DESIGN-BASED LEARNING AND SECONDARY SCIENCE ACHIEVEMENT

by

Matiah Shaman

A Project Submitted to the Faculty of

The Evergreen State College

In Partial Fulfillment of the Requirements

for the degree

Master in Teaching

2013

This Project for the Master in Teaching Degree

by

Matiah Shaman

has been approved for

The Evergreen State College

by

Terry Ford, Ph.D., Member of the Faculty

June 2013

ACKNOWLEDGEMENTS

Thank you to my great teachers who struggled to be allowed to keep teaching the way their students needed, and thank you to the teachers stuck in fear and apathy who helped me understand why we need to keep learning and working to do it better. Thank you to my friends and family, who support me and inspire me with their love and dedication to justice.

ABSTRACT

This paper summarizes the literature about the efficacy of design-based learning in science classrooms. The large majority of the quantitative research directly investigating the efficacy of design-based learning on science content found that students did learn in design-based science classrooms. Three of these studies compared design-based learning classrooms with other classroom types, and found favorable efficacy. While a few studies showed that students were not able to learn science content in design-based classrooms, other studies showed particular methods of scaffolding that are effective, which included design journals and many iterations of student designs. This research indicated overall that design-based learning in science classrooms has the potential to be an effective way to motivate and organize science teaching, however, teachers need to very carefully create iterative learning experiences and consistent and comprehensive social and learning scaffolds. Future research should focus on identifying not just if particular design-based science classrooms are effective, but what makes them more effective than other pedagogies, and what makes them more effective than other design-based curricula. These studies may benefit from implementing more iterative designs in order to optimize scaffolds.

TABLE OF CONTENTS

CHAPTER ONE: INTRODUCTION

Introduction

In recent years, fueled by fears of slipping American technological dominance, a collapsing pool of engineers, and students who cannot apply their knowledge in ever-changing vocational settings and collaborative work, there have been many calls to create new curriculum that will prepare students for engineering careers, as well as for more flexible blue-collar technical careers, and along with other careers in math and science. The project-based, pre-engineering, and technical educational, and most recently design-based learning programs that have resulted have their roots in constructivism and John Dewey's philosophies of progressive education. These models, along with models such as problem-based learning, focus on collaborative design and construction, and on inquiry that is student-directed and teacher-scaffolded. Most of these programs include or claim to include traditional subject knowledge in math, science, and other fields within their inquiry and design processes, and also to teach critical thinking skills, independent problem-solving skills, and effective collaboration. However, studies have shown varied results on whether these models effectively teach basic science and math skills, or if particular implementation of these models more or less effectively teach these skills. This paper investigates the research into the efficacy of the models of teaching that include creative design practices in particular on science and math skills and knowledge.

Rationale

The rationale for project-based and design-based learning originates with the educational philosophies of John Dewey and the progressives of the early twentieth century, and extends forward to contemporary concerns that the United States is no longer internationally dominant in engineering and engineering education. These concerns culminated in the prominence of engineering design in the development of the Next Generation Science Standards and continued research indicating the usefulness of curriculum centered on student completion of a scientific inquiry process, a design process, or on the creation of other projects.

Dewey (1938) saw a problem with the compartmentalization of knowledge and learning that tended to go on in the public schools, claiming that it would lead to students who could not apply their knowledge outside of their classroom contexts, and who could not participate in Democratic collaboration with other workers and citizens. He advocated for a vocational education program in which students learned universal collaboration skills and the lessons of other academic subjects within vocational education, so that students would have flexibility in their future careers (Knoll, 1997). However vocational programs stayed largely segregated from other subjects in the United States until the 1980s. At this time, many began to see a problem with vocational programs that were largely based on the apprenticeship model; they examined the likelihood that future jobs would change constantly, and future citizens could no longer rely on job-training in a single trade for lifelong security. This led to calls both in vocational education and general education for new programs in which students learn the collaboration, critical thinking, systems thinking, and broad problem solving skills that enable them to thrive in a variety of technology-centric careers and pursuits (Lynch, 1997). For this reason, educators seek new models for education in which students can learn these skills.

