnow, J., & Austin, G. (1956). *A study of thinking*.

Carbonell, J. (1982). Learning by analogy : formulating and generalizing plans from past experience.

Chi, M. T. H., Glaser, R., & Rees, E. (1981). Expertise in problem solving.

Colhoun, J., Gentner, D., & Loewenstein, J. (2008). Learning abstract principles through principle-case comparison. In *Proceedings of Cognitive Science Society* (pp. 1659-1664).

Collins, A., & Ferguson, W. (1993). Epistemic forms and Epistemic Games: Structures and Strategies to Guide Inquiry. *Educational Psychologist*, *28*(1), 25–42. doi:10.1207/s15326985ep2801_3

Collins, A., & Halverson, R. (2009). *Rethinking education in the age of technology: The digital revolution and schooling in America*. Teachers College Press.

Cross, N., & Cross, A. C. (1998). Expertise in Engineering Design 2 . An Outstanding Designer, 141–149.

Csikszentmihalyi, M. (1992). *Flow : the psychology of happiness*. London: Rider.

Ding, C., & He, X. (2004). K-means clustering via principal component analysis. *Proceedings of the Twenty-First International Conference of Machine Learning.* p. 29.

diSessa, A. (1993). Toward an epistemology of physics. *Cognition and Instruction*, *10*(2), 105–225.

diSessa, A. (2008). Can Students Re-Invent Fundamental Scientific Principles? Evaluating the Promise of New-Media Literacies. *Children's Learning in a Digital World*.

diSessa, A., Gillespie, N., & Esterly, J. (2004). Coherence versus fragmentation in the development of the concept of force. *Cognitive Science*, *28*(6), 843–900. doi:10.1016/j.cogsci.2004.05.003

Dougherty, D. (2013). The maker mindset. *Design, Make, Play: Growing the Next Generation of STEM Innovators.* Routledge.

Dougherty, D., Thomas, P., Chang, S., Hoefer, S., Alexander, I., & Mcguire, D. (2013). *Makerspace Playbook.*

Drescher, G. L. (1987). Object-oriented Logo. In *Artificial Intelligence and Education* (Vol. 1, pp. 153–165).

Elby, A., & Hammer, D. (2010). 13 Epistemological resources and framing: a cognitive framework for helping teachers interpret and respond to their students' epistemologies. *Personal Epistemology in the Classroom*, (3), 1–32.

Epstein, R. (1999). Generativity theory. *Encyclopedia of Creativity*.

Francis, W., Fernandez, N., & Bjork, R. (2010). Conceptual and non-conceptual repetition priming in category exemplar generation: Evidence from bilinguals. *Memory*, (915).

Gabrielson, C. (2013). *Tinkering: Kids Learn by Making Stuff*. Maker Media, Inc.

Gentner, D. (2004). Analogical Encoding : Facilitating Knowledge Transfer and Integration.

Gentner, D., & Holyoak, K. J. (1997). Reasoning and Learning by Analogy, *52*(1), 32–34.

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology*, *95*(2), 393–408. doi:10.1037/0022-0663.95.2.393

Gick, M., & Holyoak, K. (1980). Analogical problem solving. *Cognitive Psychology*, *355*, 306–355.

Hammer, D. (2000). Student resources for learning introductory physics. *American Journal of Physics*, *68*(S1), S52. doi:10.1119/1.19520

Hammer, D. (2004a). The Variability of Student Reasoning , Lecture 1 : Case Studies of Children ' s Inquiries, 1–15.

Hammer, D. (2004b). The Variability of Student Reasoning , Lecture 2 : Transitions.

Hammer, D. (2004c). The Variability of Student Reasoning , Lecture 3 : Manifold Cognitive Resources, 1–15.

Hammer, D., Elby, A., Scherr, R. E., & Redish, E. F. (2005). Resources, framing, and transfer. *Transfer of Learning from a Modern Multidisciplinary Perspective*, 89–120.

Hammer, D., Russ, R., Mikeska, J., & Scherr, R. (2008). Identifying inquiry and conceptualizing students' abilities. *Teaching scientific inquiry: Recommendations for research and implementation*, 138-156.

Harel, I. E., & Papert, S. E. (1991). *Constructionism.* Ablex Publishing.

Hartley, L. M., Wilke, B. J., Schramm, J. W., D'Avanzo, C., & Anderson, C. W. (2011). College Students' Understanding of the Carbon Cycle: Contrasting Principle-based and Informal Reasoning. *BioScience*, *61*(1), 65–75. doi:10.1525/bio.2011.61.1.12

Honey, M., & Kanter, D. E. (2012). *Design, make, play: Growing the next generation of Science Innovators*.

Honey, M., & Kanter, D. E. (2013). *Design, Make, Play: Growing the Next Generation of STEM Innovators*. Routledge.

Hutchison, P., & Hammer, D. (2009). Attending to student epistemological framing in a science classroom. *Science Education*, n/a–n/a. doi:10.1002/sce.20373

Kafai, Y. B. (1995). *Minds in play: Computer game design as a context for children's learning*. Routledge.

