ARGUMENTATION IN THE CONTEXT OF SCIENCE EXPERIMENTATION AS PREPARATION FOR INFORMED DESIGN DECISION-MAKING

by

Ying Ying Seah

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Computer and Information Technology

West Lafayette, Indiana

May 2021

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Alejandra J. Magana, Chair

Department of Computer Information Technology

Dr. Sanjay Rebello

Department of Physics and Astronomy

Dr. Carina M. Rebello

Department of Physics and Astronomy

Dr. Lisa Bosman

Department of Technology, Leadership and Innovation

Approved by:

Dr. Kathy Newton

To my loving family, Héctor, Kok Hooi, Siew Eng, Yang Yang, and Yang Ling:

I could not have done this without your unconditional love and constant support.

ACKNOWLEDGEMENTS

Throughout my graduate school experience at Purdue University, I have been blessed with people who continuously and unconditionally love, support, and assist me. First of all, I would like to thank my advisor, Dr. Alejandra J. Magana for her guidance and support all these years. She is the one who inspired me to take on the research journey and has faithfully been there to guide and mentor me. I have learned so much from her and I wouldn't be where I am today without her. I would also like to thank the other members of my committee for their support and contribution. I enjoy and appreciate the opportunities to work with Dr. Sanjay Rebello and Dr. Carina Rebello, especially for allowing me to collect data from their classrooms, and for their motivating and constructive feedback. I also value and appreciate working with Dr. Lisa Bosman, especially for her feedback and encouragement for me to think and reflect more from perspectives that I have not thought of before.

Besides, I would also like to thank all my colleagues from the Rocketed research group, for all the opportunities that we got to work and learn together, and for all the helpful feedback and discussion at our weekly group meetings. I am also grateful for all other professors and administrative staff at Purdue University that I have the honor of working with, for playing their crucial parts to help get me to where I am today. In addition, I am thankful for the funding from the National Science Foundation for supporting my research projects, and also the travel grants from the Purdue Polytechnic Institute for supporting my professional involvement at conferences.

Many thanks to all my friends at Purdue University and church family at Faith Church who made my graduate life so much easier and enjoyable. Thank you for walking along side me to encourage me and help me grow. Special thanks to the Krismartanto family for being my closest family away from home, for making me meals when I was too busy, and for all the sacrificial love you pour into my life.

Big thank you to all my family who unconditionally love and support me throughout these years. I would like to thank my loving husband, Héctor, for always being there for me (for the good times and the bad times), and for being so patient with me. I sure love and appreciate a partner who loves having nerdy conversations with me, which makes research so much more fun. I would

also like to thank my parents, for their love and care all my life, and for being understanding with me being so far away from home. I am thankful for my two brothers, who are always there for me when I need them, and for cheering me for each of my accomplishments. Additionally, I would like to thank my family in-law for their care and support as well.

Above all, I would like to thank and praise my Lord and Savior, Jesus Christ, for all that He has provided, including His words, the people He has placed in my life, and for allowing me to achieve this accomplishment.

"Enter His gates with thanksgiving and His courts with praise; Give thanks to Him and praise His name."

Psalm 100:4

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The ability to make informed decisions is a skill considered as one of the 21st century skills and is crucial as part of the critical thinking and problem-solving process in science and engineering. Despite its importance, students (e.g., beginning designers) often struggle with making informed design decisions that are well supported by relevant scientific principles. It is not uncommon to see disconnection between students' design decisions and their scientific knowledge. This type of disconnection is also described as the "design-science gap". Different approaches such as scaffolding have been done in trying to bridge this gap, however there is still limited scaffold that could seamlessly help students connect their scientific knowledge to their design experiences, and consequently help them make scientifically informed design decisions.

In this dissertation, we proposed argumentation as a scaffolding framework and investigated if the use of argumentation as a meaning-making scaffolding approach during scientific experimentation, facilitated students' generation of informed design decisions while completing a CAD-based design challenge. Specifically, we looked at the impact of the argumentation scaffold on the quality of decision-making arguments made by students, the types of claims made by students and the types of evidence and reasoning they used to back up their claims, as well as their level of performance in a final design challenge.

This study took place in a Physics for Elementary Education course in a Midwestern University in Indiana, USA. This study was part of a four-week unit that focused on the topic of heat transfer, as well as the practices of science and engineering design. The participants of this study included 54 groups of pre-service teachers (i.e., 2 to 4 students in each group) with a background in Elementary Education, from three academic semesters: Spring 2018, Spring 2019, Fall 2019. In this study, these pre-service teachers were divided into two conditions – with and without argumentation scaffold. The data analysis involved looking at the quality of students' decision-making arguments, the types of claim, evidence, and reasoning they used, as well as their final design performances.

The results of this study indicate that students in the argumentation condition were able to transfer their argumentation skills from science experimentation to design decision-making by

demonstrating better ability to justify their decisions using relevant scientific evidence and reasoning, as compared to students without argumentation scaffold. Specifically, students engaged in the argumentation scaffold generated decision-making arguments of higher quality, devoted more attention to scientific principles when they made their decision claims, used more variety of combinations of evidence and reasoning to support their claims, utilized more scientific principles to back up their claims, as well as achieved slightly better performance in their final design in terms of fulfilling the size and energy consumption requirements. Implications from this dissertation include pedagogical scaffold and assessment materials that can be easily adapted by other educators, along with suggestions based on what we learned. In addition, findings and lessons learned from this study open door to more research opportunities such as expanding and adapting the scientific argumentation framework to better fit in an engineering design context.

CHAPTER 1: INTRODUCTION

Research Problem

As preparation for facing and overcoming the challenges of the 21st century in the field of science and engineering, students need to be equipped with 21st-century skills for them to stay productive and competitive in this globalized world (Turiman, Omar, Daud, & Osman, 2012). The framework for 21st Century Learning has identified critical thinking and problem solving as one of the important learning and innovation skills that students should master (P21: Partnership for 21st Century Learning, 2019). As part of the critical thinking and problem-solving process in science and engineering, students need to be able to make an informed decision by knowing how to identify credible and reliable sources, interpret scientific information, distinguish between fact and fiction (or opinion), and construct an argument based on evidence (Turiman et al., 2012). These ways of thinking have also been identified as decision-making processes critical for 21st-century skills (Binkley et al., 2014). Regardless of their importance, classrooms today, unfortunately, still lack 21st-century learning and teaching, including skills like decision-making. Therefore, in this study, we focused on how we can pedagogically help students develop design decision-making skills through the context of science experiments.

As an important process in engineering design, decision-making was described as an iterative process of acknowledging and managing tradeoffs, due to uncertainty, complexity, interrelatedness, and situatedness of problems (Jonassen, 2012). During decision-making, designers identify alternative options, formulate guidelines to steer the directions of decision-making, and review as much relevant information as possible to evaluate the strengths and weaknesses of potential solutions before making the final design decision (Papadouris, 2012). Decision making is a complex and challenging process, especially for beginning designers. Research reveals that informed designers use strategies such as: do research, conduct deep modeling, balance tradeoffs, perform valid experiments, carry out diagnostic troubleshooting, and execute reflective design thinking when it comes to making design decisions (Crismond & Adams, 2012). Beginning designers, however, "ignore or pay too little attention to design criteria and constraints, and focus only on positive and negative aspects of their design ideas without thinking of associated benefits and tradeoffs" (Crismond & Adams, 2012, p.24).

In addition, beginning designers usually struggle to compare solutions effectively and consistently, avoid holistic methods that systematically weigh tradeoffs, as well as apply decision-making approaches in ad hoc manner (Papadouris, 2012). Moreover, some beginning designers struggle with connecting the design problems at hand with the underlying science concepts (Chao et al., 2017; Roth, Tobin, & Ritchie, 2001), as well as arguing or articulating the rationale for the different alternatives that they are considering (Crismond & Adams, 2012). It is also not unusual to see students copying design ideas either from peers or the Internet instead of relying on scientific principles and practices to aid their thinking and decision-making processes (Chao et al., 2017). Studies have also shown that students tend to demonstrate lack of effort in terms of collecting background information necessary for evaluating their designs, and even when they collected those information, the data collected was not always relevant to science (Mentzer, 2014). In cases when students were requested to provide explanations for their designs, they were likely to focus more on the physical model or function of their designs, instead of the underlying principles that make their designs work (Carlsen, 1998).

A scientific practice highly related to informed design decision making is the process of argumentation (Erduran, Ozdem, & Park, 2015). Construction and critique of scientific explanations, often described as argumentation, involve the process of rationally answering questions by shifting the focus from answer-oriented problem solving to the process-oriented practice of constructing and justifying claims (Berland & McNeill, 2010). Integrating argumentation in scientific experimentation can enhance conceptual understanding (Nussbaum, Sinatra, & Poliquin, 2008). Argumentation can be incorporated into scientific experimentation by asking students to make evidence-based claims and also providing the reasoning behind those (McNeill, Lizotte, Krajcik, & Marx, 2006). In decision-making, argumentation is seen as the process of evaluating theoretical claims based on empirical evidence or data from relevant sources (Kuhn, 1993). In other words, it is about the ability to choose and reason between different alternatives or explanations, which ultimately lead to making a choice (Jiménez-Aleixandre & Pereiro-Muñoz, 2002). Past research has shown the importance of using prompts and criteria instruction in argumentation (Nussbaum & Schraw, 2007; Zohar & Nemet, 2002). For example, Nussbaum and Schraw (2007) found that by providing students explicit instructions on the qualities of an argument as well as an example, students were able to produce arguments with higher quality and complexity. Unlike in the science domain, there is still limited research for the implementation of argumentation in the context of engineering (Mathis, Siverling, Glancy, & Moore, 2017). Even though the goal of argumentation is different for each domain (that is, scientists utilize arguments to evaluate and explain natural phenomena, whereas engineers utilize arguments to optimize the solutions of a problem), it is undeniable that argumentation is also crucial in engineering and therefore deserves research attention.

In this study, we investigated the effect of engaging students in argumentation practices during science experimentation and identify the impact of this pedagogical approach on students' decision-making processes in a related design challenge. Specifically, we implemented a 4-week long lesson on heat transfer in a physics course for pre-service teachers, where we first implemented an argumentation scaffold as students worked on physical lab experiments, and later identified possible effects as part of a design challenge. This study primarily aimed to answer the question of: how did the scaffold of argumentation in the context of science experiment influence students' design decision-making?

Research Questions

As mentioned previously, this research study was primarily led by the research question of: how did the scaffold of argumentation in the context of science experiment influence students' design decision-making? The sub-questions that follow are:

SRQ-1: What was the quality of decision-making arguments made by students?
SRQ-2: What types of claims were students making when they designed, and what types of evidence and reasoning were students using to back up their claims?
SRQ-3: What was students' level of performance in their final designs?

Scope

This research study mainly studied the effect of students' engagement in argumentation practices through science experiments, as well as the impact of this pedagogical approach on students' design decision process through a design challenge. Even though the data collected through this research study was comprehensive and rich, however, for the purpose of the research focus, we only focused on analyzing the group arguments documented by students, as well as students' final design artifacts.

Dissertation Structure

This dissertation is divided into eight chapters. The current chapter, Chapter 1, includes the research problem, discusses its purpose, significance, scope, and introduces the research questions that are guiding this study. Chapter 2 includes a literature review of relevant topics. Chapter 3 includes the theoretical framework that guides this study. Chapter 4 includes the learning design of this study, including context, learners' characteristic, learning objectives, pedagogical approach, learning materials, as well as assessment. Chapter 5 includes the methods of this study, including settings and participants, procedure, data collection method and data sources, data analysis method, as well as trustworthiness. Chapter 6 includes the results of this study, followed by discussion and implications for teaching, learning, and research in Chapter 7. Chapter 8 concludes this study, in addition to describing its limitations and potential future work.

CHAPTER 2: REVIEW OF LITERATURE

Engineering design has become more relevant to learning and as a result it has been integrated into science learning, especially in the K-12 settings. This is evident through the implementation of various science standards or accreditation standards (e.g., ABET, 2018; CAEP, 2013; Indiana Department of Education, 2010; NGSS Lead States, 2013). Table 1 below presents a few examples from these standards that highlight the integration of engineering design into science learning.

Table 1. Examples from science or accreditation standards that highlight the integration of engineering design into science learning.

Standard	**Item**	**Target Population**	**Criteria**
Next Generation Science Standards (NGSS)	MS-PS3-3 Energy	K-12	*"Apply scientific ideas or principles to design, construct, and test a design of an object, tool, process or system."*
Indiana's Academic Standards for Science - 2010 Science and Engineering Process Standards (SEPS)	SEPS 6 - Constructing explanations (for science) and designing solutions (for engineering)	K-8	*"Scientists and engineers use their results from the investigation in constructing descriptions and explanations, citing the interpretation of data, connecting the investigation to how the natural and designed world(s) work. They construct or design logical coherent explanations or solutions of phenomena that incorporate their understanding of science and/or engineering or a model that represents it and are consistent with the available evidence."*
Council for the Accreditation of Education Preparation (CAEP) 2013 Standards	Standard 1 - Content and Pedagogical Knowledge	Educators	*"The provider ensures that candidates develop a deep understanding of the critical concepts and principles of their discipline and, by completion, are able to use discipline-specific practices flexibly to advance the learning of all students toward attainment of college- and career-readiness standards."*
ABET – 2018 - 2019 Criteria for accrediting applied and natural science programs	Criterion 3 - Student Outcomes	Baccalaureate degree students	*"An ability to formulate or design a system, process, or program to meet desired needs."*

Despite its popularity in science learning, incorporating engineering design into science classrooms can be a challenging task (i.e., design-science gap). Most research attention has been given to K-12 setting and limited effort has been allocated in the space of preservice teacher education. The following section presents the current design-science gap, what is engineering design and its relation to science learning and decision making, as well as how the proposed framework – scientific argumentation, could be helpful to prepare and train preservice teachers on argumentation and decision-making skills.

Design-science Gap

Incorporating engineering design into science learning has been shown to be helpful in promoting deep science learning, by providing relevant contexts to inquiry learning (Kolodner et al., 2003; Vattam & Kolodner, 2008). However, challenges related to this effort still exist. For example, not all teachers have the proficiency or training to execute it effectively, in part due to a lack of guidance in incorporating science understanding in design (Fortus, Dershimer, Krajcik, Marx, & Mamlok-Naaman, 2004), or not all students having the ability to relate their science knowledge to their design experiences (Kolodner et al., 2003). This disconnection between design experiences and science concepts is described as the "design-science gap" (Vattam & Kolodner, 2008).

Various efforts have been done in an attempt to bridge this gap, including different scaffold approaches such as contrasting cases (e.g., Chase, Malkiewich, & S. Kumar, 2019; C. Rebello, Beardmore, & Towle, 2015; Urueña, Rebello, Dasgupta, Magana, & Rebello, 2018), scaffold for constructing explanation embedded in design software (e.g., Sandoval, Crawford, Bienkowski, Hurst, & Millwood, 2003; Sandoval, 2003), reflection prompts (e.g., Berland et al., 2013; Fortus et al., 2004), "just-in-time" benchmark lessons (e.g., Kanter, 2010; Kolodner et al., 2003), just to name a few. However, there is still limited scaffolding methods that could seamlessly integrate the practice of design and science by helping students make the connection between their design experiences and science concepts (Apedoe & Schunn, 2013).

Engineering Design

Not only is engineering design important in engineering education, but it is also becoming increasingly important in science education, mainly because of the knowledge and skillsets that engineering design provides through design practices. For very similar reasons, the *Next Generation Science Standards* (NGSS) has integrated engineering design into K-12 science education (NGSS Lead States, 2013).

Before going far, it is important to spend some time looking into what makes engineering design different from the other types of design. Historically, before the rise of modern technology and mechanical industrialization, design was characterized as being ingrained in the craft of making (Mitcham & Holbrook, 2006), and that there was no separate process between "designing" and "making" (Dym, 1994). What it meant was that the designer of an artifact is also the maker of the artifact. For example, a carpenter would design the shape of the chair that he wants to make and actually makes it himself using suitable and available materials. If something doesn't go as intended along the way, the carpenter could always modify as he finishes the chair. The modification is fine because back in those days, the end user of an artifact is usually the maker himself.

However, with the rise of modern technology and industrialization, it sprouted mass production and mass consumption (Sparke, 1986). Due to that change, a separation between the process of "designing" and "making" became apparent. For example, in order to mass produce a specific kind of bicycle, the designer first needed to specify the specifications of the design such as materials, mechanics, functions, etc. (i.e., designing). Then, these specifications would be used to assemble the bicycle (i.e., making). This change is especially true in engineering design.

Continuing on that point, what makes engineering design different from other types of design (i.e., graphic design) is that the end product generated by engineering designer is not usually a piece of artifact. Instead, the engineering designer focuses on generating *"a set of fabrication specifications for that artifact"* (Dym, 1994, p.15). In other words, engineering designers mainly generate a detailed description of how an artifact can be put together or manufactured, and it is important for designers to make sure that the description is clear, complete, specific, makeable, and assessable. Because of this reason, engineering design is defined *as "the systematic, intelligent generation and evaluation of specifications for artifacts whose form and function achieve stated* *objectives and satisfy specific constraints"* (Dym, 1994, p.17). This definition of engineering design from Dym (1994) comes with the assumption that *"some sort of representation, formalism, or language is inherently and unavoidably involved in every part of the design process"* (Dym, 1994, p.20). This means that throughout the different stages of engineering design, the design artifact must be described, and that often is done in some form of representation. Other scholars defined engineering design as the process where engineers *"apply their scientific and engineering knowledge to solution of technical problems, and then to optimize those solutions within the requirements and constraints set by material, technological, economic, legal, environmental and human-related considerations"* (Pahl, Beitz, Feldhusen, & Harriman, 2007, p.1).

Types of Engineering Design

In general, engineering design can be categorized into three different types based on the knowledge needed as well as the complexity of the design problems. The different types of engineering design are (a) creative design, (b) variant design, and (c) routine design (Dym, 1994). Creative design or sometimes also known as original design (Pahl et al., 2007), is design that uses new solution principles to come up with new design products or solutions. This type of design is usually characterized by lack of both domain and problem solving knowledge (Dym, 1994; Pahl et al., 2007). On the other hand, variant design is design with fixed principles (Pahl et al., 2007), meaning, a variant design typically keeps the general principles of the original design while changing the dimension of particular parts (e.g., subsystem) to meet specific tasks. This type of design is usually characterized by rich domain knowledge but with the lack of knowledge to apply those known domain knowledge (Dym, 1994). Next, routine design is design characterized by both rich domain and problem solving knowledge (Dym, 1994). Meaning, a routine designer would usually have the ability to effectively apply the knowledge they have into a design.

Though suggestive, the classification of these engineering design types can be too limiting due to the subjectivity of some of the ideas used (e.g., routineness is subject to the experience, standards, or even the "brain" of the designer). Other efforts have been done to develop other engineering design taxonomies. However, they won't be discussed in detailed here due to its minimal relevance to this research project.

Relating Engineering Design to Decision-making

Decision-making is one of the most critical processes in engineering design (Crismond & Adams, 2012; McDowell et al., 2010; Ullman, 2001). Some even state that engineering design is primarily a decision-making process (McDowell et al., 2010). Early studies on information processing during design tasks showed that designers, other than using if-then rules, spend most of their time searching through the design space by comparing alternatives to design criteria. They do these comparisons by using available and relevant resources, and deciding the next steps based on those comparisons (Stauffer & Ullman, 1991). A decision in engineering design is sometimes referred as *"a commitment to use resources"* (Ullman, 2001). Ideally, this commitment should be made wisely and therefore, it makes sense to put heavy emphasis of decision-making during the process of engineering design.

In addition, decision is also seen as indicators to identify the progression of a design from the beginning to the end (McDowell et al., 2010). Ultimately, design decisions are made by human designers, potentially with the aid of computational or analysis tools, and not just the final products of some computer-ran analysis. While computational tools can be very helpful when it comes to decision-making, it is still very important for designers to possess the skills of knowing how to utilize data, information, or knowledge that they have to make informed design decisions. Therefore, it is helpful to take a closer look at informed decision-making in engineering design.

Decision-making and Engineering Decision-making

Making a decision usually means making a choice from a number of alternatives. Some perspectives describe decision making as a change of probability – set one option to the probability of 1 and the rest to 0, or a change of state – from "I am not sure" to actually making a decision (Hatamura, 2006).

Decision making is a very crucial processes in design. As a core process in solving complex and ill-structured problems, it is defined as an iterative process of acknowledging and managing tradeoffs, due to uncertainty, complexity, interrelatedness, and situatedness of problems (Jonassen, 2012). According to Jonassen (2012), decisions are central to human cognitive processing and problem solving, and the role of designer in decision making is to turn ideas into reality. Essentially, design is a process that involves iterative decision-making, with the designer reducing the complexities of decisions down to a point that fulfills the design need, based on constraints and biases (Jonassen, 2008). During this process, designers identify alternative options, formulate guidelines to lead the decision-making process, and gather and evaluate as much related information as possible to analyze the strengths and weaknesses of potential solutions before making the final design decision (Papadouris, 2012).

Types of Decisions

Based on literatures, there exists different types of decisions. Very broadly and simply, decisions can be classified into three types: (a) go or no go – only one choice available and the choice is whether to do it or not, (b) single selection – multiple choices available and the choice is to pick one, and (c) structured decision – multiple nodes (each node with multiple choices) are available and the choice is to pick one node that leads to a structured route (Hatamura, 2006). Hansen and Andreasen (2004) described a decision node as a basic decision-making activity containing of six sub-activities: (a) specify, (b) evaluate solution alternatives, (c) validate a design solution, (d) navigate through the solution space, (e) unify the decision into consistent whole, and (f) decide. In design, structured decision is most commonly seen because from beginning to end, design involves decision phases that include making many different choices (Crismond & Adams, 2012), and are always bounded by constraints.

In addition, decisions can also be classified into several other types such as (a) choices from a set of alternatives, (b) acceptance or rejection, (c) evaluation, or (d) construction – efforts to generate workable solutions based on available resources (Jonassen, 2012). Choices from a set of alternatives is similar to the aforementioned single selection (Hatamura, 2006) whereas acceptance or rejection is similar to go or no go (Hatamura, 2006). The judgements used in making these decisions can be based on different values that the designer find relevant, such as functional value, aesthetic value, originality value, or personal value and belief (Christensen & Ball, 2016).

Approaches to Decision-making

Understanding the types of thinking that go into decision-making helps to better make sense of the different approaches to decision-making. Based on literature, there are different types of thinking involved in decision making. Two common ones are (a) experiential mode of thinking and (b) analytical mode of thinking (Papadouris, 2012). Experiential mode of thinking is an automatic, fast, and effortless process that relies on heuristics and is susceptible to biases and fallacies; whereas analytical mode of thinking seeks to conform to rational norms and is often shaped by time constraints (Papadouris, 2012). According to Papadouris (2012), it is not uncommon for a designer to experience multiple modes of thinking when making decisions.

Based on these types of thinking, most typically, decision-making approaches can be classified into two types: (a) normative approaches and (b) naturalistic approaches. Jonassen (2012) described normative approaches (sometimes referred as rational approaches) as rational processes of making decision based on option that brings the most benefits when facing uncertainty, usually following a set of norms or rules, or using numerical values to help make the best decision (i.e., practicing analytical mode of thinking (Papadouris, 2012)). Key assumptions of normative decision making include knowing potential courses of actions and the likelihood of outcomes. The goal of normative decision making is to optimize, and usually the huge variety of options and time involved in mathematical calculation are not a huge concern (Zannier, Chiasson, & Maurer, 2007). Examples of normative approaches include rational choice, cost-benefit, and risk assessment. On the other hand, naturalistic approaches are usually based on emotions, previous experiences or personal identities (i.e., practicing experiential mode of thinking (Papadouris, 2012)). A naturalistic decision makers typically approach ill-defined problems with the purpose fulfilling constraints, as opposed to optimizing the solutions (Zannier et al., 2007). Examples of naturalistic approaches include narrative-based decision making and identity-based decision making. Normative approaches are commonly used for more structured problems whereas naturalistic approaches are more commonly used for less structured problems (Zannier et al., 2007).