Possible learning models that integrate real-world pursuits, traditional academics, and democratic collaboration include inquiry-based learning and problem based learning. In these learning models the goal is usually to find the answer to a question or solve a problem in the way that scientists solve problems. On the other hand, project-based and design-based learning models complement inquiry-based and problem-based learning models by directing students' goals toward creative design and construction of their own technologies and artifacts, and through that process learning knowledge and skills (Silk, Schunn, & Carry, 2009).

Meanwhile, engineers have recently claimed that the pool of future engineers is shrinking, and that it particularly lacks people of color and women. The centrality of engineering to the industrial economy of the United States, as well as to University academics, has meant that these concerns have garnered much attention (Lewis, 2004). Also, claims that students in the United States are falling behind in math and science and research showing that it is not effective to simply teaching more math and science have led to interest in new, more effective ways of teaching these subjects (Stone, Alfeld, & Pearson, 2008).

Responding to these concerns, the National Research Council (2012) included engineering practices alongside scientific practices in their framework for the Next Generation Science Standards to be finalized early in 2013. They argued for including both within science education because engineering design projects are one of the most important motivations behind scientific research, and the tools produced by engineering design allow scientists to examine natural phenomena effectively and precisely. Their framework requires that engineering design be integrated with scientific practices in K-12 education. Roth (1995) argued that it is impossible to separate technological design and science as professions, largely because scientists often design new technologies to enable their research, and engineers engage in scientific research to help them understand phenomena related to their designs. The National Research Council does not explain exactly how teaching engineering can enhance science education, but some insight might be gained from delving into the distant history of science. Conner (2005) argued that both the scientific concepts and the empirical method of the scientific revolution emerged largely from the private knowledge and methods of artisans prior to the middle of the 18^{th} century. He showed that these artisans developed the empirical method to help them improve their craft. Perhaps while students engage in their own creative design work, they can be taught the scientific method as a practical way of understanding and overcoming the problems inhibiting their real-world projects.

Learning models that may help students to learn science as a way of enhancing their creative design work and that integrate science with engineering include project-based learning, technology-education, and secondary pre-engineering curricula. Unfortunately, research into these models' efficacy is either very limited or has shown inconsistent results. Hmelo-Silver, Duncan, and Clark (2007) claimed that the related problem-based learning and inquiry learning models are not generally effective when unguided, but can be very effective for many learners when carefully scaffolded and well-taught. There has been less rigorous research into design-based learning. The research that has been done seems to show that similar to problem-based learning and inquiry, design-based learning practices often teach students more effectively than traditional science and math courses to use science and mathematics in problem solving and creative work. However, the results have been mixed in research investigating students' acquisition of basic math and science knowledge skills in these programs (Brophy, Klein, Portsmore & Rogers, 2008). In addition, it is not clear that simply teaching engineering alongside science actually enhances science learning. Therefore, it is necessary to delve deeper into not just whether these new teaching models work, but what kinds of design-based and problem-based models work, and in what domains. It is also necessary to understand how and why those models work so that they can be improved and more efficiently targeted. Doing so will allow for the realization of some of the most important philosophies of education initiated in the 20th century, along with the goals of the Next Generation Science Standards.

Historical Background

The history of design-based learning originates with the Italian architecture schools in the 17th century. While these schools initially employed mostly teacher centered lecture instruction, they soon found that students could not apply the information about architecture that teachers taught using these methods to actual architectural scenarios, and so began to give students challenging design problems, and institute design competitions among students. Teachers taught architecture through these design projects and so were able to emphasize the development of student creativity as well as technical knowledge (Knoll, 1997). Following this tradition, engineering in the United States in the early 20th century was taught as an art through student design projects (Crismond, 2001). Stillmore Robinson initiated schools in which engineers actually built their machines in addition to just designing them in order to understand the practical limitations of their designs (Knoll, 1997). In addition, designing instruments that were both effective and aesthetically pleasing was an essential part of the practice of scientists in the late 19th century and early 20th century (Resnick, Berg & Eisenberg, 2000). However, Crismond wrote that when large numbers of engineers were needed in World War II, engineering and engineering education were reconceived as systematic and deterministic so that this new pool of engineers could trained more efficiently. Resnick et al. wrote that science during this era also became more industrial and scientists began to mostly use more efficient industrially-produced instruments for their work.