Kirschner, P. A. (2002). Cognitive load theory : implications of cognitive load theory on the design of learning, *12*, 1–10.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2010). Why Minimal Guidance During Instruction Does Not Work : An Analysis of the Failure of Constructivist , Based Teaching Work : An Analysis of the Failure of Constructivist , Discovery , Problem-Based , Experiential , and Inquiry-Based Teaching, (November 2013), 37–41. doi:10.1207/s15326985ep4102

Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science*, *18*(4), 513–549.

Kolodner, J. L. (1992). An Introduction to Case-Based Reasoning *.

Kress, G. (2001). Multimodal teaching and learning: The rhetorics of the science classroom.

Kurtz, K., & Loewenstein, J. (2007). Converging on a new role for analogy in problem solving and retrieval: When two problems are better than one. *Memory & Cognition*.

Lawler, R. W. (1987). Learning environments: now, then and someday. In *Artificial intelligence and education* (Vol. 1, pp. 1–26).

Lawler, R. W., & Yazdani, M. (1987). *Artificial Intelligence and Education: Learning environments and tutoring systems* (Vol. 1). Intellect Books.

Lehrer, R., & Schauble, L. (1998). Reasoning about structure and function: Children's conceptions of gears. *Journal of Research in Science Teaching*, *35*(1), 3–25. doi:10.1002/(SICI)1098-2736(199801)35:1<3::AID-TEA2>3.0.CO;2-X

Levenshtein, V. I. (1966). Binary Codes Capable of Correcting Deletions, Insertions and Reversals. *Soviet Physics Doklady*, *10*(8), 707–710.

Loewenstein, J. (2010). How one's hook is baited matters for catching an analogy. *Psychology of Learning and Motivation*, *53*, 1–63.

Maier, N. R. F. (1931). Reasoning in humans. II. The solution of a problem and its appearance in consciousness. *Journal of Comparative Psychology*. US: Williams & Wilkins Company. doi:10.1037/h0071361

Maloney, J., Burd, L., Kafai, Y., Rusk, N., Silverman, B., & Resnick, M. (2004). Scratch: a sneak preview [education]. In *Creating, Connecting and Collaborating through Computing, 2004. Proceedings. Second International Conference on* (pp. 104–109).

Martin, F. (1988). *Children, cybernetics, and programmable turtles*. Massachusetts Institute of Technology.

Martin, F., & Resnick, M. (1993). LEGO/Logo and Electronic Bricks: Creating a Scienceland for Children. In D. Ferguson (Ed.), *Advanced Educational Technologies for Mathematics and Science* (Vol. 107, pp. 61–89). Springer Berlin Heidelberg. doi:10.1007/978-3-662-02938-1_2

Martinez, S. L., & Stager, G. (2013). *Invent to learn: Making, tinkering, and engineering in the classroom*. Constructing Modern Knowledge Press.

McLaren, C. (2012). Making Makers: An Interview with Dale Dougherty.

Menekse, M., Stump, G. S., Krause, S., & Chi, M. T. H. (2013). Differentiated Overt Learning Activities for Effective Instruction in Engineering Classrooms. *Journal of Engineering Education*, *102*(3), 346–374. doi:10.1002/jee.20021

Mosborg, S., Adams, R., & Kim, R. (2005). Conceptions of the engineering design process: An expert study of advanced practicing professionals. *The 2005 Annual Conference of the American Society for Engineering Education*.

Moss, J., Kotovsky, K., & Cagan, J. (2006). The role of functionality in the mental representations of engineering students: Some differences in the early stages of expertise. *Cognitive Science*, *30*, 65–93.

Newell, A. (1994). *Unified theories of cognition*. Harvard University Press.

Nokes, T., Schunn, C. D., & Chi, M. T. H. (2010). Problem solving and human expertise. *International Encyclopedia of Education*, *5*, 265–272.

Obama, B. National Day of Making (2014). Washington, D.C.: Executive.

Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. Basic Books, Inc.

Papert, S. (1987a). Constructionism: A new opportunity for elementary science education.[National Science Foundation Award Abstract No. 8751190]

Papert, S. (1987b). Microworlds: transforming education. In *Artificial intelligence and education* (Vol. 1, pp. 79–94).

Papert, S. (2002). Hard Fun.

Papert, S., & Harel, I. (1991). Situating constructionism. In I. Harel & S. Papert (Eds.), *Constructionism* (Vol. 36, pp. 1–11).

Pea, R., Mills, M., Rosen, J., Dauber, K., Effelsberg, W., & Hoffert, E. (2004). The diver project: Interactive digital video repurposing. *MultiMedia, IEEE*, *11*(1), 54–61.

Piaget, J. (1973). To understand is to invent: the future of education (G. Roberts, Trans.). *NY: Grossman Publishers*.

Piech, C., Sahami, M., Koller, D., Cooper, S., & Blikstein, P. (2012). Modeling How Students Learn to Program. In *Proceedings of the 43rd ACM Technical Symposium on Computer Science Education* (pp. 153–160). New York, NY, USA: ACM. doi:10.1145/2157136.2157182

Polya, G. (1945). *How to Solve It*.