Ideally, decision makers want to make decision that are as rational, optimal, and unbiased as possible. However, in reality, that is not always the case. Decision makers often settle with decision that satisfice, rather than optimize, and are subject to various types of biases (Jonassen, 2012). In addition, going after this kind of perfect utility-maximizing rationality (e.g., the theory of subjective expected utility (SEU)) do not deal with aspects outside of the decision itself such as setting goals and developing new alternatives (Simon et al., 1987).

Assessment of Decision-making

The skill or the ability for making informed decision is not easy to acquire. Therefore, various research efforts have been put into assessing decision-making to better understand the process, with the hope of improving decision-making skills (e.g., through instruction, etc.). Most commonly, decision-making is assessed for decision-making ability as well as the quality or effectiveness of decision (e.g., Almendra, 2019; Jonassen & Kim, 2009). Other constructs of interest in decision making relate to the approaches designers use to make decisions during the various stages of the design process and the types of information utilized by designers during decision making (e.g. (Dwarakanath & Wallace, 1995).

Besides, decision-making has also been assessed and characterized by decision activities – activities that *"capture what kinds of changes were made to a design state"* (Adams & Atman, 2000, p.4). Decision activities include monitoring changes to a design plan, problem scoping, and solution revision (Adams & Atman, 2000). In addition, other aspects such as the roles of intuition and emotions in decision-making have also been examined (e.g., Alexander, 1964; Lerner, Li, Valdesolo, & Kassam, 2015).

The assessment of decision-making can be very challenging given the complexities that go into the decision-making process (Jonassen, 2012). To really understand what goes into a decision made, it is not enough to look at just the information about the decision itself. Instead, it is important to record the thought process involved (e.g., what goes into the mind, what was wondered, what was tried, etc.) (Hatamura, 2006). That being said, the approaches used to assess decision-making need to be carefully considered and crafted to make sure they effectively capture the desired constructs to be measured.

Due to the complexity of decision-making, qualitative methods are usually the approaches of choice. For instance, descriptive research (i.e., to describe a phenomenon and its characteristics) is a common approach taken to examine design decision making (e.g., decision making in engineering design - Dwarakanath & Wallace, 1995). Qualitative methods such as think-aloud or verbal protocol analysis (e.g., Adams & Atman, 2000; Almendra, 2019) and collecting design sketches are commonly seen in this type of research (e.g., Dwarakanath & Wallace, 1995).

Case studies utilizing interviews and content analysis is another way of studying design decision making. For example, in a study aimed at constructing a model of design decision-making, specifically to provide answer to the ways software designers construct design decisions, Zannier et al. (2007) interviewed 25 software designers about the design decisions made by them, and use content analysis as well as cross-case analysis to analyze their data. Specifically, they comprehensively coded and analyzed their transcripts using Small Window Code (SWC), Large Window Code (LWC), Generic Relational Coding (GRC), Specific Relational Coding (SRC), Case Study Summary, and Cross Case Comparison (Zannier et al., 2007).

Moreover, argumentation (i.e., decision makers write statements of arguments supported by evidences to defend a decision) is claimed to be one of the most valid and reliable ways to access decisions (Jonassen, 2012). One way to access decision-making arguments is the use of rubrics that examine the quality of decision, adequacy and credibility of premises, presence and quality of counterargument, as well as organization of arguments (Jonassen, 2012).

Decision making can be done both individually and in group. Therefore, it is not uncommon to see research efforts in trying to understand group decisions using qualitative methods. For example, in a study aimed at finding out how groups of children used evidence in decision making as well as the patterns of argumentation they used, Maloney (2007) engaged students in discussions and debates, using activities that were relevant to students, accompanied with evidence in different formats, as well as alternative choices. Students' discussions were recorded in the form of videos and were transcribed as part of data analysis. Using three different frameworks – Toulmin's (1958) framework of argumentation, Andrews' (1995) sequence of stages in forming arguments, and Belbin's (1981) team roles, Maloney (2007) explained the characteristic of children's roles and how those roles impacted the ways they used evidence to make decisions.

Identifying decisions is central to most assessments of decision-making. However, extracting and coding decisions is not an easy process because decision-making is not usually explicitly stated by the designer. Therefore, especially during the analysis of decision, it can be helpful to split decision into two types, (a) explicit decisions – when designer explicitly say or write down a decision made, and (b) implicit decisions – when designer did not explicitly say or write anything but a decision can be identified retrospectively based on his actions or the final design (Dwarakanath & Wallace, 1995).

Challenges of Decision-making

Decision making is a complicated and challenging process, especially for beginning designers. Research reveals that experienced or informed designers do research, conduct deep modeling, balance trade-offs, perform valid experiments, carry out diagnostic troubleshooting, and execute reflective design thinking when it comes to making design decisions (Crismond & Adams, 2012). These skills are hard for beginning designers because they usually *"ignore or pay too little attention to design criteria and constraints, and focus only on positive and negative aspects of their design ideas without thinking of associated benefits and trade-offs"* (Crismond & Adams, 2012, p.24).

In addition, beginning designers usually rely on shortcuts that may lead to predictable mistakes and misconceptions, struggle to compare their solutions effectively and consistently, avoid holistic methods that systematically weigh trade-offs, as well as apply decision-making approaches inconsistently – meaning, they tend to address tasks at hand but not always apply a consistent strategy (Papadouris, 2012). Moreover, some beginning designers struggle with connecting the design problems they are solving with the underlying science principles (Chao et al., 2017), as well as arguing or articulating the rationales for the different alternatives that they are considering (Crismond & Adams, 2012). It is also not unusual to see students copying design ideas either from peers or the Internet instead of turning to scientific principles and practices (Chao et al., 2017).

Teaching Decision-making

Regardless of how challenging decision-making can be, it is suggested that decision making is teachable with appropriate scaffolding (Christensen & Ball, 2016). Crismond and Adams (2012) suggested that teachers can scaffold students' decision-making, especially about analyzing and balancing pros and cons of different design alternatives, by consistently prompting students to explain and justify their design decisions, using relevant science and engineering principles, supported with evidence (e.g., computational or analytical evidence).

Along the same line, it is important to help students fluently communicate the logic and reasoning underlying the strategies used, and explain and correct potential misconceptions, as well as provide practice environment where they can combine strategies, using inductive, deductive, and abductive reasoning (Christensen & Ball, 2016). Other recommended teaching strategies include teaching and helping students to use decision diagrams (e.g., Chooser Chart and The House of Quality Diagram) so that they can visually represent their ideas and alternatives in order to further analyze and weigh them based on priorities and relevant values (Crismond & Adams, 2012).

Relating Design Decision-making to Argumentation

Argumentation plays a crucial role in decision making (Patronis & Spiliotopoulou, 1999). In design, designers often need to make decisions to solve complex issues. The complex nature of design problems usually cannot be resolved simply with straightforward answers (Jiménez-Aleixandre & Pereiro-Muñoz, 2002), and therefore, requires the proposed solutions to be based on argumentation for it to be fruitful and valuable (Patronis & Spiliotopoulou, 1999).

In decision-making, argumentation can be seen as the process of evaluating theoretical claims based on empirical evidence or data from relevant sources (Kuhn, 1993). In other words, it is about the ability to choose and reason between different alternatives or explanations which ultimately lead to making a choice (Jiménez-Aleixandre & Pereiro-Muñoz, 2002). To further elaborate, the process of argumentation in decision-making starts with the decision maker identifying claims (options), followed by identifying alternative claims (other options), and examining arguments for and against each claim to make a final decision, based on relevant evidence (Jonassen, 2012).

As briefly described in the previous section, there are various ways of teaching and supporting decision-making. Argumentation is one of the ways that can support decision-making because argumentation allows people to settle differences among competing alternatives (Hogarth & Kunreuther, 1995), by generating arguments for and against each options based on their knowledge, instead of only relying on probabilities or weights (i.e., decision matrix aka Pugh method). Not only is being able to reason about the trade-offs and benefits as well as different alternatives is imperative to design decision making (Crismond & Adams, 2012), being able to think metacognitively while making arguments is also important to design decision-making in terms of better design performance and design artifact quality (Adams & Atman, 2000; Crismond & Adams, 2012).

Argumentation has also been shown to be useful in making decision to address problems that are both well-structured and ill-structured (Jonassen & Kim, 2009). To emphasize, argumentation skill is especially important in solving problems that are ill-structured because such problems lack convergent answers and require the designers to use arguments to justify their decisions (Cho & Jonassen, 2002). For example, evidence from a study done by Cho & Jonassen (2002) showed that students who solved ill-structure problems generated more extensive arguments.

Based on literature, there are two kinds of arguments that are common to design decision-making, which are (a) rhetorical arguments and (b) dialectic arguments (Jonassen & Kim, 2009). The purpose of rhetorical arguments is persuasion, that is to convince others to accept a claim or position regardless of the positions of others, whereas the goal of dialectic arguments is resolution of differences in opinions (Jonassen & Kim, 2009). The key difference between these two types of arguments is the presence of counterarguments and rebuttals in the dialectic arguments (Jonassen & Cho, 2011). Most design problems (i.e., engineering ethics) have no one fixed solution, instead can be interpreted and supported by multiple perspectives, and therefore, they are usually more dialectic in nature (Jonassen & Cho, 2011).

Practicing argumentation in the process of design decision-making is still a challenge for many students, and therefore, continuous efforts have been done, especially in the context of STEM classrooms to help students develop argumentation skills, so that they can make better informed decisions (e.g., Cho & Jonassen, 2002; Crismond & Adams, 2012; Jiménez-Aleixandre & Pereiro-Muñoz, 2002; Jonassen & Cho, 2011). For instance, based on an environmental science study done by Jiménez-Aleixandre and Pereiro-Muñoz (2002), the authors suggested that students do not only make their decisions solely based on scientific knowledge and evidence, but also on value judgements – meaning, accessing something as good or bad in terms of one's standards or priorities. Regardless, it is still important for students to provide justifications to support their decisions.

In another study done by Jonassen and Cho (2011) on promoting argumentation in the context of solving engineering ethics problems, the authors found that students constructed better counterarguments to arguments generated by someone else, and are better at identifying and rebutting alternative solutions when they generated their own counterarguments, instead of when they were given examples of counterarguments. It is implied that students engage and argue better when they are engaged in generative learning activities where they can take ownership of their own works (Jonassen & Cho, 2011).

Current research and efforts are helpful in studying and promoting argumentation in design decision-making, however, there is still a lack of efforts in examining the differences between argumentation in the context of science and argumentation in the context of design, as well as the relationship between the two. Specifically, how does the way students argue influence their scientific conceptual understanding, and how does that understanding consequently influence the way students argue and make design decisions. This is important because in design, it is important for designers to make knowledge-driven decisions – meaning, making decisions based on scientific knowledge and insights learned from experiments, instead of purely from common sense or informal knowledge (Crismond & Adams, 2012). That being said, further research on this area is worth exploring.

CHAPTER 3: THEORETICAL FRAMEWORK

Argumentation

Very broadly, argumentation can be described as the systematic process of supporting a theory or action by reasoning. One of the earliest and seminal contributions made in argumentation was by Toulmin (1958). A model of argumentation was presented by Toulmin to describe the elements of argumentation and the relationship between them. The major elements of Toulmin's model are (a) claim – conclusion that are to be proven, (b) data – facts used to back up a claim, (c) warrants – reasons used to explain and connect the data and the claim, (d) backing – assumptions involved to justify warrants, (e) qualifiers – conditions specifiers for the claim to be considered as correct and for representation of constraints on the claim, and (f) rebuttals – conditions specifiers for when the claim will not be considered as correct.

Argumentation versus Explanation

Based on literature, it seems like the word "argumentation" or "argument" and "explanation" are sometimes used interchangeably or conflated. An example of this conflation can be seen in science education. For example, in a work where McNeill and Krajcik (2008) characterized and evaluated the impacts of teachers' pedagogical practices on student learning, they intentionally used the word "explanation" even though their work built on literature from both explanation and argumentation. The authors later on characterized explanation with three components – *"a claim (a conclusion about a problem); evidence (data that supports the claim); and reasoning (a justification, built from scientific principles, for why the evidence supports the claim)"* (McNeill & Krajcik, 2008, p.55), these components are part of the argument elements described by Toulmin (1958). Even though the authors did mention the adaptation of Toulmin's work, this conflation still makes it challenging for readers to fully understand these constructs. It was pointed out that the lack of clarity around the meaning of these words is a weakness in the field of research, especially in education (Osborne & Patterson, 2011). Hence, it is important to spend some time distinguishing argumentation and explanation.

In the world of science education, an explanation is used to understand a phenomena based on other scientific facts whereas an argument is used to justify a claim (Osborne & Patterson, 2011). A key characteristic of an explanation is that the targeted phenomenon is not open to question. On the other hand, a key characteristic of an argument is that there is typically a degree of tentativeness and uncertainty. However, explanation and argument are interconnected in the way that an explanation attempts to provide reasons or explanations for a given phenomenon whereas an argument examines whether or not the explanation is valid – meaning, whether it generates understanding successfully or whether it is better than competing explanations. To put it simply, an explanation is to answer "Why?" or to increase a sense of understanding, whereas an argument is to persuade or to justify a claim (Osborne & Patterson, 2011). In other words, *"an argument well-established premises are used to support a less-than-certain conclusion, whereas in an explanation a well-established fact is accounted for by a less-than-certain explanation"* (Osborne & Patterson, 2011, p.634).

With a clear difference made between explanation and argument, it is clear to see the importance of the process of argument. This is because at the very beginning, an explanation usually starts as a hypothesis that could potentially provide an explanation to a phenomenon. This kind of hypothesis needs to be analyzed and survive rounds of different arguments to demonstrate its incorrectness in order to become an explanation that is consensually and commonly accepted (Longino, 1990). For this reason, argumentation becomes a skills that is important for fostering a discussion that involves constructing and critiquing in order to justify beliefs (Ford, 2008). Hence, it is crucial to distinguish explanation and argumentation so that the goal of enhancing argumentation skill can be made clearer.

Types of Arguments

With the definition of argumentation made clear, it makes sense to look more closely at the different types of arguments. Based on literature, arguments can generally be categorized into three different types: (a) apodictic arguments, (b) rhetorical arguments, and (c) dialectical arguments (Jonassen & Kim, 2009). The purpose of apodictic arguments is to express absolute and reliable knowledge, leaving no doubt for the truthfulness of a claim. On the other hand, the purpose of rhetorical arguments is to persuade or convince others of a claim or position regardless of the positions of others, whereas the purpose of dialectic arguments is to settle conflict of opinions. Due to the naturalistic nature of everyday questions and problems that human are trying to solve, rhetorical arguments and dialectic arguments are most commonly seen and used (Jonassen & Kim, 2009).

In education, it is claimed by some that dialectic arguments is more applicable than rhetorical arguments because it allows argumentative function such as making a decision to happen within groups setting (Driver, Newton, & Osborne, 2000). Furthermore, presumptive arguments is a type of dialectic arguments that allow the advancement of arguments by proving or disproving it (Walton, 2014). Due to the nature of dialectic argumentation, counterarguments are usually given equal importance as the original arguments.

Some other ways arguments are categorized include a stasis approached taken by Fahnestock and Secor (2003) where arguments were studied and classified based on the issues they address. The results include four types of arguments: (a) Definitional arguments – "what is this?", (b) Causal arguments – "How did it get this way?", (c) Evaluation arguments – "Is it good or bad?", and (d) Proposal arguments – "What should we do about it?" Other approaches such as empirical approach has been taken by cognitive psychologists to study argumentation (Wolfe, 2011). An argument is defined as a claim backed up by a reason (Wolfe, Britt, & Butler, 2009), similar to "data" in the Toulmin's (1958) model. Reason can be facts, fallacies, or other types of arguments and that it is associated to the claim with a warrant.

Argumentation in Different STEM Education Domains

In education, argumentation is often associated with meaning making and inquiry learning (Jonassen & Kim, 2009). Teaching students to participate in argumentation, for example in science, is to train their scientific thinking like those of scientists (Kuhn, 1993). By learning how to argue for the claims they made, students get to engage in learning that is deeper and more epistemologically mature (Newton, Driver, & Osborne, 1999). Even though science educators have given more focus on the roles of argumentation than other disciplines, it does not mean that science is the only domain or discipline that benefits from argumentation. In fact, argumentation is seen as an crucial way of thinking about any discipline (Jonassen & Kim, 2009). That is, argumentation *"stresses the evidence-based justification of knowledge claims and it underpins reasoning across STEM domains"* (Erduran, Ozdem, & Park, 2015, p.1). Briefly presented below the general roles that argumentation plays in different STEM education domains.

Science Education

Scientific practices related to argumentation have been made the focus of science education. An example include the practices of arguing from evidence as well as getting, analyzing, and conveying information from the Framework for K-12 Science Education (NRC, 2012). Many science education scholars also argue that argumentation plays an important role in scientific thinking as scientists usually take part in argumentation to communicate and refine their science knowledge (Driver et al., 2000; Jonassen & Kim, 2009; Kuhn, 1993). Therefore, it is not surprising to see the growth in frequency of using argumentation as a pedagogy approach in classroom, especially science classroom.

In science, argumentation is commonly practiced by scientists in interpreting experimental results or in writing to convince the scientific community to accept their work (Kuhn, 1993). In other words, argumentation is often related to knowledge justification and persuasion. That is, it is defined as scientific arguments referred in evidence, and descriptions that work by persuasion (Erduran & Jiménez-Aleixandre, 2007). An argument connects claims to data via justification or the assessment of knowledge based on evidence, be it empirical or theoretical. That being said, arguments are different than pure opinions.

Various studies have been done to examine the relationship between argumentation and science learning. For example, it was found that argument-centered lessons impact students positively in the ways they elicit their previous knowledge and ideas at higher level of abstraction (Von Aufschnaiter, Erduran, Osborne, & Simon, 2008). Even though Von Aufschnaiter et al., (2008) argued that argumentation do not seem to help students directly in terms of developing new understanding. However, it does allow students to continuously consolidate their ideas and use similar ideas in different circumstances, which eventually leads to faster advancement of and utilization of ideas across contexts. Other work have also been done to develop framework that helps students to learn ways to develop and evaluate a scientific argument (e.g., Mcneill & Krajcik, 2008; Sampson, Enderle, & Grooms, 2013).

In general, argumentation appeals to the scientific community because of its high potential of supporting (a) the understanding of cognitive and metacognitive processes featuring expert performance and allowing modelling for students, (b) the development of communicative skills, (c) the accomplishment of scientific literacy (d) the gradual acquisition into the scientific practices, and (e) the development of reasoning, especially in reasoning using logical and rational criteria (Erduran & Jimenez-Aleixandre, 2007).

Technology and Engineering Education

Even though research about argumentation in technology and engineering education is less extensive that those in other areas such as science education, it is undeniable that engineers need to make evidence-based decisions (Van Epps, 2013). In other words, engineers use arguments to find the best solutions to solve problems within given constraints. In the engineering world, argumentation is sometimes referred to as evidence-based reasoning (e.g., Mathis, Siverling, Glancy, & Moore, 2017) or design rationale (e.g., Fischer, Lemke, McCall, & Morch, 1991).

Argumentation is seen as an imperative part of decision making in engineering (Patronis & Spiliotopoulou, 1999). Specifically, argumentation is described as the process of evaluating design alternatives based on empirical evidence from relevant sources (Kuhn, 1993). The ill-structredness of most engineering problems makes argumentation a skill that is important to be fostered into engineering education (Jonassen & Kim, 2009).

Increasing efforts have been done to incorporate argumentation into engineering curriculum. Though challenging, some instructional strategies were found useful in promoting the practice of argumentation, which included situating students in client-driven problems where they have to communicate and justify their ideas (e.g., client proposal) as well as asking "why" questions instead of "what" and "how" (Mathis et al., 2017). In addition, efforts on using engineering design to facilitate science learning is also on the rise. However, students are likely to apply trial and error approach when it comes to solving those design tasks. Because of that, argumentation has been used as a scaffold to guide students through this process, with the purpose of promoting informed decision-making as well as improving conceptual knowledge (e.g., Rebello, Barrow, & Rebello, 2014; Rebello & Rebello, 2013; Urueña, Rebello, Dasgupta, Magana, & Rebello, 2018).

Mathematics Education

For a long time in math education, "drill and practice" is the most commonly known practice. However, the main objective that math teachers are striving for nowadays is to foster math reasoning and understanding (Littleton & Howe, 2009). Even though there is no common or formal definition for argumentation in math education, argumentation is seen as a way to develop meaning-making and understanding in a math classroom (Littleton & Howe, 2009).

Using an argumentation perspective, Yackel (2002) conducted a study in investigating the roles of the teacher in math classroom. It was found that the math teacher (a) initiated the negotiation of classroom norms, (b) fostered argumentation as the core of students' math activity, (c) provided support for students to interact with peers in constructing collective arguments, and (d) provided argumentative supports such as data and warrants to enhance the argumentation process (Yackel, 2002). Based on these findings, the role and importance of argumentation in math education is emphasized.

In addition, argumentation is also used as a theoretical framework to describe the interaction and "sociomathematical norms" of a math classroom, identified through the kinds of justification and argumentation used – how students agree on whether a solution is acceptable, different, or convincing, as well as the roles of the math teacher (Yackel & Cobb, 2006).

One of the most important elements in math education is the practice of proving (Littleton & Howe, 2009). Proofs are generally used to verify mathematical statements and show their universality. According to Littleton and Howe (2009), there are two types of proof in math: (a) proofs that show a theorem is true – provide evidential reasons only, and (b) proofs that explain why the theorem is true – provide a set of reasons to convince. The latter is important in math classroom to promote understanding and therefore, students need to learn how to prove meaningfully, or in other words, to argue meaningfully.

Assessment of Argumentation

Multiple approaches have been taken to access argumentation and protocol analysis of students' responses to questions or arguments is one of the most commonly used methods for accessing argumentation (Jonassen & Kim, 2009). One of the most commonly assessed aspects of argumentation is the role of argumentation in fostering conceptual change (Jonassen & Kim, 2009). Previous research show that embedding argumentation in science learning environment leads to enhancement in conceptual understanding by making scientific reasoning visible (Jonassen & Kim, 2009; Von Aufschnaiter et al., 2008).

Another construct that is commonly assessed is argumentation skills – the ability to develop and assess various forms of argumentation. For instance, Blair and Johnson (1987) recognized three criteria that must be fulfilled by an argument: (a) relevance – "do premises match conclusion?", (b) sufficiency – "are there sufficient evidence?", and (c) acceptability – "are premises acceptable?". In addition, Kuhn (1993) identified and described the essential abilities needed to make strong arguments: (a) the ability to develop relevant theories to support claims, (b) the ability to provide evidence to back up theories, (c) the ability to construct alternative theories, (d) the ability to make counterarguments, (e) the ability to refute alternative theories.