Also in the early 20th century, progressive philosophers such as John Dewey (1938) began to argue for a reconceptualization of traditional education in public schools and elsewhere. Dewey argued that the contemporary division of different academic subjects and manual training in school, and the vast difference between the context of learning in schools and the context of application in the larger world, meant that students would not be able to apply their learning outside of the particular contexts of their subject-area classrooms. Thus, Dewey and his students, such as William Kilpatrick, created a philosophy that students should learn critical thinking skills and integrate various subjects in the classroom, through projects oriented by the interests of the student and facilitated by the wiser teacher (Arends, 1997).

However, Kilpatrick began to advocate a new definition of project learning, in which students should ideally direct their own purpose in creating the project, directing its planning,and judging the project. This conception began to devolve into a definition of project that included almost any activity in which students' directed themselves, and included almost no guidance by the teacher. Dewey rejected Kilpatrick's new definition, saying that students would suffer because they would not be able to plan and execute their activities; teachers should coordinate student efforts, making sure students' process continued without interruption, and that students continued to have valuable educative experiences during the process. However, partly in response to Kilpatrick's extremely free-form vision of the model, American educators began to largely reject the project method in general (Knoll, 1997).

The debate continued between Dewey's philosophy, in which culture and life skills would be taught through vocational training, and the older philosophy of vocational schools, in which these schools should teach students a specific trade by a very experienced craftsman. The latter philosophy, advanced in particular by William Prosser, won out when the United States Congress passed the Smith-Hughes act of 1917. This act created separate funding for vocational programs, and vocational teachers taught job-related skills completely separately from the rest of the curriculum. These teachers could demonstrate qualifications almost exclusively through extensive experience in their field, and with very little teacher training (Lynch, 1997).

In the 1960's project-based learning re-emerged in Germany and the United States as students sought teaching methods that did not enforce what they saw as unjust systems (Knoll, 1997). Educators created a method of learning called "discovery learning" that was based on the constructivist learning theories of Jean Piaget and Jerome Bruner (Arends, 1997). However, Roth (1995) wrote that these educators likely had an incomplete understanding of constructivism, and instead believed that students would construct ideas that would necessarily match an objective model of the real world, so students would discover scientific ideas if given appropriate scientific apparatus and objects. However, radical constructivists such as von Glasserfield acknowledged the implications of constructivism as an epistemology as well as learning process, in which students' ideas evolve and only ones that are helpful (whether or not they are accurate) survive. This means that students could not be expected to naturally discover scientific "truths," but would have to be led intentionally and systematically to a scientific way of thinking and understanding (Roth).

Since this time, several other approaches have emerged, also rooted in constructivist learning theory and project-based learning. At this point, these approaches have begun to diverge and the boundaries between them are rather hazy. Hmelo-Silver et al. (2007) described two main contemporary models called problem-based learning, and inquiry learning. They argued that these models are largely equivalent but have different historical roots. Problem-based learning was created as an alternative method for training medical students in which students were presented with a hypothetical scenario of a patient with particular symptoms and complications, and students were supposed to diagnose the patient's disease and figure out how to treat it. Problem-based learning also exposed students more to real clinical environments (Hmelo-Silver et al.). Arends (1997) described research that showed that problem-based instruction in medical training tends to have no effect on basic science knowledge relative to the traditional training models, but a more positive effect on problem-solving and clinical skills in future doctors over traditional methods. Kirschner, Sweller, and Clark (2006), however, cite evidence that this kind of learning actually has a negative effect on basic science knowledge compared to traditional training.

Hmelo-Silver et al. (2007) also described inquiry learning, in which students also engage in investigations collaboratively as they solve problems and in the process learn science content. Duncan and Clark wrote that in both inquiry learning and problem based learning, students direct their own learning and investigations to some degree, but this learning is strongly scaffolded by the teacher. In scaffolding, a term coined by Bruner, teachers often guide students through the specific tasks and processes that are necessary for students to complete their investigations, and provide their students with examples and directions for working and investigating (Hmelo-Silver et al., 2007).