Rabiner, L. R., Rosenberg, A. E., & Levinson, S. E. (1978). Considerations in dynamic time warping algorithms for discrete word recognition. *The Journal of the Acoustical Society of America*, *63*(S1), S79–S79.

Reeves, L. M., & Weisberg, R. W. (1994). The Role of Content and Abstract Information in Analogical Transfer, *1*(3).

Resnick, M. (1990). MultiLogo: A study of children and concurrent programming. *Interactive Learning Environments*, *1*(3), 153–170.

Resnick, M. (1996). StarLogo: An environment for decentralized modeling and decentralized thinking. In *Conference companion on Human factors in computing systems* (pp. 11–12).

Resnick, M., Berg, R., & Eisenberg, M. (2000). Beyond black boxes: Bringing transparency and aesthetics back to scientific investigation. *The Journal of the Learning Sciences*, *9*(1), 7–30.

Resnick, M., & Rosenbaum, E. (2013). Designing for tinkerability. *Design, Make, Play: Growing the Next STEM Innovators.*

Roll, I. (2009). Structured Invention Tasks to Prepare Students for Future Learning: Means, Mechanisms, and Cognitive Processes, (December).

Roschelle, J., & Teasley, S. (1995). The construction of shared knowledge in collaborative problem solving. *Computer Supported Collaborative Learning*.

Russ, R. S., Coffey, J. E., Hammer, D., & Hutchison, P. (2009). Making classroom assessment more accountable to scientific reasoning: A case for attending to mechanistic thinking. *Science Education*, *93*(5), 875–891. doi:10.1002/sce.20320

Russ, R. S., Lee, V. R., & Sherin, B. L. (2012). Framing in cognitive clinical interviews about intuitive science knowledge: Dynamic student understandings of the discourse interaction. *Science Education*, *96*(4), 573–599. doi:10.1002/sce.21014

Scherr, R. E., & Hammer, D. (2009). Student Behavior and Epistemological Framing: Examples from Collaborative Active-Learning Activities in Physics. *Cognition and Instruction*, *27*(2), 147–174. doi:10.1080/07370000902797379

Schwartz, D. L. (1992). Constructivism in an age of non-constructivist assessments.

Schwartz, D. L. (1995). The emergence of abstract representations in dyad problem solving. *The Journal of the Learning Sciences*, *4*(3), 321–354.

Silver, J., Rosenbaum, E., & Shaw, D. (2012). Makey Makey: Improvising Tangible and Nature-based User Interfaces. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction* (pp. 367–370). New York, NY, USA: ACM. doi:10.1145/2148131.2148219

Sipitakiat, A., Blikstein, P., & Cavallo, D. (2002). The GoGo Board: Moving towards highly available computational tools in learning environments. In *Proceedings of Interactive Computer Aided Learning International Workshop*.

Smith, J., diSessa, A., & Roschelle, J. (1994). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences*.

Steele, C. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*(6), 613–629.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. doi:10.1016/0364-0213(88)90023-7

Tisue, S., & Wilensky, U. (2004). Netlogo: A simple environment for modeling complexity. In *International Conference on Complex Systems* (pp. 16–21).

Toulmin, S. (1999). Knowledge as shared procedures.

Turkle, S., & Papert, S. (1992). Epistemological Pluralism and the Revaluation of the Concrete. *Journal of Mathematical Behavior*, *11*(1), 1–30.

Tyack, D., & Cuban, L. (1995). *Tinkering toward utopia: a century of public school reform*. Cambridge, Mass.: Harvard University Press.

VanLehn, K. (1996). Cognitive skill acquisition. *Annual Review of Psychology*, *47*, 513–39. doi:10.1146/annurev.psych.47.1.513

Werner, H. (1937). Process and achievement—a basic problem of education and developmental psychology. *Harvard Educational Review*.

Wilensky, U. (1991). Abstract meditations on the concrete and concrete implications for mathematics education.

Wilson, C. D., Anderson, C. W., Heidemann, M., Merrill, J. E., Merritt, B. W., Richmond, G., … Parker, J. M. (2006). Assessing students' ability to trace matter in dynamic systems in cell biology. *CBE Life Sciences Education*, *5*(4), 323–31. doi:10.1187/cbe.06-02-0142

Windsor, D. (2013). ConstructionKids. Retrieved from http://dmp.nysci.org/#case-study/41

Worsley, M., & Blikstein, P. (in press). Analyzing Engineering Design through the Lens of Computation. *Journal of Learning Analytics*, *1*(2).

Worsley, M., & Blikstein, P. (2011). What's an Expert? Using Learning Analytics to Identify Emergent Markers of Expertise through Automated Speech, Sentiment and Sketch Analysis. In *Proceedings of the Fourth Annual Conference on Educational Data Mining (EDM 2011)* (pp. 235–240). Eindhoven, Netherlands.

Worsley, M., & Blikstein, P. (2013). Towards the Development of Multimodal Action Based Assessment. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 94–101). New York, NY, USA: ACM. doi:10.1145/2460296.2460315