In addition, the types of STEM related arguments have also been assessed (e.g., Wolfe, 2011). Some of the results are as the following. Empirical arguments are arguments resulted from the collection and analysis of data, with the attempt to support claims by incorporating empirical data into reasons (Wolfe, 2011). These arguments usually involve casual relationships. Decision-based arguments are arguments used to support a verdict or decision (Wolfe, 2011). The decision is usually incorporated into the claim, and the reasons backing up the claim has the potential of either agree or disagree with the selected decision, when comparing with other alternatives. Proposal arguments are arguments used to persuade something to be granted or permitted (Wolfe, 2011). One example of this is research grant proposals, where the goal is to persuade grant giver that a research is worth the funding.

Quality of argumentation is another aspect of argumentation that is commonly measured. It is inevitable to see the roots of Toulmin's (1958) work in these measurement. Some researchers measured and classified arguments into five levels – a hierarchy of increasing quality based on reasons and grounds, as well as rebuttals (Erduran, Simon, & Osborne, 2004; Von Aufschnaiter, Erduran, Osborne, & Simon, 2008). Other researchers accessed arguments using a three-level rubric, with increasing quality based on the elements of claim, evidence, and reasoning (McNeil & Krajcik, 2008; Krajcik & Merritt, 2012).

Other way argumentation has been characterized was through the various dimensions of learning progression – instructional context, argumentative product, and argumentative process (Berland & McNeill, 2010). Argumentative product and argumentative process took the assessment of argumentation a step further from the aforementioned five-levels or three-levels assessments, by including the appropriateness and sufficiency of claims and information to defend those claims, as well as the complexity of argumentative functions (e.g., state, defend, evaluate, question).

Moreover, the functions of argumentation outside of science field have also been investigated and characterized. Mathis, Siverling, Glancy, and Moore (2017) examined the ways argumentation can support engineering in STEM integration units, using the Framework for Quality K-12 Engineering Education (Moore et al., 2014), and found that argumentation has great potential in facilitating the process of design, application of science, engineering, and math, engineering thinking, as well as communication in engineering.

Other relevant assessment methods include using a framework developed in TAPping into argumentation: Developments in the application of Toulmin's argument pattern for studying science discourse (Sibel Erduran, Simon, & Osborne, 2004) as well as an adapted Evidence-Based Reasoning (EBR) framework by Furtak and colleagues (Furtak, Hardy, Beinbrech, Shavelson, & Shemwell, 2010) to measure the assessment of variables for argumentative structure, quality of backing, as well as teacher contribution.

Analyzing argumentation using Toulmin-based coding schemes can sometimes be challenging because it can be difficult for coders to establish what counts as a claim, premise, warrant, or backing (e.g., what looks like a claim to a coder might look like a premise to another coder) (Brown, Furtak, Timms, Nagashima, & Wilson, 2010). An attempt has been done to overcome this challenge by having the coders write a storyline for each classroom discourse prior to the actual coding to established an agreed-upon interpretation of statements (Furtak et al., 2010).

Challenges in Argumentation

Argumentation is not an easy skill to acquire. Some of the major challenges faced by students in argumentation include the lack of understanding in argumentative discourse, struggles in differentiating between claim and evidence, failures to counterargue, and "my-side bias" – inclination to support arguments based on personal beliefs rather than confirming or disconfirming evidence (Jonassen & Kim, 2009).

In addition, students also struggle with problems with validity, naïve conception of argument structure, inappropriate sampling of evidence, and alternation of representation of argument and evidence (Zeidler, 1997). Other relevant challenges include confirmation bias (i.e., only look for data that support their ideas), generalization based on limited information, failure to justify evidence, as well as hesitance to engage in argumentative discussion (e.g., culturally consider questioning others' ideas as disrespectful) (Sampson et al., 2013).

Teaching Argumentation

Some of the reasons that contribute to the challenges faced by students in argumentation include the lack of pedagogical skills from teachers to promote argumentation in classroom, and even if there are opportunities to develop those skills, they are usually under external pressure to cover others materials, not allowing sufficient time for professional development (Driver et al., 2000; Newton et al., 1999; Von Aufschnaiter et al., 2008).

Argumentation is a way that allows people to rationally address questions and tackle problems that occur daily life and therefore, it is important to embed and foster argumentative activities in learning environments (e.g., project-based learning, learning by design, etc.) so that it can promote fruitful ways of thinking, conceptual change, as well as problem solving (Jonassen & Kim, 2009). If students are immersed in activities where they only need to memorize information, there is no need for argumentation and therefore, it is very unlikely for them to gain argumentation skills.

Various approaches have been developed to support the teaching and learning of argumentation. One of the most direct approaches is to provide clear directions for students to construct arguments, including counterarguments in classroom activities (Nussbaum & Kardash, 2005). In a study, Nussbaum and Kardash (2005) found that explicit direction for students to argue in different ways improved students' performance in argumentation. Apart from providing clear direction and instruction, it is as important to clearly explain to students what it means to provide arguments as well as to provide examples of how to do it. Using an argumentation framework (i.e., argument consists of claim, evidence, and reasoning), Mcneill and Krajcik (2008) suggested five strategies to help students in the argumentation process, which include (a) explicitly explain the framework, (b) demonstrate and evaluate explanations, (c) provide a reason for making explanations, (d) relate to everyday explanations, and (e) evaluate and give feedback to students.

Often, it can be useful to scaffold argumentation by helping students to visualize arguments using graphical tools (Jonassen & Kim, 2009). Relating to providing clear direction for arguments, such direction can potentially be presented visually using graphic organizer (e.g., Nussbaum & Schraw, 2007). Besides, it is also crucial for teachers to scaffold argumentation by consistently prompting and asking questions. Some of the useful strategies in doing this include the (a) refutation strategy – "what alternative might someone else recommend?", (b) synthesizing strategy – "is there a different or out-of-the-box alternative?" , and (c) weighing strategy – "which alternative is better and why?" (Nussbaum & Schraw, 2007). Research shows that by prompting students to participate in the assessment of alternative arguments, students are able to more elaborately discuss and justify their solutions (Jonassen & Cho, 2011). These practice should not only be done on an individual level, but in group setting as well, as it is found that when engaged in collaborative argumentation, students are able to generate more arguments as well as construct and apply argument to new contexts (Jonassen & Kim, 2009).

It can also be beneficial to engage students in argumentative discourse guided by a set framework (Driver, Newton & Osborne, 2000). Argumentation is a process that helps students to think like a scientist or engineer by enabling process-oriented practice of constructing and justifying claims (Jonassen & Kim, 2010). Prior studies have shown that argumentation can enhance conceptual understanding, investigational capability, and problem solving in science (Driver, Newton & Osborne, 2000, Nussbaum & Schraw, 2007). That being said, the efforts of developing students' knowledge through argumentation must consider the significance of students' prior knowledge and the complexity of the tasks (Von Aufschnaiter et al., 2008) as students advance their understanding based on what is situated in their relevant zone of proximal development (Vygotsky, 1962).

CHAPTER 4: LEARNING DESIGN

Context

For the purpose of this research study, a four-week unit was designed and implemented in a Physics for Elementary Education course in a Midwestern University in Indiana, USA. The context of this unit focused on the knowledge of physical science, specifically heat transfer, as well as the practices of science and engineering design. The implementation of this unit involved one Physics professor, one graduate teaching assistant, as well as two professors and a graduate student who helped in the design and implementation of the curriculum.

Leaner's Characteristic

The target learners of this unit were pre-service teachers who were preparing to be future elementary teachers. These learners were expected to come with fundamental and prior knowledge on the topic of heat transfer.

Learning Objectives

As discussed in the previous chapters, students in general, including pre-service teachers, struggle with having deep understanding of abstract and difficult physical science concepts such as heat transfer. In addition, they also struggle to perform effective engineering design practices, such as providing effective scientific argumentation to support or justify their experiment outcomes, as well as making informed design decisions based on their experiment outcomes. These struggles often result in challenges in gaining deep conceptual understanding of the relevant physical science concepts.

To address these challenges, this unit aimed to help students develop deeper understanding of physical science concepts, as well as gain and improve on the practices of science and engineering design. Specifically, the learning objectives of this unit were students will be able to: (a) experiment with and develop deeper conceptual understanding of heat transfer processes, which include conduction, convection, and radiation, and (b) engage in and learn science and engineering practices such as design skill, experimentation skill, argumentation skill, and informed decision-making skill. Table 2 below shows the mapping between the difficult concepts that we would like to help students overcome and the learning objectives to help us achieve that.

Table 2. Mapping between difficult concepts and learning objectives.

Difficult concepts	Learning Objectives
Heat transfer	Students will be able to experiment with and develop deeper conceptual understanding of heat transfer processes, which include conduction, convection, and radiation.
Effective science and engineering practices	Students will be able to engage in and learn science and engineering practices such as design skill, experimentation skill, argumentation skill, and informed decision-making skill.

The success in acquiring these learning objectives could be evidenced by students' improved understanding of knowledge on heat transfer, their ability to argue scientifically by making informed design decisions. That is, by making the right claim, justifying it scientifically by citing specific evidence, and writing explanatory texts to examine ideas through appropriate selection, organization, and analysis of information collected from experiments, simulations, or multimedia sources students would demonstrate their understanding of heat transfer and their ability to make informed design decisions. The collection of these evidences would be discussed more in the upcoming assessment section.

Learning Theory

The learning theory that is guiding the design of this unit is constructivism. As learned from literature, there are two constructivist's views on learning: individual and social. According to the individual constructivist's view, human construct knowledge and form meaning based on their own experiences (Wadsworth, 1996). New schemes are formed by modifying old ones through adaptation to more complex experience based on the internalization of a person's action on realities in the world (Driver R., 1994). On the other hand, social constructivist's view recognizes the role of cultural community in learning. As described by Bruner on Vygotsky's work, *"that world is a symbolic world in the sense that it consists of conceptually organized, rule bound belief systems about what exists, about how to get to goals, about what is to be valued.*

There is no way, none, in which a human being could possibly master that world without the aid and assistance of others for, in fact, that world is others" (Bruner, 1985, p.32). Therefore, learning involves a dialogic process where people interact. For example, a more experienced person can support the learning of a less experienced person by organizing tasks, guiding the less experienced person to carry out and internalize the process (Driver R., 1994).

Both individual and social constructivist's views play influential roles in learning. In science learning, constructivists believe that students construct their own knowledge influenced strongly by social environments and therefore, meaning making involves both personal and social processes (Driver R., 1994). Constructivism emphasizes on a few key elements that are crucial for conceptual change to happen in science education. These elements are: (a) the significance of prior knowledge, (b) the significance of students' activities, (c) the significance of contextualization, (d) the significance of collaboration within learning community, and (e) the significance of teacher's role in a classroom.

Students usually come to a science classroom with strongly held personal views from previous experience and it is crucial for teacher to understand their prior knowledge because most of the time, students will build new knowledge based on their prior knowledge (Leach & Scott, 2003). Students are more likely to develop deep understanding if they are given the chance to construct their own knowledge. That being said, they need to be engaged in activities that allow them to "explain, muster evidence, find examples, generalize, apply concepts, analogize, and represent in a new way." (Perkins, 1993, p.29). In addition, students' activities need to be appropriately contextualized as it promotes higher motivation and encourages the transfer of knowledge (Perkins, 1993). Contextualization means bringing in real-life situation and connecting science to students' personal lives. Besides, students would benefit from learning in a collaborative community where they learn from one another. These interactions promote the use of appropriate language to express and justify students' own conceptions, clarify and reflect on their peers' perspectives, and negotiate new, joined meanings (Kearney, 2004). However, these interactions wouldn't be possible without the guidance of teacher in the classroom. Hence, teacher plays a crucial role to mediate scientific knowledge for students, to assist them to make personal meaning and construct knowledge using formal scientific discourse (Driver R., 1994). Therefore in this study, we would like to use constructivism and its important elements to guide us through the design of the whole learning experience.

Pedagogical Approach

The pedagogical approach used in this unit is a modified engineering design cycle incorporated into an 3E's inquiry learning cycle (C. M. Rebello, 2019). In general, this unit followed this 3E's pedagogical structure: (a) Explore – students learn and gain experience through their actions and reactions in a new environment, (b) Explain – students learn new concepts or principles that help them reason their unique experiences, and (c) Elaborate – students ally the newly discovered concepts and reasoning to new situations (Karplus & Butts, 1977; C. M. Rebello, 2019). Figure 1 below depicts the process of this modified engineering design/inquiry learning cycle. This learning cycle was enacted in this study as follows: (a) Explore – students were provided an opportunity in the lab to explore science ideas through conducting physical experiments, (b) Explain – students were provided with an opportunity to discuss and clarify scientific ideas with instructors and peers, and (c) Elaborate – students were provided with an opportunity to apply science ideas learned to design through working on a design challenge.

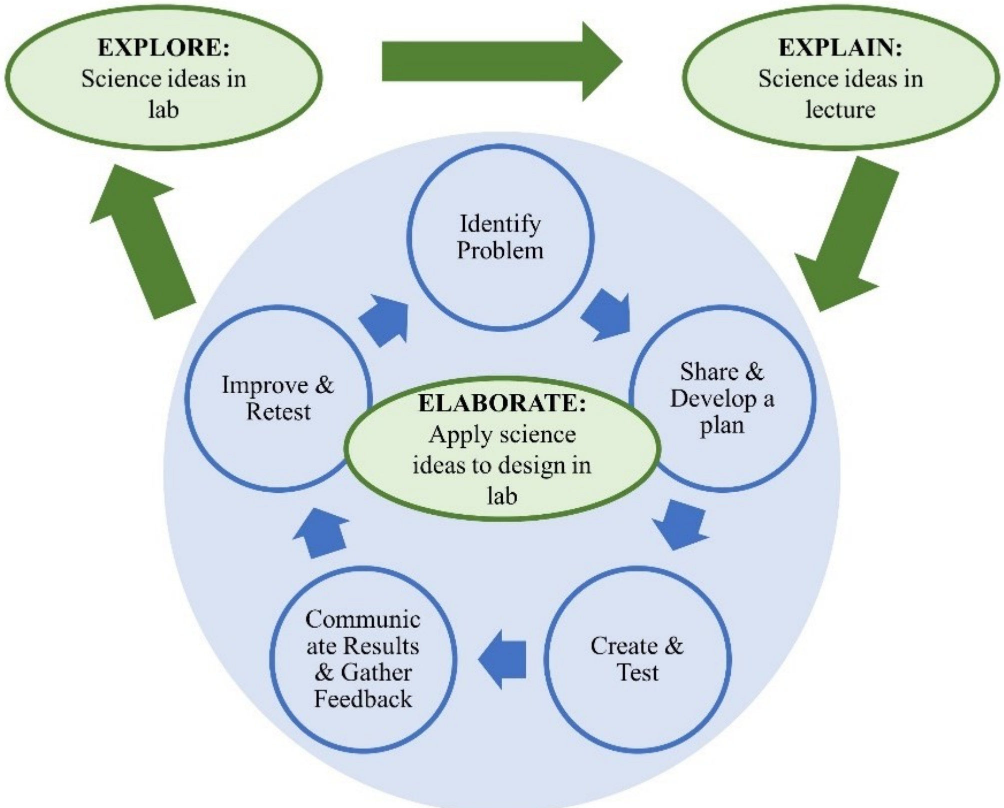


Figure 1. The adapted modified engineering design cycle incorporated into an 3E's inquiry learning cycle (Karplus & Butts, 1977; C. M. Rebello, 2019)

This modified engineering design cycle/inquiry learning cycle is inspired by the approach of Learning by Design ™ (LBD) (Kolodner et al., 2009). Grounded in cased-based reasoning and problem-based learning, LBD is a form of project-based inquiry approach to science where students learn science in the context of attempting to complete a design challenge. Unlike traditional approach where students memorize facts and formulas, LBD allows students to engage in the process of designing and conducting experiments, evaluating data and generating conclusions, constructing informed decisions and backing them up with evidence, working in teams, and communicating. Some of the LBD principles that were incorporated into the modified engineering design/inquiry learning cycle are to (a) give students opportunity to solve a design challenge using their existing knowledge, (b) encourage instructors to help students compare and contrast ideas, determine what they need to learn, and choose idea to focus on, (c) allow students to go through a process of exploratory and experimental work to generate potential solutions, (d) allow students to build, test, analyze, and revise models to optimize their solutions, and (e) allow students to share their work for feedback as well as to justify their design decisions.

We approached this unit with this pedagogical approach because by working on a relatable and realistic design challenge, students would be given the opportunity to understand and develop new knowledge based on their personal experiences in interacting with various variables (Piaget, 1937). In addition, with these approaches, instructors get to actively play the role of an experienced person to support the learning of a less experienced person (i.e. students) by guiding them to carry out and internalize the process (Driver et al., 1994). Besides, classroom interaction such as feedback or presentation promote the use of appropriate language to express and justify students' own conceptions, clarify and reflect on their peers' perspectives, and negotiate new, collaborated meanings (Kearney, 2004).

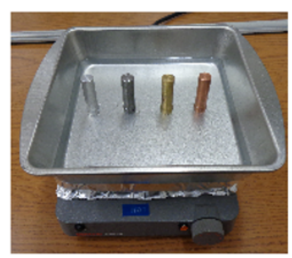
In addition to the aforementioned approaches, to help students make explicit informed design decisions by connecting them to their science knowledge, we implemented a scientific argumentation framework as a scaffolding approach. The selected argumentation framework is the Claim – Evidence – Reasoning framework by McNeil and Krajcik (2011). When applied effectively, this framework has found to be effective in improving students conceptual understanding and enhancing their competency to think and communicate more scientifically. As described by McNeil and Krajcik (2011), claim is a statement that answers a question, evidence is the data collected and used to support a claim, and reasoning is the justification used to explain why and how evidence supports a claim using science concepts. A fourth component – rebuttal can be introduced as students acquire more familiarity and expertise in this framework so that they can generate counter claim, evidence, and reasoning to explain why a counter alternative is not appropriate (McNeil & Krajcik, 2011). For this particular unit, we only included the first three components and they will be implemented throughout the various components of this unit, including lectures, physical lab experiments, and design challenge. We chose to do this because we considered that our students were not familiar enough with the framework yet.

Learning Materials

As previously mentioned, this unit was divided into three main components: lectures, physical lab experiments, and design challenge. In addition to using physical manipulatives for the lab experiments, an educational CAD tool called Energy3D (Xie, Schimpf, Chao, Nourian, & Massicotte, 2018) was used. Energy3D is an easy-to-use tool designed for the goal of educational research that enables students to design and construct energy efficient buildings. Energy3D collects data such as clickstream data and student actions in the form of JSON, in the background as student designs. In this unit, Energy3D is the main tool (i.e., each provided to students in an USB drive) students will use for the design challenge.

In addition, learning materials such as lecture notes, lab reports, and design challenge description were also included. Specifically, students were provided with four Word lab worksheets, two Excel sheets for data collection during the physical experiments, two PowerPoint templates for students to document their design, and two lecture PowerPoint slides. Appropriate prompts from the scientific argumentation framework will be embedded in each of these documents to provide appropriate scaffold to students. Besides, reflection questions will also be included to provide students an opportunity to reflect on their work and learning. Figure 2 below provides an example of how these documents would look like.

EXPERIMENT 1: Heat Transfer through Different Materials
Place each rod in tray with a small amount of water as shown. Turn on the hot plate (level 5) and wait until the water is hot (cover the tray such that the water becomes hot faster). Put on the tray the rods when the water is boiling. Wait at least one minute and record the temperature of at the top and bottom of any two rods using the **TURKEY** thermometer. Calculate temperature difference between the top and bottom.



Material	Temp. at Top (°C)	Temp. at Bottom (°C)	Temp diff. (Bott.– Top)

- Based your data, through which material is it easiest to transfer the heat and through which material is it hardest? **Provide all relevant evidence and reasoning to support your idea**.

Figure 2. Example of a Word lab worksheet (words in purple indicate prompts from the scientific argumentation framework).

Assessment

In terms of assessment, either in class or through their online submission, students were provided with both formative feedback and summative feedback. In both the lecture and the lab, students were provided with on-the-spot feedback through either question and answer (Q&A) or when the instructors saw a need for feedback or correction. In addition, students will be assessed for their i>Clicker responses in lecture (i.e., 3 points each lecture), group discussions (i.e., 5 points each lecture), and lab reports (i.e., 40 points each lab). These artifacts were assessed using a generic rubric that looked at completeness, clarity, and correctness. Students were provided with written feedback for these submissions if necessary. Figure 3 below shows the generic rubric used.

	Levels of Achievement				
Criteria	0-30% Achievement	30-50% Achievement	50-70% Achievement	70-90% Achievement	90-100% Achievement
COMPLETENESS	**0 Points** 0-30% of questions answered.	**2.5 Points** 30-50% of questions answered.	**5 Points** 50-70% of questions answered.	**7.5 Points** 70-90% of questions answered.	**10 Points** 90-100% of questions answered.
CLARITY	**0 Points** 0-30% of answers clearly explained.	**2.5 Points** 30-50% of answers clearly explained.	**5 Points** 50-70% of answers clearly explained.	**7.5 Points** 70-90% of answers clearly explained.	**10 Points** 90-100% of answers clearly explained.
CORRECTNESS	**0 Points** 0-30% of questions answered correctly.	**2.5 Points** 30-50% of questions answered correctly.	**5 Points** 50-70% of questions answered correctly.	**7.5 Points** 70-90% of questions answered correctly.	**10 Points** 90-100% of questions answered correctly.

Figure 3. Rubric for assessing student artifacts.

Curriculum Design

This unit is spread across the span of four weeks. Each week, there were a 50 minutes lecture and a 2 hours and 50 minutes lab session. Heat transfer was the focus of this unit. All the heat transfer concepts delivered through these activities were:

- Heat
- Temperature
- Radiation
- Albedo
- Solar Heat Gain Coefficient (SHGC)
- Conduction
- Thermal Conductivity, U-value, R-value
- Convection

The activities for each week are described below.

In week one, this unit started with a lab session. In this lab, students were introduced to a design challenge that they needed to work on and complete throughout this unit. The purpose of the design challenge was for students to construct an energy efficient home under certain requirements and constraints. With the remaining of time, students were asked to work on a suboptimal design where they evaluated and revised a pre-built home by making it more energy efficient. The goal of this activity was to allow students to familiarize themselves with the CAD tool (i.e., Energy3D) as well as to initiate their thinking on the design strategies that they plan to use for the design challenge.

In week two, this unit started with a lab session. This time, students were given time during the entire lab session to perform physical experiments related to radiation. As they worked on these experiments, they were required to complete a corresponding lab report. Then, in week three when students came back to the class, they were given time in the lecture to discuss and review the radiation concepts that they performed the week before with their peers and instructors. In the same week, students went to another lab session where they were given time to perform physical experiments related to conduction and convection. Again, similar to week two, students were required to complete a corresponding lab report as they work on those experiments.

In week four, this unit started with a lecture. This time when students came back to the lecture, they were given time to discuss and review the conduction and convection concepts that they worked on from the week before with their peers and instructors. The purpose of these reviews and discussions was to make sure that students have good understanding of the topics they learn as well as to clarify any misconceptions that they might have, if needed. Finally, in the last lab session, students were given time to complete the design challenge where they had to build an energy efficient home from scratch. Figure 4 below depicts the procedures of this four-weeks unit.