Meanwhile, there have been increasing political and academic calls to rebuild vocational education that have led to many overlaps between problem-based instruction and similar methodologies, and the activities in vocational education classes (Lynch, 1997). In 1994, the National Assessment of Vocational Education reported that extensive industry experience in particular occupations did not improve vocational teachers' efficacy, although a few years of experience did improve their efficacy. Lynch also described a report by the Secretary of Labor in 1991 called the SCANS report, which emphasized that jobs are now changing rapidly and unpredictably, and will continue to do so in the future. This report emphasized the need to integrate work-place training with other school subjects, so that students would have the cognitive flexibility to enter multiple fields. Lynch claimed that many recommendations for vocational education reform emerged in the late 1980s and early 1990s, and tended to emphasize integration of curriculum requirements such as English, math, and science in vocational schools, partnerships with the private sector, and an ability to prepare students for higher education should they seek it. Additional reports recommended that vocational learning should be integrated into general school programs, shouldand teach students skills in problem solving, collaboration, and continued learning and investigation. These reforms are remarkably consistent with Dewey's initial philosophies of industrial arts education and education in general, of constructivist learning theorists, and of the problem-based, project-based and inquiry learning models that were developing. However, additional reports indicated that vocational teachers were not ready to make these changes, perhaps limited by their lack of educational training (Lynch, 1997).

As Lewis (2004) detailed, the technology education programs of the past have more recently been giving way to pre-engineering curricula. In these programs, students learn mathematics and science through designing and sometimes building engineering-type projects. Through these projects, students must use inquiry and problem solving to find answers, and use technology and design tools that had previously been used in technology education courses. However, Lewis claimed that new courses in pre-engineering are particularly appealing to the academics who largely design and implement these curricula for several reasons. First, engineers increasingly claim that the engineering pool is shrinking, and the United States needs more engineers in general and more minority engineers in particular. Secondly, Lewis claimed that pre-engineering is generally more palpable to academics and funders because the goal is white-collar careers instead of blue collar jobs that academics and funders may not value or understand technical blue collar jobs. While this model may include inquiry and problem-based instruction, those are included within the larger design process, thus distinguishing design-based learning from those models. There have been a number of pre-engineering initiatives in the 2000's. The most widespread is Project Lead the Way (or PLTW), which consists of a three course sequence in high school that exposes students to various kinds of engineering and related processes through a scaffolded development, and which attempts to teach science, mathematics, and other topics through the process (Brophy et al., 2008). On the other hand, design-based learning emerged in the early 1990's as its own model of science and math education that more closely resembles project-based and problem-based teaching than technology education, and that like the former teaching models, usually aims to teach science and math content rather than technology content (Apedoe, Reynolds, Ellefson, and Schunn, 2008).

Definitions

There are numerous terms for teaching models and structures, some of which overlap or are used interchangeably. These include project-based learning, problem-based learning, inquiry learning, the discovery method, constructivist education, design-based learning, engineering education, pre-engineering, technical education, vocational education, and industrial arts. Hmelo-Silver et al. (2007) argued that problem-based learning and inquiry learning are distinguished almost entirely by their roots, and both involve posing an open ended or ill-formed question or problem, planning a route to solve this problem, using scaffolding, research tools and to collaboratively investigate the problem and find a solution, creating artifacts that demonstrate student understanding, and then judging the students' processes and artifacts. This is strikingly familiar to Kilpatrick's vision of project learning. However, although students direct much of their own investigations in problem-based learning and inquiry learning, these are often highly scaffolded by the teacher, whereas Kilpatrick's project learning is almost entirely open ended and student-directed. The discovery method, created to follow Bruner's constructivist philosophy, tends to more closely follow Kilpatrick's student centered vision and has largely fallen into disfavor. All of these models (except Kilpatrick's) can be considered constructivist learning, because they are based on constructivist and social constructivist learning theories (Knoll, 1997). More recently, the term project-based learning has been used in a variety of ways. Marx, Blumenfeld, Krajcik, Blunk, Crawford, Kelley, and Meyer and (1994) seemed to define it largely in the same way as the above understanding of problem-based learning, in that the goal is an investigation of a problem or guiding question. However, it may be that project-based learning is broader than problem-based learning, encompassing problem-based learning and similar approaches, but directed more towards the design and creation of artifacts than on solving a problem (“project-based learning,” 2011).