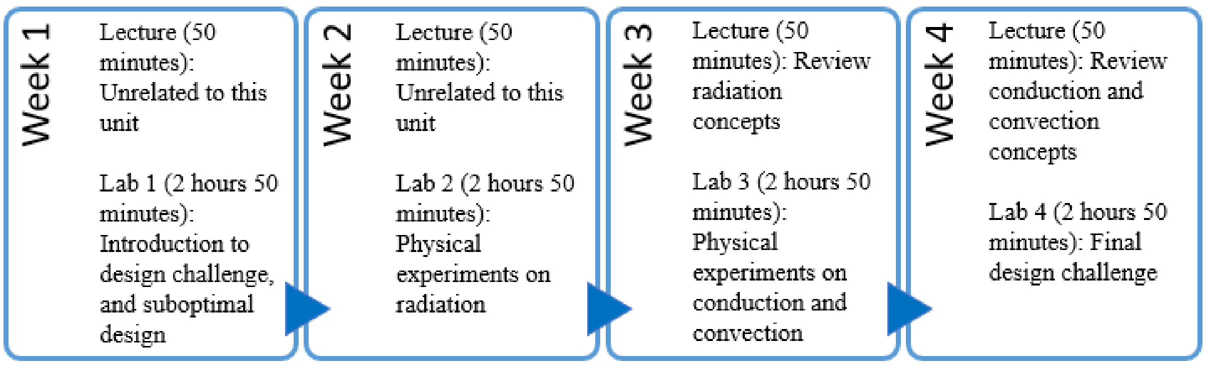


Figure 4. Procedures.

CHAPTER 5: METHODS

The research method used in this study was mixed method, specifically the mixed method convergent design (Creswell & Plano Clark, 2017). The main goal of convergent design is to "obtain different but complementary data on the same topic" (Morse, 1991, p. 122) with the purpose of better understanding the research problem. The convergent design procedures typically include four major steps: (a) the collection of both quantitative and qualitative data on topic of interest, (b) analysis of both quantitative and qualitative data separately and independently, (c) merging of two results, and (d) interpretation of results to meet study's overall purposes (Creswell & Plano Clark, 2017). We chose the convergent design because we were interested in collecting qualitative as well as quantitative data in order to help us answer our research questions. Since our data collection happened in an actual classroom, the convergent design provided us the ability to collect all the evidence we needed during one phase of the research, with the freedom of collecting and analyzing each data separately and independently, before having to draw a conclusion from all the data (Creswell & Plano Clark, 2017).

Setting and Participants

This study was conducted in a Physics for Elementary Education course for pre-service teachers for the span of three academic semesters: Spring 2018, Spring 2019, Fall 2019. The same activities were carried out in the three respective semesters. The total participants involved were one professor and 54 groups of pre-service teachers (i.e., 2 to 4 students in each group) with a background in Elementary Education. This study took place during a 4-week unit of the course, with Heat Transfer being the targeted science topic of that unit. In terms of learning outcomes of the unit, students were expected to (a) able to experiment with and develop deeper conceptual understanding of heat transfer processes (i.e., conduction, convection, radiation), and (b) able to engage in and learn science and engineering practices (e.g., design, experimentation, argumentation, informed decision-making).

Procedures

There were three main components involved in this study, which included lectures, lab experiments, as well as a design challenge (which took place during the lab sessions). As described in the previous chapter, the pedagogical approach for this study was a modified engineering design cycle incorporated into a 3E's inquiry learning cycle (Karplus & Butts, 1977; C. M. Rebello, 2019b). This 3E's cycle was enacted in the classroom in the following ways: (a) Explore – students were given opportunity in the lab to explore science concepts through performing physical experiments, (b) Explain – students were given opportunity to discuss and clarify scientific ideas with peers and instructors, and (c) Elaborate – students were given opportunity to apply the science ideas that they learned through working on a design task (i.e., a challenge to build an energy efficient home).

During week 1, students were given a pre-test to complete. The purpose of the pre-test was to assess students' conceptual understanding related to heat transfer. After that, students were introduced to a design challenge which they would be working on later in the study. In this challenge, students were asked to design an energy efficient home using an educational CAD software called Energy3D (Xie et al., 2018). Energy3D is a design CAD tool that allowed students to build prototype of their home, conduct analysis, and revise their prototype based on feedback received from those analysis. A screenshot of Energy3D is showed in Figure 5 below.

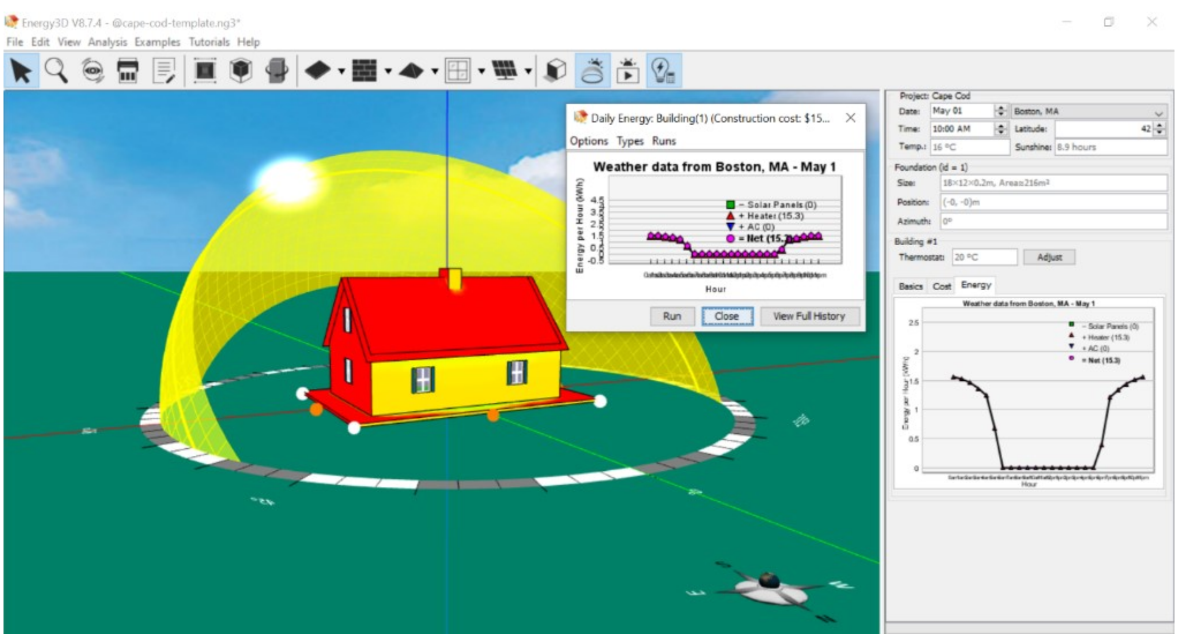


Figure 5. Screenshot of the user interface of Energy3D.

For the design challenge, students were given a list of design constraints and requirements that they should fulfil as they work on designing their energy efficient home. Some of these requirements included a size requirement of $200m^2$, a budget of \$120,000, and 0 kW/h or below for net energy consumption. Before working on this design challenge (which happened in week 4 of this study), students were provided with opportunities in week 1 to improve a sub-optimal (provided to them) using Energy3D based on their existing knowledge. As they worked on the challenge, students made notes of their ideas, design decisions, as well as strengths and weaknesses of their designs in their lab reports.

During week 2 and 3 of this study, students participated in two lab sessions where they conducted physical experiments. Students were divided into two conditions – the control condition and the experimental condition. Five groups of students, with each comprised of three to four students, were assigned as the control condition. On the other hand, seven groups of students, also with each comprised of three to four students, were assigned as the experimental condition. The assignment of conditions was simply based on the lab sessions that students were in. For example, students in the first lab session were assigned to be in the control condition whereas students in the second lab session were assigned to be in the experimental condition. Students in both conditions received identical instructions, with the exception that students in the experimental condition were given an argumentation scaffold. From past research, we learned that students, even since they were young, have the ability to engage in the practice of argumentation when provided with effective instruction, when prompted, when engaged in real-life problems they can relate to, as well as when given tasks that connect to their interests (Lee, 2017). Therefore, we believe that students in the control condition in this study were not discriminated from the practice of argumentation. Even though they were not provided with the spelled-out argumentation framework and examples, they were still provided with prompts that allow them to think about how to support their claims with relevant evidence and reasoning.

That being said, the argumentation scaffold used in this study was adapted from the Claim-Evidence-Reasoning framework by McNeill & Krajcik (2011). The purpose of the scaffold was to help students make claims based on valid evidence and supporting reasoning. The way we provided this scaffold to students was to provide them prompts throughout their lab reports. Students were also provided with examples of what a convincing argument would look like. After each lab session, all students met in the lecture where they discussed and clarified the science concepts learned with their instructors. Figure 6 below demonstrates an example of the argumentation scaffold provided in the science experiment lab reports. The phrase “predict what you think will happen…” aimed to help students elicit their claim, whereas the phrase “provide all relevant evidence, reasoning…” in purple aimed to remind students to support their claim with evidence and reasoning.

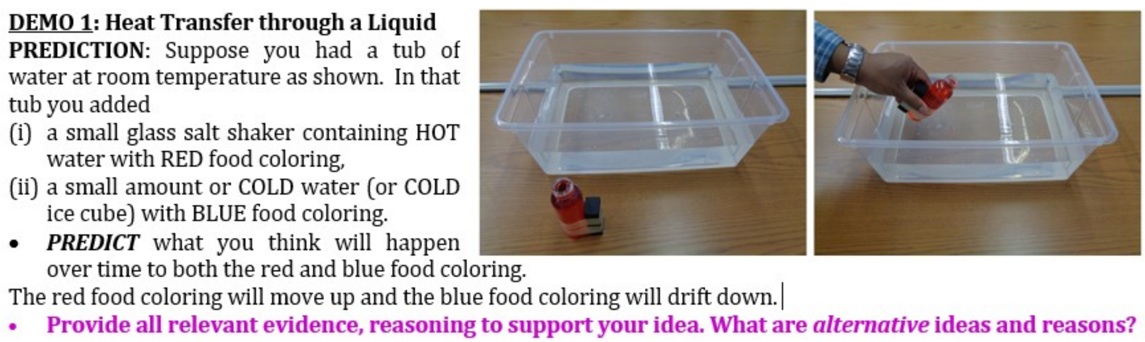


Figure 6. Example of argumentation scaffold in lab reports.

On the other hand, Figure 7 below demonstrates an example of the argumentation scaffold in the design challenge report. Similar prompts as those in Figure 6 were provided in the texts highlighted in yellow. In addition, the “example” provided students an example of what a good and convincing argument would look like, by individually labelling each component (i.e., claim, evidence, reasoning).

FACTOR	ARGUE PREDICTION	OBSERVATION	EXPLANATION - JUSTIFICATION
Which factor of the design are you going to change and how are you going to change it? *Example*: Solar panel tilt the orientation upward	**BEFORE YOU MAKE THE CHANGE: Provide reasoning for what *you think* will happen due to the change? (Construct an argument justifying why you are changing the factor and the outcome you expect. <u>Consider</u>: What evidence and reasoning supports your prediction?)** *Example*: Solar panels tilted upward generate more electrical energy [CLAIM] because they receive a greater energy from solar radiation when facing the sun directly [EVIDENCE], and convert the energy from solar radiation to electrical energy [REASONING]	**AFTER THE CHANGE: What *actually happened* due to the change?** *Example*: Solar panels facing the south side caused the annual energy cost of the home to decrease.	**Explain *why you think* it happened. (Construct an argument justifying your explanation by providing all relevant reasons <u>Consider</u>: What evidence and reasoning supports your explanation? Link science or other concepts *you think* are relevant with observations or other evidence. Are there alternative explanations?)** *Example*: Solar panels generated more energy [CLAIM] because when facing the sun directly, they received more energy from the sunlight [EVIDENCE], and converted the energy from sunlight to electrical energy [REASONING]

DO NOT FILL THIS OUT HERE. FILL THIS OUT IN THE POWERPOINT TEMPLATE PROVIDED

Figure 7. Example of argumentation scaffold in design challenge report.

During week 4 (i.e., the last week) of this study, students began working on their final design challenge where they had to build an energy efficient home from scratch using Energy3D. While they did that, they also documented their design ideas and justification in the form of lab reports. At the end of this study, students took a post-test assessing their conceptual understanding related to heat transfer, as well as a transfer task as part of the course exam. This transfer task required students to design a solar cooking, using the scientific principles that they had learned from the energy efficient home challenge, but in a slightly different context. Through the entire study, students worked in groups of three of four students. They each completed an individual report and compiled a document summarizing their design as a team.

Data Collection Method and Data Sources

All the data for this study was collected in the classroom throughout the entire four weeks of this unit. The data sources included in this study were: (a) four lab reports, including students' ideas and arguments– collected through students' submission on Blackboard, and (b) final design log files logged by Energy3D – collected through students' submission on Blackboard. Since we were interested in studying students' arguments to see how they make informed design decisions at a group level as well as how do their argumentation skills relate to their final design performance, we focused on only analyzing students' group arguments documented during their final design challenge and their final design artifacts.

Data Analysis Methods

The data analysis methods of this study included both quantitative (i.e., statistical analysis) and qualitative methods (i.e., content analysis). To begin, we extracted students' group decision-making arguments from their final design reports. Once we obtained those arguments, we grouped them according to their types of arguments: prediction or implementation. Design ideas that students wanted to test were classified as prediction arguments; design ideas that students actually tested were classified as implementation arguments. Once grouped, these arguments were separated into three components, which were Claim, Evidence, and Reasoning. Each of these components were then scores independently by raters using the adapted rubric from (McNeill et al., 2006), as shown in Table 3 below. It is crucial to note that only the first three component of the rubric were included for our scoring (i.e., Claim, Evidence, and Reasoning). The fourth component from the original rubric – Rebuttal, was removed because we considered it to be more suitable for students who have gained more experience and expertise with the framework, which was not yet the case for our students.

Table 3. Assessment rubric for scoring scientific arguments (McNeill et al., 2006).

Component	Level 0	Level 1	Level 2
Claim: an assertion or conclusion that answers the original question.	Does not make a claim or makes an inaccurate claim.	Makes an accurate but incomplete claim.	Makes an accurate and complete claim.
Evidence: scientific data that supports the claim; the data is appropriate and sufficient to support the claim.	Does not provide evidence or only provides inappropriate evidence for a claim.	Provides appropriate but insufficient evidence to support claim; may include inappropriate evidence.	Provides appropriate and sufficient evidence to support claim
Reasoning: a justification that links the claim and evidence and shows why the data counts as evidence to support the claim by using the appropriate principles.	Does not provide reasoning or only provides reasoning that does not link evidence to claim.	Provides reasoning that links the claim and evidence; repeats the evidence and includes some scientific principles but not sufficient.	Provides reasoning that links evidence to claim; includes appropriate and sufficient scientific principles.

Once all the arguments were scored, we calculated the average scores for each group and for each condition (i.e., control group and experimental group). Based on the scores, we were able to get an overall sense of students' performances. In addition, we also ran Mann-Whitney U tests to compare the performance between the control and the experimental conditions.

In addition to scoring students' arguments in general, we were interested in investigating the types of claims made by students, and the types of information they chose to use to justify their decision-making arguments. To accomplish that, we scored students' claim, evidence, and reasoning using two rubrics adapted from Wilson-Lopez and colleagues (2020), as provided in Table 4 and Table 5 below. The original rubric (Wilson-Lopez et al., 2020) has included more categories, but we have only selected a few that we thought were most relevant to our data.

Table 4. Rubric for categorizing the types of claims made by students. Adapted from (Wilson-Lopez et al., 2020).

Code	Category	Definition	Sample Quote
CONS	Constraint-oriented	Student argues that a design or a design element for a product or process, performs better or worse than other designs; and/ or that it meets specified criteria or constraints.	*"The cost of the home increased because our starting house was much smaller in area, but didn't reach the requirements, so we had to increase the size to the minimum."*
SCIE	Science-oriented	Student argues using a particular scientific or mathematical concept, to explain how and why a design works.	*"We increased the solar efficiency of the panel to get more energy through sunlight."*

Table 5. Rubric for categorizing the types of evidence and reasoning made by students. Adapted from (Wilson-Lopez et al., 2020).

Code	Category	Definition	Sample Quote
ECON	Economical constraint-oriented	Factors related to budget, economic impacts, or revenue to justify a claim.	*"The price of the house went from \$160,000 to \$151,708. Reducing the amount of walling reduces the price."*
ENGY	Energy constraint-oriented	Factors related to energy consumptions, or energy sustainability to justify a claim.	*"The net energy went from 7130 to -1135. "*
SCIE	Science-oriented	Scientific or mathematical concepts to justify a claim.	*"Since our solar panels standing more straight up to face the sun, more solar radiation was absorbed through the solar panel. Solar panels generated more energy because when ...directly facing the path of the sun there was more energy absorbed and this energy was converted into electrical energy."*

Once we coded the types for claim, evidence, and reasoning, we produced visual representations to help us see the results more clearly, including the transitions from the types of claims to the types of evidence, as well as from the types of evidence to the types of reasoning used by students. We also looked the how students used a combination of data to support their claims.

Finally, we were interested in measuring students' final design performances. To accomplish that, we scored the main elements of their final design solutions, which included the area (m^2), final cost ($), and final annual energy consumption (kW/h) of their final home designs. Each of these elements were scored independently before we added them all up to derive a final design performance score. Each element was weighted equally to make up 100% (i.e., 33.33% each). The formula used for calculating these scores is presented in Table 6 below. In addition, we ran Man-Whitney U test to compare students' performances scores.

Table 6. Calculation of students' final design performance scores.

Criteria	**Formula**
Area Score	$\frac{Final\ area}{200}$
Cost Score	$\frac{Final\ cost}{Max\ final\ cost}$
Energy Consumption Score	$\frac{Final\ energy\ consumption}{Max\ final\ energy\ consumption}$
Final Design Performance Score	*(Area Score/3) + (Cost Score/3) + (Energy Consumption Score/3)*

Trustworthiness

In order to ensure the validity and reliability of the instruments and the scoring process of this study, an inter-rater reliability was performed by three raters. The first rater (Rater 1) was doctoral student in cyberlearning, the second rater (Rater 2) was an interdisciplinary postdoctoral researcher, and the third rater (Rater 3) was an undergraduate research assistant. Rater 1 and Rater 2 both have had experience carrying out studies relevant to CAD-based design challenge, while the Rater 3 were trained by the first rater about the study. As part of the inter-rater reliability, all three raters were trained on the scientific concepts covered in this study, in addition to how to score the different components of arguments using the rubrics stated in the previous section. Once trained, all three raters were assigned specific data to score.

Rater 1 and Rater 2 scored the group arguments using the rubric presented in Table 3, for the data from Spring 2019, which contained 336 arguments (i.e., about 28% of the entire data set). On the other hand, Rater 1 and Rater 3 scored the arguments using the rubric presented in Table 4 and Table 5, for the data from Fall 2019, which contained 713 arguments (i.e., about 60% of the entire data set). All three raters scored these data independently and compared the degree of agreement subsequently. The raters repeated this scoring activity for three times. During the scoring process, the three raters would meet and have discussion whenever there was an uncertainty. Besides, Pearson's Correlation tests were performed for the final rounds of scoring. Results from the Pearson's Correlation tests showed strong positive correlations between all the three raters. Respectively, Rater 1 and Rater 2 obtained a r-score of 0.834 ($p < 0.001$), whereas Rater 1 and Rater 3 obtained a r-score of 0.994 ($p < 0.001$). Table 7 below presents the correlation score for the agreement between Rater 1 and Rater 2 on their scoring for items using rubric in Table 3, whereas Table 8 below presents the correlation score for the agreement between Rater 1 and Rater 3 on their scoring for items using rubric in Table 4 and Table 5.

Table 7. Results from Pearson's Correlation test for agreement between Rater 1 and Rater 2 on their scoring of items using rubric in Table 3.

		Rater 1	**Rater 2**
Rater 1	Pearson Correlation	1	.843**
	Sig. (2-tailed)		.000
	N	336	336
Rater 2	Pearson Correlation	.843**	1
	Sig. (2-tailed)	.000	
	N	336	336

**. Correlation is significant at the 0.01 level (2-tailed)

Table 8. Results from Pearson's Correlation test for agreement between Rater 1 and Rater 3 on their scoring of items using rubrics in Table 4 and Table 5.

		Rater 1	Rater 3
Rater 1	Pearson Correlation	1	.994**
	Sig. (2-tailed)		.000
	N	713	713
Rater 3	Pearson Correlation	.994**	1
	Sig. (2-tailed)	.000	
	N	713	713

**. Correlation is significant at the 0.01 level (2-tailed)

CHAPTER 6: RESULTS

As previously mentioned in Chapter 1, this research study was mainly guided by the following main research question and four sub-questions:

RQ: How did the scaffold of argumentation in the context of science experimentation influence students' design decision-making?

SRQ-1: What was the quality of decision-making arguments made by students?
SRQ-2: What types of claims were students making when they designed, and what types of evidence and reasoning were students using to back up their claims?
SRQ-3: What was students' level of performance in their final design?

In this chapter, we present the results for each sub research question in separate sections, and we then summarized all the results to give a big picture of the findings before moving on to the next chapter for a more detailed discussion.

SRQ-1: What was the quality of decision-making arguments made by students?

The results in this section describe the quality of decision-making arguments made by students. These results are reported based on the conditions that students were in, which were: control (i.e., without argumentation scaffold) and experimental (i.e., with argumentation scaffold); as well as the types of decision arguments students made, which were: prediction (i.e., arguments where students predicted their design decisions) and implementation (i.e., arguments where students explained their actual design decisions). There was a total of 21 groups of students from the control condition, and a total of 33 groups of students from the experimental condition. The quality of these arguments was scored using the rubric from Table 3. To recall, there are three score levels in this rubric: Level 0 (i.e., score 0), Level 1 (i.e., score 1), and Level 2 (i.e., score 2). Score 0 was assigned to elements of arguments that were of poor quality, score 1 was assigned to elements of arguments that were of moderate quality, and score 2 was assigned to elements of arguments that were of good quality. Figure 8 and Table 9 below present the statistics for prediction arguments for both conditions, whereas Figure 9 and Table 10 below present the statistics for implementation arguments for both conditions.

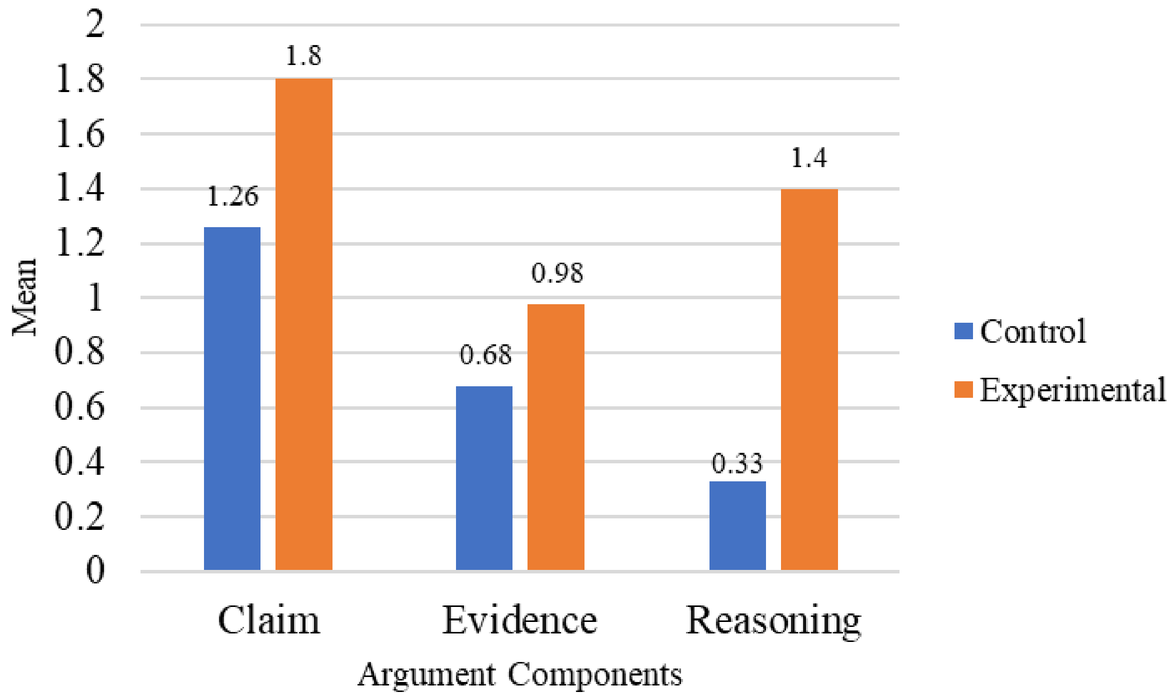


Figure 8. Mean scores for prediction arguments.

From Figure 8, we could see that students in the control condition scored a mean of 1.26 for their prediction claims, which denoted a moderate performance. This meant that most of these students were making claims that were somewhat accurate, but incomplete. In terms of prediction evidences, we could see that students in the control condition scored a mean of 0.68, which denoted a poor, somewhat close to moderate performance. This meant that most of the students either didn't provide any evidence, or only provided evidence that were inappropriate. In terms of reasoning, we could see that students in the control condition scored a mean of 0.33, which also denoted a poor performance. This meant that most of the students either didn't provide any reasoning, or only provided reasoning that didn't link their evidences to their claims.

On the other hand, we could see that students in the experimental condition scored a mean of 1.8 for their prediction claims, which denoted a moderate, somewhat close to good performance. This meant that most of the students makes accurate and complete claims. In terms of evidence, we could see that students in the experimental condition scored a mean of 0.98, which denoted a somewhat moderate performance. This meant that most of the students either provided appropriate but insufficient evidence to support their claims, or may include some inappropriate evidence. In terms of reasoning, we could see that students in the experimental condition scored a mean of 1.4, which denoted a moderate performance. This meant that most of the students provided reasoning that connect their claim and evidence, but repeated the evidence and included insufficient scientific principles.

As comparisons, it can be observed that the overall prediction arguments quality of the control condition was weaker than the overall prediction arguments quality of the experimental condition, based on their mean scores. Specifically, the control condition scored an average of 1.26 for the claim component, which was lower than the average of 1.8 scored by the experimental condition. Similar trends were noticeable for the evidence and the reasoning components. That was, the control condition scored an average of 0.68 for the evidence component, whereas the experimental condition scored higher at an average of 0.98. In addition, the control scored an average of 0.33 for the reasoning component, whereas the experimental condition scored higher at an average of 1.4. Based on these mean scores, students in the control condition made lower quality prediction arguments, as compared to the students in the experimental condition. Specifically, students in the control condition struggled most with providing good evidence and reasoning. Even though students in the experimental condition generated arguments of better quality overall, as compared to the control condition, it was still clear to see that these students struggled with providing good evidence, given that they had the lowest score for the evidence component, compared to the claim and the evidence components.

Table 9. Mann-Whitney U test statistics for prediction arguments.

	Condition						**Mann-Whitney U**	**Sig.(2-tailed)**
	Control			**Experimental**				
	M	**MR**	**n**	**M**	**MR**	**n**		
Claim	1.26	14.95	21	1.80	35.48	33	83.00	.000
Evidence	.68	18.62	21	.98	33.15	33	160.00	.000
Reasoning	.33	14.40	21	1.40	35.83	33	71.50	.000

In addition, the Mann-Whitney U test statistics from Table 9 provides further details. From these results, it was clear that there was a significant difference in the mean scores for the control and experimental conditions. For instance, there was significant difference in the scores for control – claim (M=1.26, MR=14.95) and experimental – claim (M=1.80, MR=35.48), with Mann-Whitney U=83, p=0.000; there was significant difference in the scores for control – evidence (M=0.68, MR=18.62) and experimental – evidence (M=0.98, MR=33.15), with Mann-Whitney U=160, p=0.000; there was significant difference in the scores for control – reasoning (M=0.33, MR=14.40) and experimental – reasoning (M=1.40, MR=35.83), with Mann-Whitney U=71.50, p=0.000. In short, these results suggested that students in the experimental condition produced prediction arguments that were significantly better than those of students in the control condition.

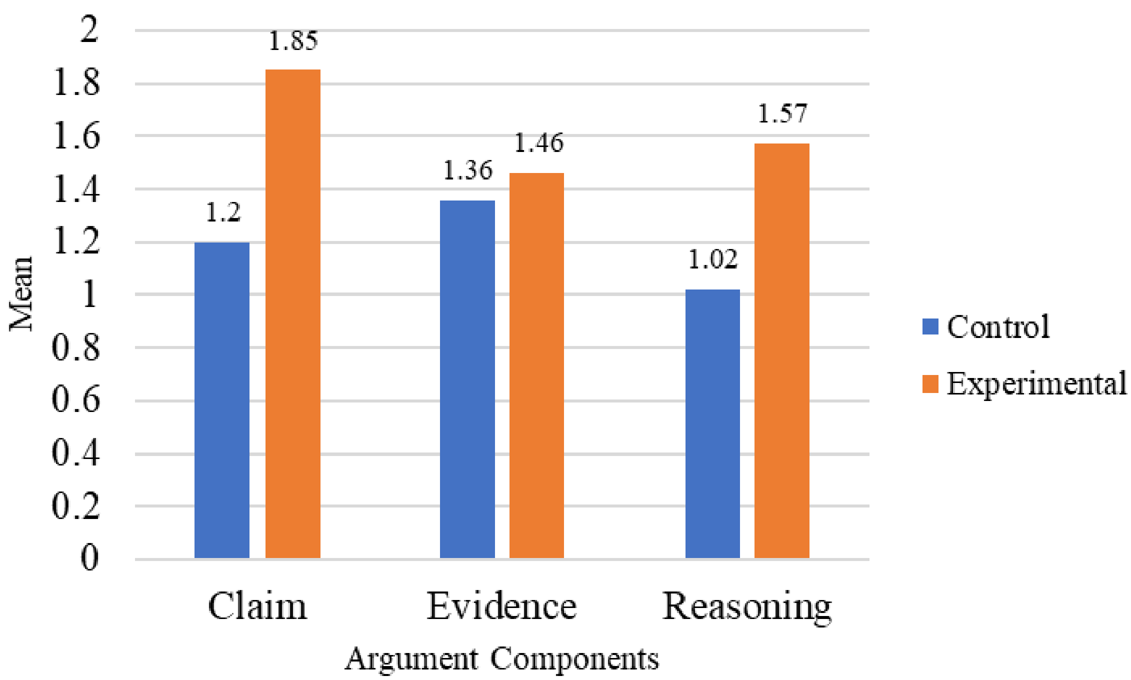


Figure 9. Mean scores for implementation arguments.

On the other hand, from Figure 9, we could see that students in the control condition scored a mean of 1.20 for their implementation claims, which denoted a moderate performance. This meant that most of these students were making claims that were somewhat accurate, but incomplete. In terms of prediction evidences, we could see that students in the control condition scored a mean of 1.36, which denoted a moderate performance. This meant that most of the students either provided appropriate but insufficient evidence to support their claims, or may include some inappropriate evidence. In terms of reasoning, we could see that students in the control condition scored a mean of 1.02, which also denoted a moderate performance. This meant that most of the students provided reasoning that connect their claim and evidence, but repeated the evidence and included insufficient scientific principles.

On the other hand, we could see that students in the experimental condition scored a mean of 1.85 for their implementation claims, which denoted a moderate, somewhat close to good performance. This meant that most of the students makes accurate and complete claims. In terms of evidence, we could see that students in the experimental condition scored a mean of 1.46, which denoted a somewhat moderate performance. This meant that most of the students either provided appropriate but insufficient evidence to support their claims, or may include some inappropriate evidence. In terms of reasoning, we could see that students in the experimental condition scored a mean of 1.57, which denoted a moderate, somewhat good performance. This meant that most of the students provided reasoning that connect their claim and evidence, but repeated the evidence and included insufficient scientific principles.

As comparisons, it can be observed that the overall implementation arguments quality of the control condition was weaker than the overall implementation arguments quality of the experimental condition, based on their mean scores. Specifically, the control condition scored an average of 1.20 for the claim component, which was lower than the average of 1.85 scored by the experimental condition. Similar trends were noticeable for the evidence and the reasoning components. That was, the control condition scored an average of 1.36 for the evidence component, whereas the experimental condition scored higher at an average of 1.46. In addition, the control scored an average of 1.02 for the reasoning component, whereas the experimental condition scored higher at an average of 1.57. Based on these mean scores, students in the control condition made lower quality implementation arguments, as compared to the students in the experimental condition. Specifically, students in the control condition struggled most with providing good claim and reasoning. Even though students in the experimental condition generated arguments of better quality overall, as compared to the control condition, it was still clear to see that the scores for the evidence element for both conditions were very close. This seems to suggest that students from both control and experimental conditions provided evidence of similar quality.

Table 10. Mann-Whitney U test statistics for implementation arguments.

	Condition						Mann-Whitney U	Sig.(2-tailed)
	Control			Experimental				
	M	**MR**	**n**	**M**	**MR**	**n**		
Claim	1.20	16.45	21	1.85	34.53	33	114.50	.000
Evidence	1.36	25.38	21	1.46	28.85	33	302.00	.417
Reasoning	1.02	17.93	21	1.57	33.59	33	145.50	.000

In addition, the Mann-Whitney U test statistics from Table 10 provides further details. From these results, it was clear that there was a significant difference in the mean scores for the claim and reasoning components for the control and experimental conditions. For instance, there was significant difference in the scores for control – claim (M=1.20, MR=16.45) and experimental – claim (M=1.85, MR=34.53), with Mann-Whitney U=114.50, p=0.000; there was also significant difference in the scores for control – reasoning (M=1.02, MR=17.93) and experimental – reasoning (M=1.57, MR=33.59), with Mann-Whitney U=145.50, p=0.000. However, there was no significant difference in the scores for control – evidence (M=1.36, MR=25.38) and experimental – evidence (M=1.46, MR=28.85), with Mann-Whitney U=302, p=0.417. In short, these results suggested that students in the experimental condition produced claim and reasoning as part of their implementation arguments that were significantly better than those of students in the control condition, with the exception for the evidence component where there was no significant difference between the two conditions.

Lastly, when comparing the mean scores between prediction arguments and implementation arguments within conditions, we could see that students from the control condition showed improvement in their implementation evidences (i.e., 1.36 vs. 0.68) and reasonings (i.e., 1.02 vs. 0.33), with the exception of their implementation claims (i.e., 1.20 vs. 1.26). On the other hand, we could see that students from the experimental condition showed improvement in their implementation claims (i.e., 1.85 vs. 1.8), evidences (i.e., 1.46 vs. 0.98), and reasonings (1.57 vs. 1.4). Despite the improvement showed by students in the control condition when they moved from prediction to implementation, students in the experimental group still demonstrated an overall performance that was significantly better.

SRQ-2: What types of claims were students making when they designed, and what types of evidence and reasoning were students using to back up their claims?

Moving forward, we dived deeper and took a closer look at the types of claims made by students when they designed, as well as the types of evidence and reasoning they used to back up their claims. The types of claims made by students were coded using the rubric in Table 4, whereas the types of evidence and reasoning provided by students were coded using the rubric in Table 5. The types of claim, evidence, and reasoning generated by students were then calculated in percentages and reported, as below.

Figure 10 below presents the different types of claim, along with percentages, made by students from both control and experimental conditions. From Figure 10, it could be seen that majority of the students from the control condition made constraint-oriented claims (i.e., 69%) whereas majority of the students from the experimental condition made science-oriented claims (i.e., 66.45%) when they designed. Only 24% of the students from the control condition made science-oriented claims and only 31.45% of the students from the experimental condition made constraint-oriented claims. There was a small number of students from both conditions who did not make any claims as they designed, which included 7% of students from the control condition and 2.01% of students from the experimental condition.

These results provided more insights into how students from both the control and experimental conditions made their design decisions. In other words, what factors did they consider as they made these design decisions. These results suggested that most students from the control condition focused more on whether or not they met the design constraints (i.e., constraint-oriented claim), whereas most students from the experimental condition focused more on how they could apply scientific concepts to make their designs work (i.e., science-oriented claim). Based on these results, it also seems that a small number of students did not make any clear claims about their design decision. However, comparatively, there was less students in the experimental condition who did not make any clear claims, as compared to the control condition.

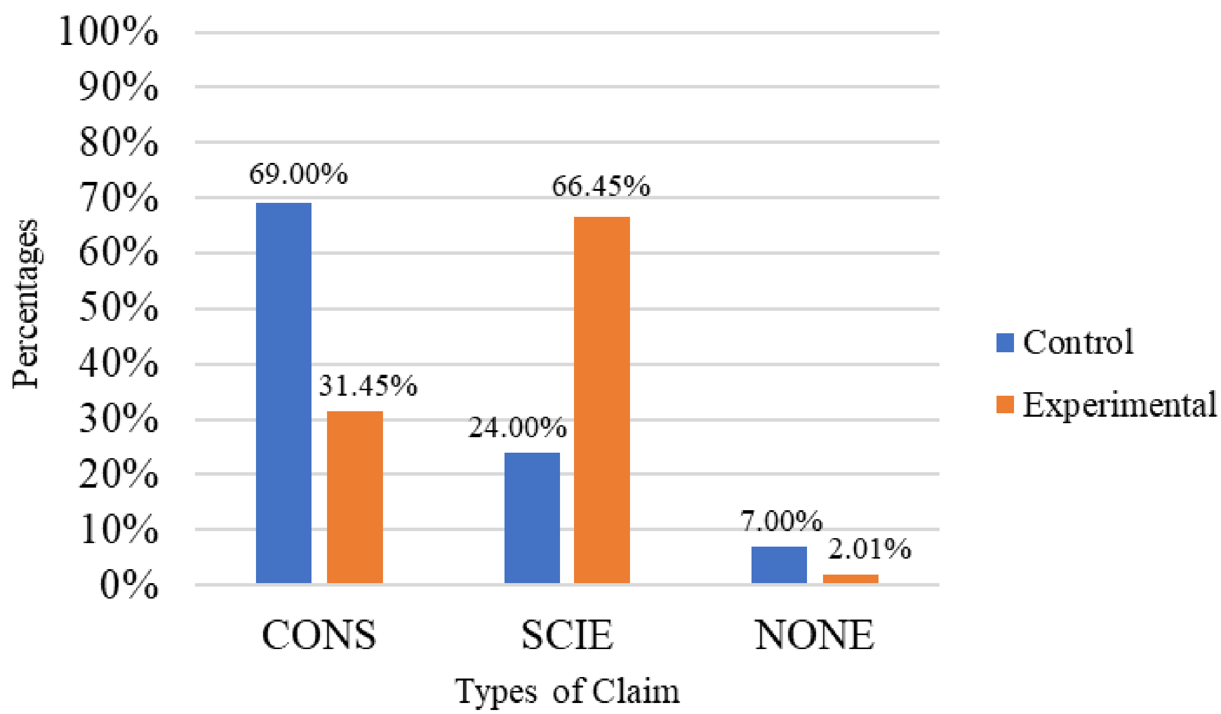


Figure 10. The types of claim made by students.

After understanding the types of claims made by students from each condition, we moved on to looking at the types of evidence and reasoning they used to back up their claims. The types of evidence and reasoning made by students were coded using the rubrics in Table 4 and Table 5. In the following paragraphs, we would first discuss the types of evidence and reasoning provided by students in the control condition, followed by the transition of combinations of claim, evidence, and reasoning made by students. Following that, we would discuss the same for the experimental condition.

Figure 11 below presents the types of claim, evidence, and reasoning, along with percentages for each category, made by students in the control condition. Figure 11 also demonstrates the transition of types from claim to evidence, and evidence to reasoning. This demonstration would give a better picture of the types of evidence and reasoning students used to support a certain type of claim.

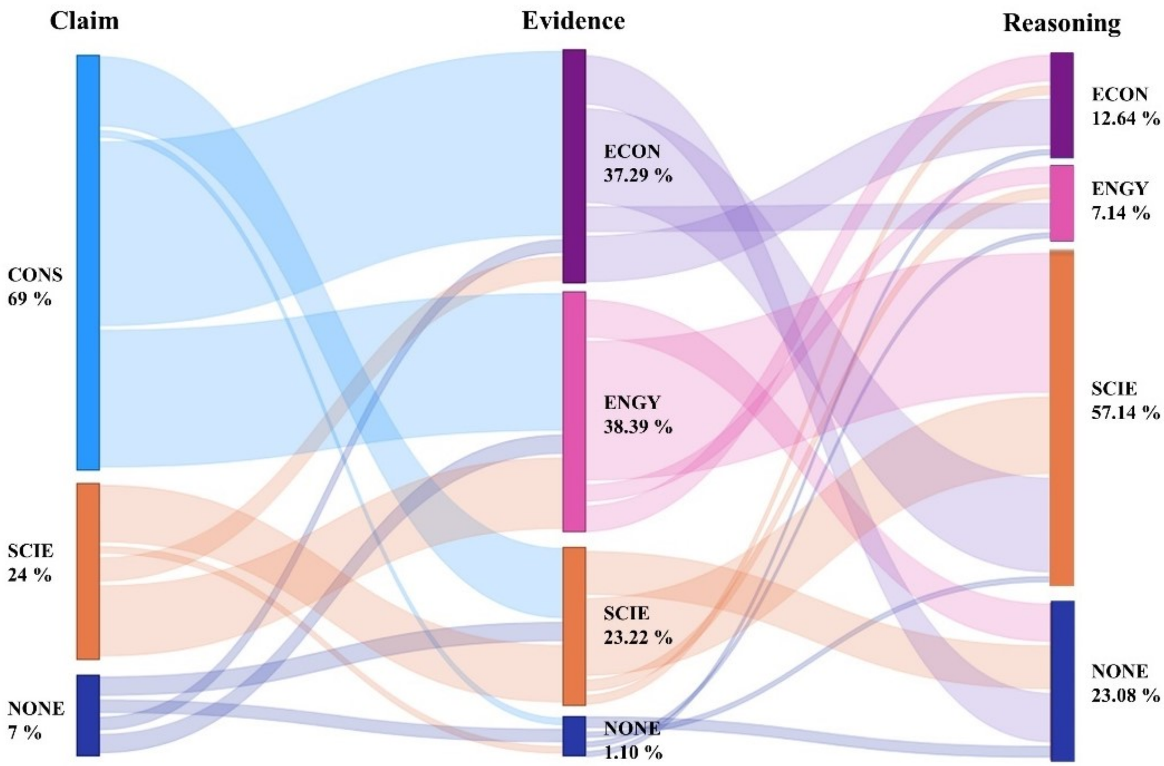


Figure 11. The types of claim, evidence, and reasoning made by students in the control condition, as well as the transition of types from claim to evidence, and from evidence to reasoning.

From Figure 11, in terms of claim, we could see that 69% of the students in the control condition tend to make constraint-oriented claims, whereas 24% of them tend to make science-oriented claim, and 7% of them did not make any claim at all. In terms of evidence, we could see that 37.29% of students in the control condition used economical constraint-oriented evidence, 38.39% of students used energy constraint-oriented evidence, 23.33% of students used science-oriented evidence, and 1.10% of students did not use any evidence to support their claims. In terms of reasoning, we could see that 12.64% of students used economical constraint-oriented reasoning, 7.14% of students used energy constraint-oriented reasoning, 57.14% of students used science-oriented reasoning, and 23.08% of students did not use any reasoning to further connect their claims and evidences as they construct their decision-making arguments.

Taking a closer look at the transition of types from claim to evidence, and from evidence to reasoning, we could see that most of the students who made constraint-oriented claims ended up using economic constraint-oriented evidence (i.e., 33.09%) and energy constraint-oriented evidence (i.e., 24.31%) to support their claims. Most of these students subsequently used science-oriented reasoning (i.e., 43.23%) to justify their decision-making arguments by linking their claims and evidences. On the other hand, we could see that most of the students who made science-oriented claims ended up using energy constraint-oriented evidence (i.e., 11.88%) and science-oriented evidence (i.e., 9.32%) to support their claims. Most of these students subsequently used science-oriented reasoning (i.e., 49.93%) to justify their decision-making arguments by linking their claims and evidences. Table 11 below presents sample arguments made by students from the control condition that represent the selected different types of transitions as shown in Figure 11.

Table 11. Sample arguments from selected different types of transitions made by students from the control condition.

Transition Path	**Claim**	**Evidence**	**Reasoning**
C:CONS – E:ENGY&ECON – R:ECON	*"The net energy increased only a little because we took away a solar panel that had been consuming energy from the sun."*	*"The net energy increased to -6,600 kWh with a difference of 3,218 kWh from the original. The cost decreased to $176,000 with a difference of $65,341 from the original."*	*"The cost went down because we eliminated the amount to materials we used."*
C:CONS – E:ENGY – R:SCIE	*"The energy did decrease as we decreased the foundation of the house."*	*"Due to the decreased area the energy is lost quicker."*	*"As conduction can move faster over a small area."*
C:SCIE – E:ENGY – R:SCIE	*"The energy consumption went up slightly after changing the color."*	*"Energy consumption after change: 18583 kwh."*	*"The albedo of the walls were lower, so less heat went through."*
C:SCIE – E:ENGY – R:ECON&ENGY	*"Adding solar panels allowed us to absorb energy from the sun and use the energy for the home."*	*"The net energy decreased."*	*"Thus decreasing the cost to heat and cool the home and lowering the net energy."*

As we continued to examine the types of evidence and reasoning provided by students in the control condition, it is important to note that there were instances where students used a combination of more than one type of evidence and reasoning to support certain claims. Figure 12 below shows the different combinations of types of evidence used by students in the control condition. Based on Figure 12, we could see that 21.18% of students provided only economical constraint-oriented evidence, 20% of students provided only energy constraint-oriented evidence, 17.65% of students provided only science-oriented evidence, whereas 28.24% of students provided a combination of economical constraint-oriented and energy constraint-oriented evidence, 2.35% of students provided a combination of energy constraint-oriented and science-oriented evidence, and 10.59% of students provided a combination of all three types of evidence.

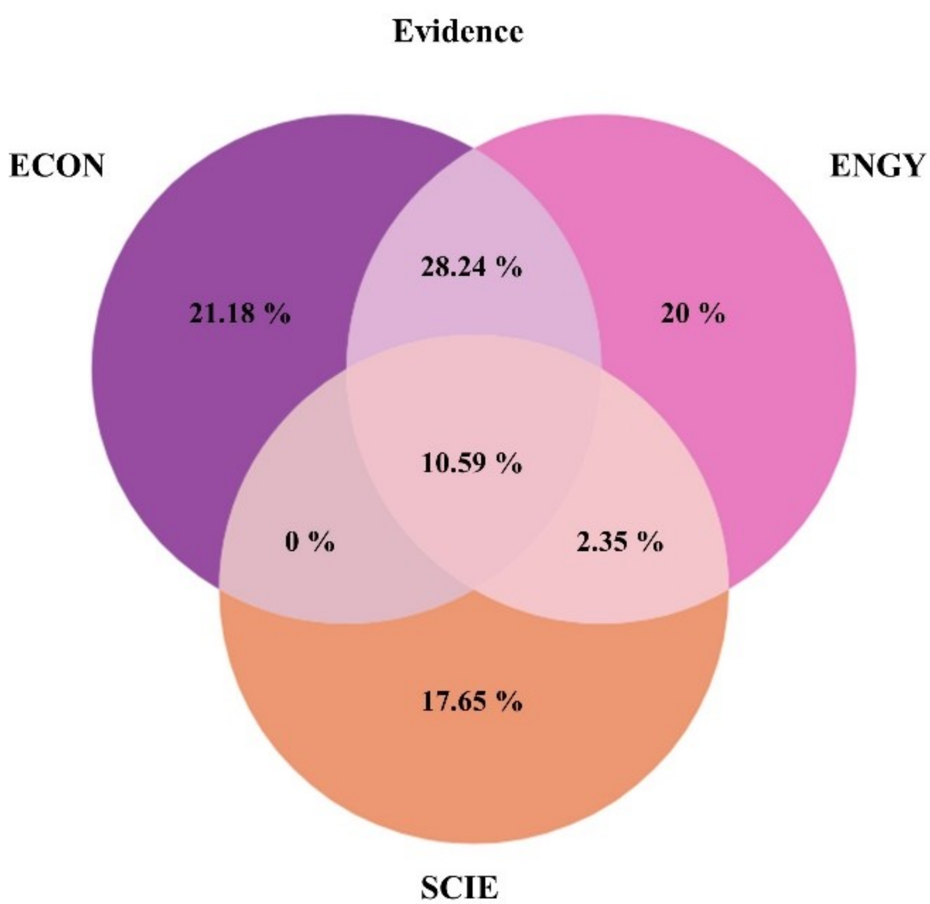


Figure 12. Combinations of types of evidence made by students in the control condition.

Similarly, Figure 13 below shows the different combinations of types of reasoning used by students in the control condition. Based on Figure 13, we could see that 14.29% of students provided only economical constraint-oriented reasoning 7.14% of students provided only energy constraint-oriented reasoning, 74.29% of students provided only science-oriented reasoning, whereas 4.29% of students provided a combination of economical constraint-oriented and energy constraint-oriented reasoning, and none of other students used a different combination of reasoning types.

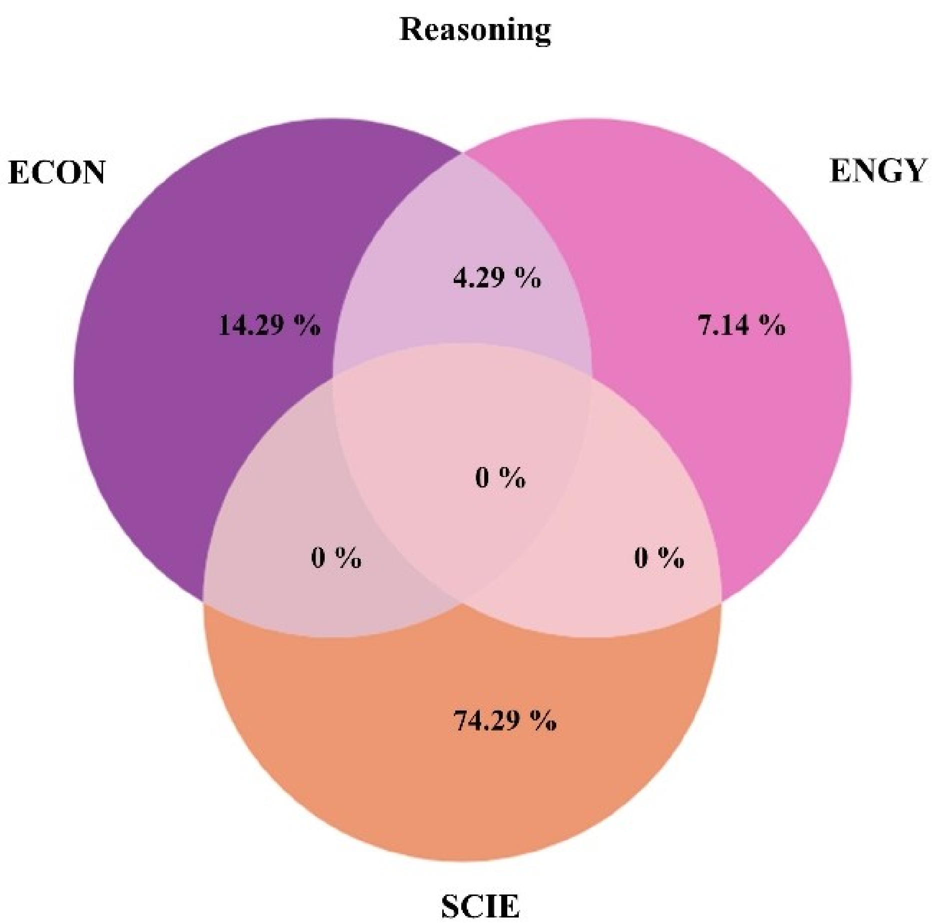


Figure 13. Combinations of types of reasoning made by students in the control condition.

Next, we looked at the results from the experimental condition. Figure 14 below presents the types of claim, evidence, and reasoning, along with percentages for each category, made by students in the experimental condition. Figure 14 also demonstrates the transition of types from claim to evidence, and evidence to reasoning. This demonstration would give a better picture of the types of evidence and reasoning students used to support a certain type of claim.

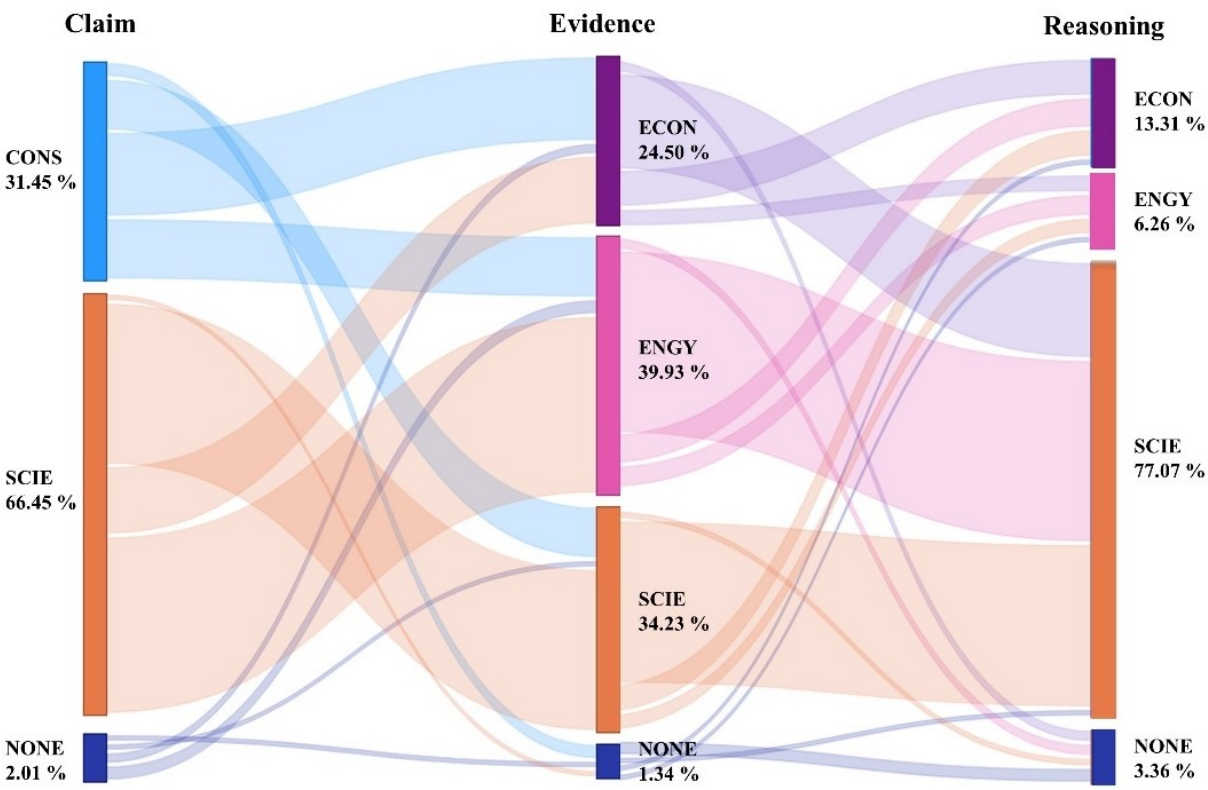


Figure 14. The types of claim, evidence, and reasoning made by students in the experimental condition, as well as the transition of types from claim to evidence, and from evidence to reasoning.

From Figure 14, in terms of claim, we could see that 31.45% of the students in the experimental condition tend to make constraint-oriented claims, whereas 66.45% of them tend to make science-oriented claim, and 2.01% of them did not make any claim at all. In terms of evidence, we could see that 24.50% of students in the experimental condition used economical constraint-oriented evidence, 39.93% of students used energy constraint-oriented evidence, 34.23% of students used science-oriented evidence, and 1.34% of students did not use any evidence to support their claims. In terms of reasoning, we could see that 13.31% of students used economical constraint-oriented reasoning, 6.26% of students used energy constraint-oriented reasoning, 77.07% of students used science-oriented reasoning, and 3.36% of students did not use any reasoning to further connect their claims and evidences as they construct their decision-making arguments.

Again, taking a closer look at the transition of types from claim to evidence, and from evidence to reasoning, we could see that most of the students who made constraint-oriented claims ended up using economical constraint-oriented evidence (i.e., 13.31%) and energy constraint-oriented evidence (i.e., 9.28%) to support their claims. Most of these students subsequently used science-oriented reasoning (i.e., 48.44%) to justify their decision-making arguments by linking their claims and evidences. On the other hand, we could see that most of the students who made science-oriented claims ended up using energy constraint-oriented evidence (i.e., 29.31%) and science-oriented evidence (i.e., 26.62%) to support their claims. Most of these students subsequently used science-oriented reasoning (i.e., 77.05%) to justify their decision-making arguments by linking their claims and evidences. Table 12 below presents sample arguments made by students from the experimental condition that represent the selected different types of transitions as shown in Figure 11.

Table 12. Sample arguments from selected different types of transition made by students from the experimental condition.

Transition Path	Claim	Evidence	Reasoning
C:SCIE – E:SCIE – R:SCIE	*"Adding insulation to the house decreased the annual net energy."*	*"Because the insulation slowed down the transition of heat and cool air through the walls."*	*"it lowered the need for heat and air usage."*
C:SCIE – E:ENGY&ECON – R:SCIE	*"Increasing the solar efficiency of the solar panel decreased the net energy of the house"*	*"Because the solar panel absorbs more sunlight .The net energy of the house decreased from -1800 to -5000."*	*"And the amount of solar energy converted into electrical energy increased."*
C:CONS – E:ECON – R:ECON	*"The cost of the house decreased after getting rid of one of the panels and decreasing its size."*	*"The cost of the house decreased by $50,000."*	*"By only putting one solar panel on the house and making it slightly smaller the cost of the house was decreased."*
C:CONS – E:CONS – R:SCIE&ECON&ENGY	*"Decreasing the total area of the walls decreased the cost of the home and the net energy."*	*"Because less materials were needed to build the house."*	*"Which decreases the cost, and less space inside the house, means less energy to heat the house and keep the energy inside of the house."*

Next, as we continued to examine the types of evidence and reasoning provided by students in the experimental condition, it is important to note that there were also instances where students used a combination of more than one type of evidence and reasoning to support certain claims. Figure 15 below shows the different combinations of types of evidence used by students in the experimental condition. Based on Figure 15, we could see that 6.25% of students provided only economic constraint-oriented evidence, 20.14% of students provided only energy constraint-oriented evidence, 23.61% of students provided only science-oriented evidence, whereas 20.83% of students provided a combination of economical constraint-oriented and energy constraint-oriented evidence, 7.64% of students provided a combination of energy constraint-oriented and science-oriented evidence, 4.86% of students provided a combination of economical constraint-oriented and science-oriented evidence, and 16.67% of students provided a combination of all three types of evidence.

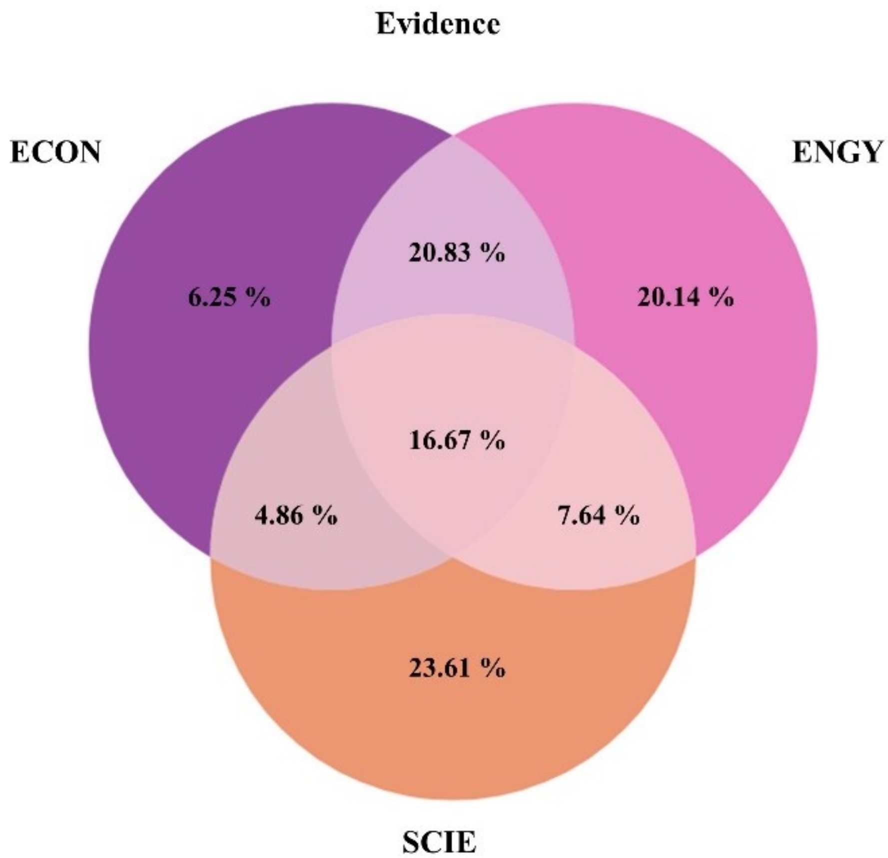


Figure 15. Combinations of types of evidence made by students in the experimental condition.

Similarly, Figure 16 below shows the different combinations of types of reasoning used by students in the experimental condition. Based on Figure 16, we could see that 9.72% of students provided only economical constraint-oriented reasoning, 3.47% of students provided only energy constraint-oriented reasoning, 77.78% of students provided only science-oriented reasoning, whereas 4.86% of students provided a combination of economical constraint-oriented and energy constraint-oriented reasoning, 0.69% of students provided a combination of energy constraint-oriented and science-oriented reasoning, 2.78% of students provided a combination of economical constraint-oriented and science-oriented reasoning, and 0.69% of students provided a combination of all three types of reasoning.

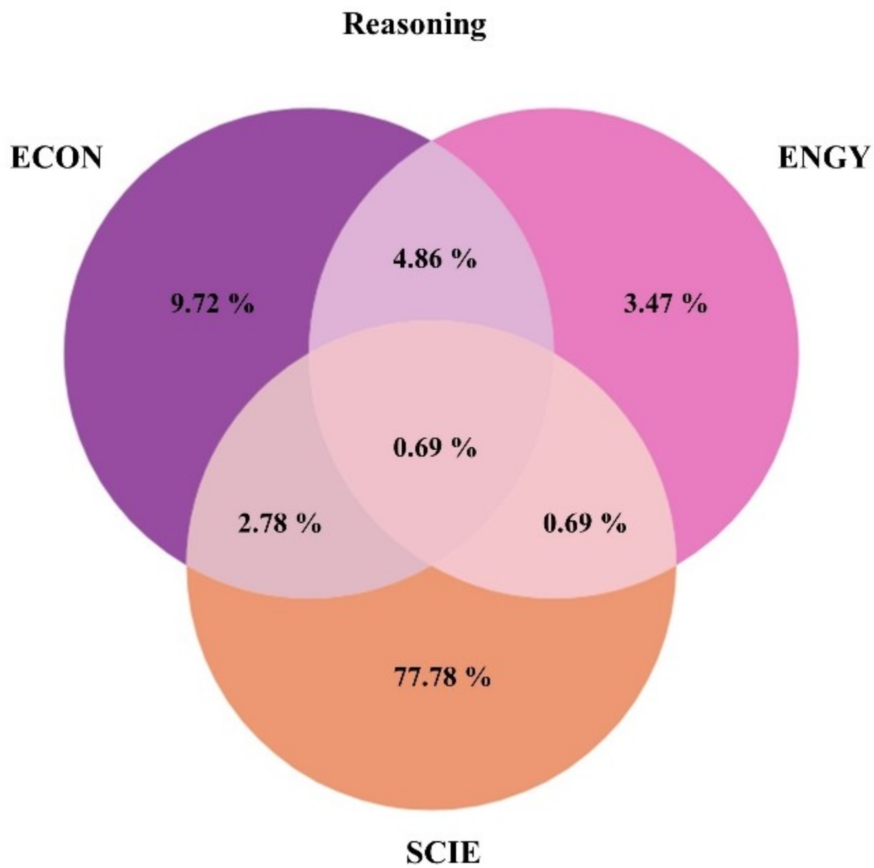


Figure 16. Combinations of types of reasoning made by students in the experimental condition.

Comparing the types of claim, evidence, and reasoning used by students from both control and experimental conditions, it was clear to see that comparatively, students from the experimental condition generated more science-oriented claims (i.e., 66.45%) as compared to students from the control condition (i.e., 24%). In addition, while students from the experimental condition focused more on providing energy constraint-oriented evidence (i.e., 39.93%) and science-oriented evidence (i.e., 34.23%), students from the control condition focused more on providing economical constraint-oriented (i.e., 37.29%) and energy constraint-oriented evidence (i.e., 38.39%). Interestingly, majority of the students from both the control and experimental conditions justified their arguments using science-oriented reasoning. However, there was still more students in the experimental condition who used science-oriented reasoning (i.e., 77.07%) as compared to the students in the control condition (i.e., 57.14%). Lastly, when comparing the combination of types of evidence and reasoning used by students, it was clear to note that students from the experimental condition utilized more combinations (i.e., 7 variations for evidence and 7 variations for reasoning) than the students in the control condition (i.e., 6 variations for evidence and 4 variations for reasoning).

SRQ-3: What was students' level of performance in their final designs?

Using the rubric in Table 6, the final design performance scores of students from both the control and experimental conditions were calculated. The scores included were size – how well students fulfilled the size requirement of the design challenge, cost – how well students fulfilled the cost requirement of the design challenge, energy consumption – how well students fulfilled the energy efficiency requirement of the design challenge, and final design performance – how well students performed in the design challenge as a whole (i.e., equally weighted the scores of size, cost, and energy consumption). Figure 17 below presents a comparison of these scores between the control condition and the experimental condition.

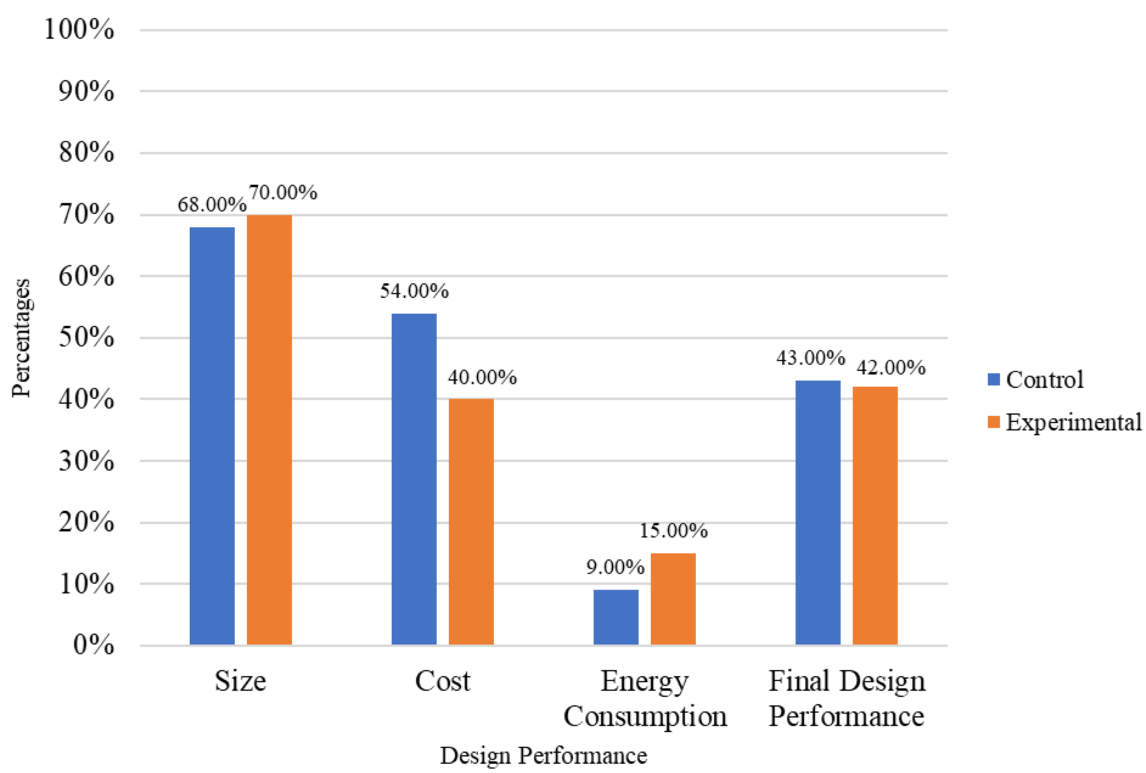


Figure 17. Students' final design performances.

Based on the results from Figure 17, it seems that students from the control condition were better at meeting the cost requirement (i.e., 54% vs. 40%), whereas the students from the experimental condition were better at meeting the size requirement (i.e., 70% vs. 68%) and the energy consumption requirement (i.e., 15% vs. 9%). In terms of the overall design performance, students from both the control and experimental conditions achieved very close scores, with only 1% of difference (i.e., 43% vs. 42%). Considering all factors, students from the experimental condition seemed to pay more attention to size and energy consumption of their designs, whereas students from the control condition seemed to pay more attention to the cost of their design. With these results in mind, we performed a Mann-Whitney U test to see if the differences between students' final design performance scores were significant. Table 13 below presents the Mann-Whitney U test statistics for students' final design performances for both conditions.

Table 13. Mann-Whitney U test statistics for students' final design performances.

	Condition						**Mann-Whitney U**	**Sig.(2-tailed)**
	Control			**Experimental**				
	M	**MR**	**n**	**M**	**MR**	**n**		
Size	.68	25.55	21	.70	28.74	33	305.50	.465
Cost	.54	28.67	21	.40	26.76	33	322.00	.642
Energy Consumption	.09	23.95	21	.15	29.76	33	272.00	.176
Final Design Performance	.43	28.57	21	.42	26.82	33	324.00	.689

Based on the results from Table 13, it was clear that there were no significance differences in the mean scores for all the design components for the control and experimental conditions, since all the p scores were greater than 0.05. Regardless, it was still interesting to observe the different factors that students from each condition performed better at (i.e., control condition was better at fulfilling cost requirement whereas experimental condition was better at fulfilling size and energy efficiency requirements).

In addition to the results presented above, we were interested in examining if the number of students working together in each group would impact the overall results (note that some groups only consisted of two students, and some others consisted of three to four students). To do that, we ran descriptive statistics as well as Mann-Whitney U tests for these groups separately (i.e., groups with only two students vs. groups with three to four students). In terms of the quality of their decision-making arguments, specifically prediction arguments, students in the experimental groups had rather similar or close mean scores in all three components, when comparing mean scores of groups with only two students versus groups with three to four students (e.g., Claim - 1.76 vs. 1.83; Evidence - .96 vs. .99; Reasoning – 1.36 vs. 1.47). However, we noticed that there were some relatively large differences when comparing the mean scores of the control groups, that was groups with only two students versus groups with three to four students. For example, control groups with only two students had a mean score of 1.69 for claim, as compared to the mean score of 1.00 of the control groups with three to four students. Similarly, control groups with only two students scored higher in reasoning (i.e., .42 vs. .18). On the other hand, control groups with three to four students scored higher in evidence (i.e., .93 vs. .44). In addition, we noticed statistically significant differences between the quality of prediction arguments made by the control and experiment groups, except for claim in groups with only two students (i.e., $p=.365$), as well as evidence in groups with three to four students (i.e., $p=.136$). Table 14 and 15 below show the Mann-Whitney U test statistics for prediction arguments, for groups with only two students and groups with three to four students, respectively.

Table 14. Mann-Whitney U test statistics for prediction arguments (groups with only two students).

	Condition						**Mann-Whitney U**	**Sig.(2-tailed)**
	Control (2 students)			**Experimental (2 students)**				
	M	**MR**	**n**	**M**	**MR**	**n**		
Claim	1.69	9.81	8	1.76	12.46	14	42.50	.365
Evidence	.44	5.56	8	.96	14.89	14	8.50	.000
Reasoning	.18	5.31	8	1.36	15.04	14	6.50	.000

Table 15. Mann-Whitney U test statistics for prediction arguments (groups with three to four students).

	Condition						Mann-Whitney U	Sig.(2-tailed)
	Control (3 or 4 students)			Experimental (3 or 4 students)				
	M	MR	n	M	MR	n		
Claim	1.00	7.08	13	1.83	22.95	19	1.00	.000
Evidence	.83	13.46	13	.99	18.58	19	84.00	.136
Reasoning	.42	9.42	13	1.47	21.34	19	31.50	.000

In terms of implementation arguments, students in the experimental groups had rather similar or close mean scores in claim and reasoning, when comparing mean scores of groups with only two students versus groups with three to four students (i.e., Claim - 1.86 vs. 1.85; Reasoning – 1.50 vs. 1.61), except evidence (i.e., 1.67 vs. 1.31). On the other hand, students in the control groups had rather similar or close mean score in reasoning (i.e., 1.01 vs. 1.03). However, there were some relatively big differences in claim and evidence (i.e., Claim – 1.45 vs. 1.05; Evidence – 1.55 vs. 1.25). In addition, we noticed statistically significant differences between the quality of implementation arguments made by the control and experiment groups, except for evidence in both groups with only two students (i.e., p=.441), as well as with three to four students (i.e., p=.910). Table 16 and 17 below show the Mann-Whitney U test statistics for implementation arguments, for groups with only two students and groups with three to four students, respectively.

Table 16. Mann-Whitney U test statistics for implementation arguments (groups with only two students).

	Condition						**Mann-Whitney U**	**Sig.(2-tailed)**
	Control (2 students)			**Experimental (2 students)**				
	M	**MR**	**n**	**M**	**MR**	**n**		
Claim	1.45	6.63	8	1.86	14.29	14	17.00	.006
Evidence	1.55	10.06	8	1.67	12.32	14	44.50	.441
Reasoning	1.01	7.75	8	1.50	13.64	14	26.00	.042

Table 17. Mann-Whitney U test statistics for implementation arguments (groups with three to four students).

	Condition						**Mann-Whitney U**	**Sig.(2-tailed)**
	Control (3 or 4 students)			**Experimental (3 or 4 students)**				
	M	**MR**	**n**	**M**	**MR**	**n**		
Claim	1.05	10.23	13	1.85	20.79	19	42.00	.001
Evidence	1.25	16.23	13	1.31	16.68	19	120.00	.910
Reasoning	1.03	10.81	13	1.61	20.39	19	49.50	.003

As for their final design performance, students in the experimental groups with only two students performed very slightly better in terms of size (i.e., .78 vs. .74), and energy consumption (i.e.,.11 vs. .07), except cost (i.e., .20 vs. .51), as compared to the control groups. For the students in the experimental groups with three to four students, they performed very slightly better in size (i.e., .65 vs. .64), and energy consumption (i.e., .19 vs. .11), except cost (i.e., .54 vs. .55), as compared to the control groups. However, it is important to note that none of these differences were statistically significant (i.e., all p scores were greater than 0.05). Table 18 below shows the Mann-Whitney U test statistics for students' final design performances, for groups with only two students and groups with three to four students, respectively.

Table 18. Mann-Whitney U test statistics for students' final design performances (groups with only two students).

	Condition						**Mann-Whitney U**	**Sig.(2-tailed)**
	Control (2 students)			**Experimental (2 students)**				
	M	**MR**	**n**	**M**	**MR**	**n**		
Size	.74	10.31	8	.78	12.18	14	46.50	.525
Cost	.51	13.44	8	.20	10.39	14	40.50	.297
Energy Consumption	.07	10.75	8	.11	11.93	14	50.00	.714
Final Design Performance	.44	14.06	8	.36	10.04	14	35.50	.165

Table 19. Mann-Whitney U test statistics for students' final design performances (groups with three to four students).

	Condition						**Mann-Whitney U**	**Sig.(2-tailed)**
	Control (3 or 4 students)			**Experimental (3 or 4 students)**				
	M	**MR**	**n**	**M**	**MR**	**n**		
Size	.64	15.85	13	.65	16.95	19	115.00	.762
Cost	.55	15.88	13	.54	16.92	19	115.50	.762
Energy Consumption	.11	13.42	13	.19	18.61	19	83.50	.126
Final Design Performance	.44	15.58	13	.46	17.13	19	111.50	.650

CHAPTER 7: DISCUSSION

Results Discussion

Chapter 6 presents the results that answer the three sub research questions of this study, under the overarching question of: How does the scaffold of argumentation in the context of science experiment influence students' design decision-making? From Chapter 6, we learned about the impact of the scaffold of argumentation on the quality of decision-making arguments made by students, the types of claim, evidence, and reasoning generated by students, as well as the final design performance of students (i.e., control vs. experimental). In this chapter, we would discuss the results of this entire study from a broader perspective, relating them back to the literature that informed this study and to other related works, as well as discuss the implications of these results for teaching and learning, as well as for integrated STEM education research.

Looking at all the results together, we could clearly see that students in the experimental condition (i.e., with argumentation scaffold) generated decision-making arguments of higher quality, devoted more attention to scientific principles when they made their decision claims, used more variety of combinations of evidence and reasoning to support their claims, utilized more scientific principles to back up their claims, incorporated more successfully the structure of the argumentation framework, as well as achieved slightly better performance in their final design in terms of fulfilling the size and energy consumption requirements.

These performance differences noticed between the control and the experimental conditions are in favor of our argument that the scaffold of argumentation has positive influence on fostering informed design decision-making in students. Specifically, it influenced and motivated students to ground their design decisions more in scientific principles, as evidenced by the better quality and the claim, evidence, and reasoning that were more science-oriented made by students in the experimental condition. This impact could potentially be contributed by the exercise of arguing using a predefined framework (in this case, the Claim – Evidence – Reasoning framework) while experimenting (McNeill & Krajcik, 2012). Using a predefined framework to help students practice how to make arguments while they worked on science experiments might have helped them to pay more attention to the connection between their claims and evidences, by justifying using the science knowledge that they already know previously or learned during the course (Erduran & Jimenez-Aleixandre, 2007). Besides, the exercise of arguing using a predefined framework could also have helped students to better articulate and refine their scientific knowledge (Driver et al., 2000; Jonassen & Kim, 2009; Kuhn, 1993). That being said, when students moved from science experiments to the design challenge, it was reasonable to observe that students in the experimental condition, who had the argumentation scaffold, were able to transfer the skills that they have learned from a science context to a design context, and performed better not only in general, but specifically more successful in making informed design decisions.

These explanations based on past research are indeed supported by what we observed from our results. Going back to the results in SRQ-1, we could see that students in the experimental condition generated decision-making arguments (i.e., both prediction and implementation arguments) that were significantly better than the ones generated by students in the control condition, with an exception for evidence component for implementation arguments. Even though the difference was not significant, students from the experimental condition still did slightly better in terms of the quality of their evidence, as compared to those of the control condition. Largely, these results are similar to findings in literature. For example, Becker (2014) found that students who were provided with explicit instruction in scientific explanation and argumentation were able to construct arguments that were significantly better than students without the instruction, in the context of biology. Other findings from various grade levels and domains such as chemistry and physics also showed positive effects of argumentation scaffold on quality of students' written arguments (Lizotte, McNeill, & Krajcik, 2004; McNeill & Krajcik, 2009; Nussbaum et al., 2008; Stark, Puhl, & Krause, 2009; Zembal-Saul, Munford, Crawford, Friedrichsen, & Land, 2002; Zohar & Nemet, 2002). Besides, based on the results in the first half of SRQ-2, in terms of claims, we could see that students in the experimental condition had higher tendency of generating more science-oriented claims, as compared to the students in the control condition who tend to generate more constraint-oriented claims. In addition, we could also see that there was a higher number of students in the control condition who did not have any specified claims in their decision-making arguments at all, as compared to those in the experimental condition. Based on these observations, we interpret students who leaned towards generating constraint-oriented claims as paying more attention on troubleshooting their designs so that they could fit and fulfil all the design constraints, without necessarily thinking much about the scientific reasoning behind each design action. On the other hand, we interpret students who leaned towards generating science-oriented claims as paying more attention on intentionally evaluating different design alternatives based on scientific reasoning, which subsequently made their design decision-making process a more informed one. Due to the complexity of the engineering design process, it is undeniable that different design aspects need to be taken into consideration, such as problem context, constraints and requirements, and scientific explanations. However, it is important to note that engineering investigations should not only focus on how well design constraints and criteria are met, but also how well the proposed solutions are supported by science (Bybee, 2011; NRC, 2012). Afterall, when applied efficiently, scientific principles can make the design solutions more informed, efficient, and effective in addressing the problems at hand (Bybee, 2011). In a way, we suggest that students who were able to produce more science-oriented claims might be more successful in making design decisions that are more informed. In addition, they might also be more successful in making decision-making arguments that are more complete (e.g., without missing information such as claims). Similar result was reported by Mastro (2017) that when provided with clear argumentation scaffold, students were able to organize and produce better and more coherent arguments.

Moreover, based on our results in the second half of SRQ-2, in terms of evidence, we could see that students in the experimental condition focused more on using energy constraint-oriented evidence and science-oriented evidence to back up their claims, whereas students in the control condition focused more on using economic constraint-oriented evidence and energy constraint-oriented evidence. This observation makes sense if we consider the tendency of students from the control condition to focus more on meeting the design constraints of the challenge, instead of using scientific principles to justify why certain design features were effective in making the home energy efficient, and some weren't. In terms of reasoning, we could see that most of the students from both the control and the experimental conditions justified their arguments using science-oriented reasoning. However, there were still more students in the experimental condition who justified scientifically compared to those in the control condition. Moreover, if we refer back to the results in SRQ-1, we could see that the reasoning generated by the students in the experimental condition were of significantly better quality. Reasoning is identified as one of the most challenging parts of constructing an argument (Bell & Linn, 2000; McNeill et al., 2006; Sandoval & Millwood, 2005); however, with the right scaffold, students' ability to reason could be strengthen, which is the case in our study. Therefore, even though most of the students from both conditions were able to justify their decision using scientific principles, students from the experimental condition showed better understanding of the scientific principles used, as reflected in the quality of reasoning made. This finding is consistent with what we observed in the claim and evidence components of this study, and is also aligned with findings from other studies (e.g., Becker, 2014). Furthermore, we could also see that students in the experimental condition utilized more variety of combination of data as evidence and reasoning to support their design claims. This is a strength demonstrated by students as they took into consideration the different aspects of design and utilized different types of supports for their design arguments (Wilson-Lopez et al., 2020). It is likely that the practice of argumentation helped students to better consider and select relevant data as evidence and reasoning to evaluate their design decisions (Kuhn, 1993). In addition, we could observe that students in the experimental condition demonstrated the ability to incorporate the structure of the argumentation framework better by having much less missing argument components, as compared to the students in the control condition. This could be contributed to the spelled-out or explicit approach where students were explained each component of the framework, and were provided scaffold that were clearly spelled-out as part of their instructions and deliverables (Nussbaum & Schraw, 2007).

Despite the major differences that we noticed in terms of the overall performances between the control and the experimental conditions, we observed that students in general still struggled with citing quantitative evidence, instead of qualitative and conceptual evidence to support their claims. This could potentially be due to the nature of heat transfer, which is mostly abstract and non-observable. Because of this nature, collecting and citing quantitative evidence even in science experiments can be challenging too, unless students were able to use high technological equipment, which is not too common in a standard classroom. Similar struggle was noticed in literature such as students' inability to include quantitative data in their arguments when it was their first time dealing with it (i.e., unfamiliarity) (Becker, 2014), they just simply failed to include data as evidence (Sandoval, 2003), or they faced difficulties in finding and evaluating data relevant to the design problems (Wertz, Purzer, Fosmire, & Cardella, 2013). It could also be possible that students struggled to find or generate the evidence that they need in Energy3D, either due to their incomplete familiarity of the software, or the lack of robustness of the software at its current stage to provide in-depth quantitative evidence. It is also possible that the skill of collecting and citing evidence effectively is something that takes longer time for students to master. In a study where progression of students' ability to construct scientific arguments over the course of a semester was examined, findings showed overall improvement in the quality of arguments (i.e., including evidence) made by students over time. (Becker, 2014). Therefore, it is possible that the skill of effectively collecting and citing evidence can be improved when given more time for students to practice.

Next, looking at the results from SRQ-3, in terms of final design performance, even though there was no significant difference between the scores of both the control and the experimental conditions, it is still interesting to see that students from the experimental condition performed slightly better in terms of size and energy consumption, while students from the control condition performed slightly better in terms of cost. We believe that these results align with the other observation that we have seen so far. In addition, we also believe that students who had the opportunity to practice arguing with the argumentation scaffold were put under a situation where they had to metacognitively and carefully weigh the different constraints that they had to fulfil, as well as the validity and accuracy of their arguments before committing to a design decision, which subsequently resulted in better performance in the areas that they prioritized (i.e., size and energy consumption) (Adams & Atman, 2000; Crismond & Adams, 2012). Nevertheless, engineering design is a complex process where it involves making many choices during decision-making and it is important that students are able to balance all benefits and trade-offs from the beginning to the end (Crismond & Adams, 2012). Therefore, it is important to note that students from neither the control nor the experimental condition were able to successfully balance all the trade-off they needed to consider. This could be due to the lack of instructional supports or structures that guide students to weigh benefits of a design against its drawbacks (Wilson-Lopez et al., 2020). Students could also have less tendency to remember to include design constraints that were previously identified into their final design if they don't systematically record them or revisit them (Wilson, Smith, & Householder, 2014), regardless of the quality of each design decision they made.

Lastly, by comparing students' performances, either in arguments quality or final design performances, based on the number of students in each group (i.e., either only two students or three to four students), we noticed that there were not much differences in terms of argument quality for students in the experimental condition. For example, experimental groups with only two students had very similar scores for their prediction arguments with the experimental groups with three to four students. However, we observed more noticeable differences in argument quality between control groups with only two students and control groups with three to four students. For example, control groups with only two students performed much better in terms of the quality of their prediction claim and reasoning, whereas control groups with three to four students performed much better in terms of the quality of their prediction evidence. In another example, control groups with only two students performed much better in terms of the quality of their implementation claim and evidence. In terms of final design performances on the other hand, not much differences were noticed between all groups of various group sizes. It is interesting to note that group sizes (i.e., either only two students or three to four students) seemed to have more impact for students in the control condition, and not much for students in the experimental condition. From the surface, it seemed control groups with only two students performed generally better in terms of arguments quality. However, we do not have enough evidence to make that claim in this study.

In summary, our results proved that engaging students with the argumentation scaffold in the context of science experiments can not only help students improve argumentation skills, but also allow them to transfer those skills to a design context, and subsequently help them to make better informed design decisions.

Implications for Teaching and Learning

Using argumentation scaffold in the context of science experiments demonstrated to be an effective approach in engaging students in the argumentation process where they connected science principles to their design experience, and consequently were able to make better informed design decisions. Specifically, students engaged in the argumentation scaffold were able to make decision-making arguments that were of significantly higher quality. These students were also able to make claims that were more scientific oriented, supported by science- and energy-oriented evidence and reasoning, which suggested a closer connection between their science knowledge and design decisions. Besides, these students also demonstrated the ability to incorporate the structure of the argumentation framework more successfully, by producing arguments that were completer and more consistent in terms of structure and component. Therefore, we recommend educators who are interested in helping students to foster and improve their informed decision-making skills to consider the argumentation approach that we proposed and implemented in this study. To ensure its effectively, we suggest educators to spend time in making the scaffold explicit, as well as clearly explain to their students how to construct good arguments, and why is it important to do so (Mcneill & Krajcik, 2008; Nussbaum & Kardash, 2005). We also think it is a good idea for educator to utilize prompting questions and examples when using an argumentation scaffold (Nussbaum & Schraw, 2007), as well as carefully select and adapt a set a framework that is most appropriate for the students that they are teaching (Driver et al., 2000).

We also learned from this study that students in general (i.e., from both control and experimental conditions) struggle with collecting and citing quantitative evidence, rather than qualitative and conceptual evidence. That is, students were generally able to use relevant evidence qualitatively to describe cause-and-effect relationships, but they struggle to cite "specific numbers or data" as evidence. As previously discussed, this could be due to the abstract nature of heat transfer, lack of high technology to produce those evidence, students' lack of familiarity with Energy3D to collect such evidence, or the potentially longer time needed to master this skill. Therefore, we suggest educators to consider providing more training and practice opportunities for students on collecting and citing appropriate evidence (e.g., explain importance and differences between qualitative and quantitative evidence), and carefully select appropriate software to pair with the subject they are teaching to avoid hinderance to evidence collection.

Additionally, we were interested to see if group sizes (i.e., groups with only two students versus groups with three to four students) affect students' overall performance in this study. This was based on the concern/assumption that groups with more students might experience better learning because more ideas get to be contributed from more minds, or the opposite that some students might experience "freeloading" in bigger groups, and hence affect their performance negatively. Regardless, results from this study showed that the performance of students in the experimental condition (i.e., with argumentation scaffold) were not really affected by the number of students in each group. On the other hand, we noticed some performances of students in the control condition seemed to be affected by the number of students in each group. Particularly, groups with only two students seemed to be performing comparatively better in some components of their arguments. However, we did not have consistent and enough evidence to make a claim that group size is a factor that affected students' performance. Nevertheless, we still suggest educators to consider group size as they assign students in groups. This could be helpful to ensure students' participation and the quality of their learning experience.

Lastly, we learned that students with the argumentation scaffold performed slightly better in their final design in terms of the size and energy consumption constraints. On the other hand, students without the argumentation scaffold performed slightly better in terms of the cost constraint. It is important to note that there was no significant difference in their overall performance. This result demonstrated that, even though students with the argumentation scaffold were able to make more scientifically informed design decisions, they did not necessarily perform better in their final design. This could be due to their lack of competency in other design skills such as balancing trade-offs and troubleshooting. Therefore, we suggest educators to consider incorporating other scaffolding approaches to help students develop other design skills as they engage in engineering design activities.

Implications for Integrated STEM Education Research

In this study, we investigated the effectiveness of the use of argumentation scaffold in the context of science experiments in fostering students' informed design decision-making skill. We introduced the scientific argumentation framework (McNeill & Krajcik, 2011) as a scaffold in a curriculum unit that implemented a modified engineering design cycle incorporated into an 3E's inquiry learning cycle. Based on our findings, we believe that we were able to bring the "design-science gap" (Vattam & Kolodner, 2008) a little closer by helping students to connect their design experience with their science knowledge. We accomplished that by using the argumentation scaffold as a medium that tied the scientific experiments and design challenge experience together, providing students framework and guidance in using their science knowledge to inform their design decisions.

The argumentation framework used in this study was largely used in the science domain. However, findings from this study suggest that it could be useful and applicable in the engineering domain as well. Therefore, it would be interesting to see more research work on implementing such framework in the engineering domain. Another interesting thing we learned from this study is that, when using the argumentation framework rubric for assessment (i.e., Table 3), it was at times difficult to clearly distinguish the components of evidence and reasoning. In addition, due to the nature of engineering, an argument may be backed up by multiple sources of evidence and reasoning for a single claim (e.g., a single decision can be made to fulfil multiple constraints at once), which is not currently accounted for in the rubric (i.e., Table 3) we used in this study. Therefore, we think it could be helpful and important to expand and adapt the scientific argumentation framework (McNeill & Krajcik, 2011) to reduce the confusion in distinguishing evidence and reasoning, as well as to incorporate elements that would fit the assessment for design arguments better.

CHAPTER 8: CONCLUSION

Limitations of the Study

In this study, even though we believed students in the control condition did not miss out on opportunities to practice argumentation, we did not provide them with information about the argumentation scaffold that the other students in the experimental condition had received. A simple remedy could be providing these students the scaffold information at the end of the curriculum unit. In addition, we did not get to study the long-term effect on transfer tasks to see if students could retain what they learned and transfer them to different contexts. We also did not assess students' conceptual change of their scientific knowledge, as well as their arguments in the context if scientific experiments. Moreover, we did not have a comparable treatment as a second experimental group for us to compare argumentation scaffold with other pedagogy approaches.

Conclusion and Future Work

In conclusion, in this study we learned the use of argumentation scaffold is effective in helping students foster the practice of depending on scientific principles to make informed design decisions. Particularly, it helped students to make decisions of higher quality by mostly depending on scientific evidence and reasoning. It also helped students to better articulate their design rationales by clearing stating their claims, evidence, and reasoning (i.e., evidenced by students' ability to incorporate the structure of argumentation better). At the same time, we also learned that more effort is still needed in terms of helping students to learn how to provide good and sufficient evidence (i.e., quantitative evidence), as well as to learn better engineering design skills such as balancing trade-offs. In addition, we learned that the argumentation scaffold could also be an effective tool to bridge the "design-science gap" by helping students to connect their design experience to their science knowledge. We believe that it is worth making effort to incorporate such scaffold more in the engineering setting, as well as to expand the current argumentation framework to incorporate elements of engineering design.

Based on the limitations of this study, we suggest the following future work. It would be interesting to investigate the long-term effect of argumentation training on transfer design tasks, as well as the effect on transfer tasks in different contexts (i.e., far transfer). It would also be interesting to measure students' change in scientific conceptual understanding before and after the teaching intervention (e.g., pre-test and post-test), or comparison of students' scientific conceptual understanding between conditions (e.g., control versus experimental). Other interesting future work could include comparing the approach suggested in our study with other pedagogical approaches to compare their effectiveness. In addition, it would be beneficial to expand the work on designing learning environments that promote engineering-related argumentation (Wilson-Lopez et al., 2020). Hence, some potential research questions might include:

- What is the long-term effect of argumentation scaffold in the context of science experiments on students' ability to make informed design decision?
- What is the effect of argumentation scaffold in the context of science experiments on students' ability to make informed design decision in various design contexts?
- How does the argumentation scaffold affect students' scientific conceptual understanding?
- How does the effectiveness of argumentation scaffold compare to other scaffolding approaches?
- How can the argumentation scaffold be modified to include multiple competing constraints of engineering design?
- What are the supports needed to successfully implement engineering-related argumentation?
- What are ways that engineering-related argumentation can be assessed?

REFERENCES

ABET. (2018). *Criteria for Accrediting Applied and Natural Science Programs, 2018 – 2019*. Retrieved from https://www.abet.org/accreditation/accreditation-criteria/criteria-for-accrediting-applied-and-natural-science-programs-2018-2019/

Adams, R., & Atman, C. (2000). Characterizing engineering student design processes: An illustration of iteration. *Proceedings of the 2000 ASEE Conference and Exhibition*.

Alexander, C. (1964). *Notes on the synthesis of form*. Harvard University Press.

Almendra, R. A., & Christiaans, H. (2009). Decision-making in design : A comparative study. *Proceedings of the 2nd International Conference On Research into Design*, Banglore, India. January 7-9, 2009.

Apedoe, X. S., & Schunn, C. D. (2013). Strategies for success: Uncovering what makes students successful in design and learning. *Instructional Science*, *41*(4), 773–791. https://doi.org/10.1007/s11251-012-9251-4

Becker, R. E. (2014). *Explicit instruction of scientific explanation and argument in an undergraduate introductory biology laboratory course using the claim, evidence and reasoning framework* [Doctoral dissertation, Louisiana State University]. Retrieved from https://digitalcommons.lsu.edu/gradschool_dissertations/3152

Bell, P., & Linn, M. C. (2000). Scientific arguments as learning artifacts: Designing for learning from the web with KIE. *International Journal of Science Education,22*(8),797-817. https://doi.org/10.1080/095006900412284

Berland, L. K., Martin, T. H., Ko, P., Peacock, S. B., Rudolph, J. J., & Golubski, C. (2013). Student learning in challenge-based engineering curricula. *Journal of Pre-College Engineering Education Research (J-PEER)*. https://doi.org/10.7771/2157-9288.1080

Berland, L. K., & McNeill, K. L. (2010). A learning progression for scientific argumentation: Understanding student work and designing supportive instructional contexts. *Science Education*, *94*(5), 765–793. https://doi.org/10.1002/sce.20402

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2014). Defining twenty-first century skills. In *Assessment and teaching of 21st century skills*. https://doi.org/10.1007/978-94-007-2324-5_2

Blair, J. A., & Johnson, R. H. (1987). Argumentation as dialectical. *Argumentation*, *1*(1).

Brown, N. J. S., Furtak, E. M., Timms, M., Nagashima, S. O., & Wilson, M. (2010). The evidence-based reasoning framework: Assessing scientific reasoning. *Educational Assessment*, *15*(3), 123–141. https://doi.org/10.1080/10627197.2010.530551

Bruner, J. (1985). Vygotsky: a historical and conceptual perspective. In *Culture communication and cognition: Vygotskian perspectives*.

Bybee, R. W. (2011). Scientific and engineering practices in k-12 classrooms: Understanding a framework for K-12 science education. *Science and Children, 49*(4),10.

CAEP. (2013). *2013 CAEP Standards*. Retrieved from http://caepnet.org/~/media/Files/caep/standards/caep-standards-one-pager-0219.pdf?la=en

Carlsen, W. S. (1998). Engineering design in the classroom: Is it good science education or is it revolting? *Research in Science Education, 28*(1), 51-63. https://doi.org/10.1007/BF02461641

Chao, J., Xie, C., Nourian, S., Chen, G., Bailey, S., Goldstein, M. H., … Tutwiler, M. S. (2017). Bridging the design-science gap with tools: Science learning and design behaviors in a simulated environment for engineering design. *Journal of Research in Science Teaching*, *54*(8), 1049–1096. https://doi.org/10.1002/tea.21398

Chase, C. C., Malkiewich, L., & S. Kumar, A. (2019). Learning to notice science concepts in engineering activities and transfer situations. *Science Education*, *103*(2), 440–471. https://doi.org/10.1002/sce.21496

Cho, K. L., & Jonassen, D. H. (2002). The effects of argumentation scaffolds on argumentation and problem solving. *Educational Technology Research and Development, 50*(3),5. https://doi.org/10.1007/BF02505022

Christensen, B. T., & Ball, L. J. (2016). Dimensions of creative evaluation: Distinct design and reasoning strategies for aesthetic, functional and originality judgments. *Design Studies, 45,* 116-136. https://doi.org/10.1016/j.destud.2015.12.005

Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and Conducting Mixed Methods Research.* SAGE Publications.

Crismond, D. P., & Adams, R. S. (2012). The informed design teaching and learning matrix. *Journal of Engineering Education*, *101*(4), 738–797. https://doi.org/10.1002/j.2168-9830.2012.tb01127.x

Driver, R., Asoko, H., Leach, J., Scott, P., & Mortimer, E (1994). Constructing scientific knowledge in the classroom. *Educational Researcher, 23*(7), 5-12. https://doi.org/10.3102/0013189X023007005

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, *84*(3), 287. https://doi.org/10.1002/(SICI)1098-237X(200005)84:3<287::AID-SCE1>3.3.CO;2-1

Dwarakanath, S., & Wallace, K. M. (1995). Decision-making in engineering design: Observations from design experiments. *Journal of Engineering Design*, *6*(3), 191–206. https://doi.org/10.1080/09544829508907913

Dym, C. L. (1994). *Engineering design: A synthesis of views*. New York, NY: Cambridge University Press.

Erduran, S., & Jimenez-Aleixandre, M. (2007). Argumentation in Science Education: Perspectives from Classroom-Based Research. *International Journal of Science Education* (Vol. 14). https://doi.org/10.1080/0950069920140508

Erduran, S., Ozdem, Y., & Park, J. Y. (2015). Research trends on argumentation in science education: A journal content analysis from 1998–2014. *International Journal of STEM Education*, *2*(1), 1–12. https://doi.org/10.1186/s40594-015-0020-1

Erduran, S.,, Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's Argument Pattern for studying science discourse. *Science Education*, *88*(6), 915-933. https://doi.org/10.1002/sce.20012

Fahnestock, J., & Secor, M. (2003). *A rhetoric of argument: Brief edition*. New York, NY: McGraw-Hill.

Fischer, G., Lemke, A. C., McCall, R., & Morch, A. I. (1991). Making argumentation serve design. *Human Computer Interactions*, *6*(3-4), 393-419. https://doi.org/10.1080/07370024.1991.9667173

Ford, M. (2008). Disciplinary authority and accountability in scientific practice and learning. *Science Education*, *92*(3), 404–423. https://doi.org/10.1002/sce.20263

Fortus, D., Dershimer, R. C., Krajcik, J., Marx, R. W., & Mamlok-Naaman, R. (2004). Design-based science and student learning. *Journal of Research in Science Teaching*, *41*(10), 1081–1110. https://doi.org/10.1002/tea.20040

Furtak, E. M., Hardy, I., Beinbrech, C., Shavelson, R. J., & Shemwell, J. T. (2010). A framework for analyzing evidence-based reasoning in science classroom discourse. *Educational Assessment*, *15*(3), 175–196. https://doi.org/10.1080/10627197.2010.530553

Hansen, C. T., & Andreasen, M. M. (2004). A Mapping of design decision-making. Proceedings of the 8th *Internetaional Design Conference*. Dubrovnik, Croatia (pp. 1409–1418).

Hatamura, Y. (2006). *Decision-making in engineering design: theory and practice*. London: Springer.

Hogarth, R. M., & Kunreuther, H. (1995). Decision making under ignorance: Arguing with yourself. *Journal of Risk and Uncertainty*, *10*(1), 15-36. https://doi.org/10.1007/BF01211526

Indiana Department of Education. (2010). *2010 Science and Engineering Process Standards (SEPS)*. Retrieved from https://www.doe.in.gov/sites/default/files/standards/indiana-chemistry-standards-2016-41116.pdf

Jiménez-Aleixandre, M. P., & Pereiro-Muñoz, C. (2002). Knowledge producers or knowledge consumers? Argumentation and decision making about environmental management. *International Journal of Science Education*, *24*(11), 1171–1190. https://doi.org/10.1080/09500690210134857

Jonassen, D. H. (2008). Instructional design as a design problem solving: An iterative process. *Educational Technology*, 21-26. https://doi.org/Related link: URL: <http://asianvu.com/bookstoread/etp/>

Jonassen, D. H. (2012). Designing for decision making. *Educational Technology Research and Development*, *60*(2), 341-359. https://doi.org/10.1007/s11423-011-9230-5

Jonassen, D. H., & Cho, Y. H. (2011). Fostering argumentation while solving engineering ethics problems. *Journal of Engineering Education*, *100*(4), 680–702. https://doi.org/10.1002/j.2168-9830.2011.tb00032.x

Jonassen, D. H., & Kim, B. (2009). Arguing to learn and learning to argue: Design justifications and guidelines. *Educatioanl Technology Research and Development, 58*(4), 439-457. https://doi.org/10.1007/sl

Kanter, D. E. (2010). Doing the project and learning the content: Designing project-based science curricula for meaningful understanding. *Science Education*, *94*(3), 525-551. https://doi.org/10.1002/sce.20381

Karplus, R., & Butts, D. P. (1977). Science teaching and the development of reasoning. *Journal of Research in Science Teaching*, *14*(2), 169-175). https://doi.org/10.1002/tea.3660140212

Kearney, M. (2004). Classroom use of multimedia-supported predict-observe-explain tasks in a social constructivist learning environment. *Research in Science Education*, *34*(4), 427-453. https://doi.org/10.1007/s11165-004-8795-y

Kolodner, J. L., Camp, P. J., Crismond, D., Fasse, B., Gray, J., Holbrook, J., … Ryan, M. (2003). Problem-based learning meets case-based reasoning in the middle-school science classroom: Putting learning by design[tm] into practice. *Journal of the Learning Sciences*, *12*(4), 495–547. https://doi.org/10.1207/S15327809JLS1204_2

Krajcik, J. S., & Merritt, J. (2012). Engaging students in scientific practices : What does constructing and revising models look like in the science classroom? *The Science Teache*r, *79*(3), 38. Retrieved from http://learningcenter.nsta.org/files/tst1203_38.pdf

Kuhn, D. (1993). Science as argument: Implications for teaching and learning scientific thinking. *Science Education*, *77*(3), 319–337. https://doi.org/10.1002/sce.3730770306

Leach, J., & Scott, P. (2003). Individual and sociocultural views of learning in science education. *Science and Education*, *12*(1), 91-113. https://doi.org/10.1023/A:1022665519862

Lee, O. (2017). Common core state standards for ela/literacy and next generation science standards: Convergences and discrepancies using argument as an example. *Educational Researcher*, *46*(2), 90–102. https://doi.org/10.3102/0013189X17699172

Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and decision making. *Annual Review of Psychology*, *66*(1), 799–823. https://doi.org/10.1146/annurev-psych-010213-115043

Littleton, K., & Howe, C. (2010). *Educational dialogues: Understanding and promoting productive interaction*. Routledge.

Lizotte, D. J., McNeill, K. L., & Krajcik, J. (2004). Teacher practices that support students' construction of scientific explanations in middle school classrooms. *In Proceedings of the 2004 International Conference of the Learning Sciences(ICLS).*

Longino, H. (1990). *Science as social knowledge*. Princeton, NJ: Princeton University Press.

Maloney, J. (2007). *Children's roles and use of evidence in science: An analysis of decision-making in small groups*. *British Educational Research Journal*, *33*(3), 371-401. https://doi.org/10.1080/01411920701243636

Mastro, G. (2017). *Claim, evidence, and reasoning: Evaluation of the use of scientific inquir scientific inquiry to suppor o support argumentativ gumentative writing in the middle riting in the middle school science classroom* [Doctoral dissertation, Dominican University of California]. Retrieved from https://doi.org/10.33015/dominican.edu/2017.edu.09

Mathis, C. A., Siverling, E. A., Glancy, A. W., & Moore, T. J. (2017). Teachers' incorporation of argumentation to support engineering learning in stem integration curricula. *Journal of Pre-College Engineering Education Research (J-PEER)*, *7*(1), 6. https://doi.org/10.7771/2157-9288.1163

Mcneill, K. L., & Krajcik, J. (2008). Inquiry and scientific explanations: Helping students use evidence and reasoning. *Science as Inquiry in the Secondary Setting*, 121–134.

McNeill, K. L., & Krajcik, J. (2008). Scientific explanations: Characterizing and evaluating the effects of teachers' instructional practices on student learning. *Journal of Research in Science Teaching*, *45*(1), 53–78. https://doi.org/10.1002/tea.20201

McNeill, K. L., & Krajcik, J. (2009). Synergy between teacher practices and curricular scaffolds to support students in using domain-specific and domain-general knowledge in writing arguments to explain phenomena. *Journal of the Learning Sciences, 18*(3), 416-460. https://doi.org/10.1080/10508400903013488

McNeill, K. L., & Krajcik, J. (2012). Framework for constructing scientific explanations. *Supporting Grade 5-8 Students in Constructing Explanations in Science: The Claim, Evidence, and Reasoning Framework for Talk and Writing*.

McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *Journal of the Learning Sciences*, *15*(2), 153-191. https://doi.org/10.1207/s15327809jls1502_1

McNeill, K. L., & Martin, D. M. (2011). Claims, evidence, and reasoning : Demystifying data during a unit on simple machines. *Science and Children, 48*(8), 52.

Mentzer, N. (2014). High school student information access and engineering design performance. *Journal of Pre-College Engineering Education Research (J-PEER), 4*(1), 4. https://doi.org/10.7771/2157-9288.1074

Mitcham, C., & Holbrook, J. B. (2006). Understanding technological design. In *Defining Technological Literacy Towards an Epistemological Framework* (pp.105-120). Palgrave Macmillan, New York. https://doi.org/10.1057/9781403983053

Moore, T. J., Glancy, A. W., Tank, K. M., Kersten, J. A., Smith, K. A., & Stohlmann, M. S. (2014). A framework for quality k-12 engineering education: Research and development. *Journal of Pre-College Engineering Education Research (J-PEER), 4*(1), 2. https://doi.org/10.7771/2157-9288.1069

Morse, J. M. (1991). Approaches to qualitative-quantitative methodological triangulation. *Nursing Research, 40*(2),120-123. https://doi.org/10.1097/00006199-199103000-00014

Newton, P., Driver, R., & Osborne, J. (1999). The place of argumentation in the pedagogy of school science. *International Journal of Science Education*, *21*(5), 553-576. https://doi.org/10.1080/095006999290570

NGSS Lead States. (2013). *Next Generation Science Standards: For States, By States (Engineering Design in the NGSS).* Washington, DC.

NRC. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. *American Physical Society*. Washington, DC. https://doi.org/10.17226/13165

Nussbaum, E. M., & Kardash, C. M. (2005). The effects of goal instructions and text on the generation of counterarguments during writing. *Journal of Educational Psychology*, *97*(2), 157. https://doi.org/10.1037/0022-0663.97.2.157

Nussbaum, E. M., & Schraw, G. (2007). Promoting argument-counterargument integration in students' writing. *Journal of Experimental Education*, *76*(1), 59-92. https://doi.org/10.3200/JEXE.76.1.59-92

Nussbaum, E. M., Sinatra, G. M., & Poliquin, A. (2008). Role of epistemic beliefs and scientific argumentation in science learning. *International Journal of Science Education*, *30*(15), 1977–1999. https://doi.org/10.1080/09500690701545919

Osborne, J. F., & Patterson, A. (2011). Scientific argument and explanation: A necessary distinction? *Science Education*, *95*(4), 627–638. https://doi.org/10.1002/sce.20438

P21: Partnership for 21st Century Learning. (2019). Framework for 21st Century Learning - P21. *P21.Org*.

Pahl, G., Beitz, W., Feldhusen, J., & Harriman, R. A. (2007). *Engineering Design: A Systematic Approach.* Springer.

Papadouris, N. (2012). Optimization as a reasoning strategy for dealing with socioscientific decision-making situations. *Science Education*, *96*(4), 600-630. https://doi.org/10.1002/sce.21016

Patronis, T., & Spiliotopoulou, V. (1999). Students' argumentation in decision-making on a socio-scientific issue: Implications for teaching. *International Journal of Science Education*, *21*(7), 745–754. https://doi.org/10.1080/095006999290408

Perkins, D. (1993). Teaching for understanding. *American Educator: The Professional Journal of the America Federation of Teachers*, *17(3)*(8), 28–35.

Rebello, C., Beardmore, D., & Towle, B. (2017). *Argumentation with contrasting cases : facilitation of deep structure learning.* Proceedings of the European Science Education Research Education 2017 Conference.

Rebello, C. M. (2019). *Scaffolding evidence-based reasoning in a technology supported engineering design activity.* Proceedings of the 13th Conference of the European Science Education Research Association (ESERA). Retrieved from http://par.nsf.gov/biblio/10105929%0A

Rebello, C. M., Barrow, L. H., & Rebello, N. S. (2013). *Effects of argumentation scaffolds on student performance on conceptual physics problems.* Proceedings of Physics Education Conference (pp. 293–296). https://doi.org/10.1119/perc.2013.pr.061

Rebello, C. M., & Rebello, N. S. (2013). *Transfer of argumentation skills in conceptual physics problem solving.* Proceedings of *AIP Conference*, January 2013 (pp.322–325). https://doi.org/10.1063/1.4789717

Roth, W. M., Tobin, K., & Ritchie, S. (2001). *Re/Constructing Elementary Science. Counterpoints: Studies in the Postmodern Theory of Education.* New York, NY: Peter Lang Publishing, Inc.

Sampson, V., Enderle, P., & Grooms, J. (2013). Argumentation in science education. *The Science Teacher*, *80*(5), 30–33.

Sandoval, W A, Crawford, V., Bienkowski, M., Hurst, K., & Millwood, K. (2003). Effects of explanation support on learning genetics . *In annual meeting of National Association for Research in Science Teaching*.

Sandoval, W. A. (2003). Conceptual and epistemic aspects of students' scientific explanations. *Journal of the Learning Sciences*, *12*(1), 5-51. https://doi.org/10.1207/S15327809JLS1201_2

Sandoval, W. A., & Millwood, K. A. (2005). The quality of students' use of evidence in written scientific explanations. *Cognition and Instruction, 23*(1), 23-55. https://doi.org/10.1207/s1532690xci2301_2

Simon, H. A., Dantzig, G. B., Hogarth, R., Plott, C. R., Schelling, T. C., Shepsle, K. A., … George, B. (1987). Decision making and problem solving. *Interfaces*, *17*(5), 11–31.

Sparke, P. (1986). *An introduction to design and culture in the twentieth century* (1st ed.). United Kingdom: Routledge.

Stark, R., Puhl, T., & Krause, U. M. (2009). Improving scientific argumentation skills by a problem-based learning environment: Effects of an elaboration tool and relevance of student characteristics. *Evaluation and Research in Education, 22*(1), 51-68. https://doi.org/10.1080/09500790903082362

Stauffer, L. A., & Ullman, D. G. (1991). Fundamental processes of mechanical designers based on empirical data. *Journal of Engineering Design*, *2*(2), 113-125. https://doi.org/10.1080/09544829108901675

Toulmin, S. E. (1958). *The Uses of Argument*. Cambridge: Cambridge UP.

Turiman, P., Omar, J., Daud, A. M., & Osman, K. (2012). Fostering the 21st century skills through scientific literacy and science process skills. *Procedia - Social and Behavioral Sciences*, *59*, 110–116. https://doi.org/10.1016/j.sbspro.2012.09.253

Ullman, D. G. (2001). Robust decision-making for engineering design. *Journal of Engineering Design*, *12*(1), 3–13. https://doi.org/10.1080/09544820010031580

Urueña, Y. B. P., Rebello, C. M., Dasgupta, C., Magana, A. J., & Rebello, N. S. (2018). Impact of argumentation scaffolds in contrasting designs tasks on elementary pre-service teachers' use of science ideas in engineering design. *Proceedings of Physics Education Research Conference* (pp.304–307). https://doi.org/10.1119/perc.2017.pr.071

Van Epps, A. S. (2013). Educating for evidence based decisions in engineering: The view as librarian and instructor. Proceedings of *ASEE Annual Conference and Exposition.*

Vattam, S. S., & Kolodner, J. L. (2008). On foundations of technological support for addressing challenges facing design-based science learning. *Pragmatics & Cognition*, *16*(2), 405–436. https://doi.org/10.1075/pc.16.2.07kel

Von Aufschnaiter, C., Erduran, S., Osborne, J., & Simon, S. (2008). Arguing to learn and learning to argue: Case studies of how students' argumentation relates to their scientific knowledge. *Journal of Research in Science Teaching*, *45*(1), 101-131. https://doi.org/10.1002/tea.20213

Vygotsky, L.S. (1962). *Thought and language* (E. Hanfmann & G. Vakar, Eds.). Cambridge, Ma: MIT Press.

Wadsworth, B. J. (1996). *Piaget's theory of cognitive and affective development: Foundations of constructivism.* Longman Publishing.

Walton, D. (2014). *Argumentation schemes for presumptive reasoning*. Psychology Press. https://doi.org/10.4324/9780203811160

Wilson-Lopez, A., Strong, A. R., Hartman, C. M., Garlick, J., Washburn, K. H., Minichiello, A., … Acosta-Feliz, J. (2020). A systematic review of argumentation related to the engineering-designed world. *Journal of Engineering Education*, *109*(2), 281–306. https://doi.org/10.1002/jee.20318

Wilson, A. A., Smith, E., & Householder, D. L. (2014). Using disciplinary literacies to enhance adolescents' engineering design activity. *Journal of Adolescent and Adult Literacy, 57*(8), 676-686. https://doi.org/10.1002/jaal.302

Wolfe, C. R. (2011). Argumentation across the curriculum. *Written Communication*, *28*(2), 193–219. https://doi.org/10.1177/0741088311399236

Wolfe, C. R., Britt, M. A., & Butler, J. A. (2009). Argumentation schema and the myside bias in written argumentation. *Written Communication*, *26*(2), 183–209. https://doi.org/10.1177/0741088309333019

Xie, C., Schimpf, C., Chao, J., Nourian, S., & Massicotte, J. (2018). Learning and teaching engineering design through modeling and simulation on a CAD platform. *Computer Applications in Engineering Education*, *26*(4), 824-840. https://doi.org/10.1002/cae.21920

Yackel, E. (2002). What we can learn from analyzing the teacher's role in collective argumentation. *Journal of Mathematical Behavior, 21*(4), 423-440. https://doi.org/10.1016/S0732-3123(02)00143-8

Yackel, E., & Cobb, P. (2006). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education*, *27*(4), 458-477. https://doi.org/10.2307/749877

Zannier, C., Chiasson, M., & Maurer, F. (2007). A model of design decision making based on empirical results of interviews with software designers. *Information and Software Technology*, *49*(6), 637-653. https://doi.org/10.1016/j.infsof.2007.02.010

Zeidler, D. L. (1997). The central role of fallacious thinking in science education. *Science Education*, *81*(4), 483–496. https://doi.org/10.1002/(SICI)1098-237X(199707)81:4<483::AID-SCE7>3.0.CO;2-8

Zembal-Saul, C., Munford, D., Crawford, B., Friedrichsen, P., & Land, S. (2002). Scaffolding preservice science teachers' evidence-based arguments during an investigation of natural selection. *Research in Science Education, 32*(4), 437-463. https://doi.org/10.1023/A:1022411822951

Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching*, *39*(1), 35-62. https://doi.org/10.1002/tea.10008