On the other hand, in design-based learning and engineering education models, the goal is explicitly the design and/or construction of a final project that is either physical or computer-based (Brophy et al., 2008). In these programs, students may use problem-based learning and inquiry to figure out problems and learn skills and information to help in their design, but the problem is always a creative one. In particular, secondary engineering and pre-engineering educational programs emphasize engineering-like creative projects, and thus are well suited to include science and math content, however they do not always explicitly teach it. Modern technical education programs, including some vocational education and industrial arts programs, include many of the same principles, but tend to be targeted at blue collar careers rather than engineering (Lynch, 1997). However, these programs often use technology in other contexts as well.

However, in the last two decades, design-based learning has emerged as its own model of education in science classes rather than engineering or technology classes, although different researchers call it by different names such as Learning by Design, Design-Based Learning, and performance Project-based Science curriculum or pPBSc (Apedoe et al., 2008; Doppelt & Schunn, 2008). While it draws inspiration from technology education, design-based learning derives from project-based learning and problem-based learning, and some describe it as a category of project-based learning (Apedoe et al, 2008; Kanter, 2010). While some also describe it as a type of inquiry-based learning, Marx, et al. (1994) and Mehalik, Dopplet and Schuun (1998) carefully differentiate it from the “scripted inquiry” that they say is the most common implementation of inquiry-based science, in which student investigations are carefully planned in the curriculum to lead students to construct particular and predetermined conclusions about the natural world. While some authors avoid using the term design-based learning as a general term because Design-Based Learning in all capital letters is the name of a particular proprietary enactment of it, this paper will refer to all of these curricula types as (lower case) design-based learning, following the nomenclature of project-base, problem-based, and inquiry-based curricula. Additionally, this paper will use the terms design-based learning, design-based science, and design-based environments mostly interchangeably.

In all of these models, new educational programs are often created by the same researchers who study their efficacy in a research process called design-based research (not to be confused with design-based learning). This has been shown to be an effective process, and these researchers have found ways to remove some of their bias, but their involvement in the design of the programs should be taken into account when considering their findings (Wang & Hannafin, 2005).

Statement of Purpose

The purpose of this paper is to investigate the efficacy of design-based science courses on secondary students' learning. The two main aspects of this investigation include determining if and in what contexts is design-based learning effective for teaching students basic science concepts and skills, and what makes it most effective if it indeed can be effective.

Limitations

Because the different learning models described above are often ill-defined and overlapping, it is difficult to concretely describe a scope for this paper. While the initial focus of this paper was on all teaching that involves design, including pre-engineering programs, vocational and technical educational programs, and within other models such as problem-based learning, inquiry learning, and project-based learning. The research has emerged a salient category of curricula intended explicitly to teach science through design called design-based science in this paper. So, while this paper will investigate whether the design activities in other curricula are effective for fostering science learning, the main focus of this paper is on curricula that set out to systematically teach science through design. The paper will also include other research in implementing project-based learning that may be directly applicable to implementing design-based learning in order to provide useful recommendations for overcoming implementation challenges in design-based learning. In addition, the focus of this paper will be science (and in some cases math) content learning. This paper will not investigate how well design-based or engineering education teaches students skills in design or engineering. On the other hand, the paper will review ways in which design-based science can support learning in science content and skills that are not traditionally valued or supported by science education, yet have nevertheless been shown to be essential to the practice of science. While skills in operating scientific technology and instruments are important for practicing science, these are skills that may be able to be supported in a variety of ways, and which may not be exclusive to science. For that reason, this paper will not investigate the efficacy of design-based models or other curricula on improving technological or manual skills.

Summary

Project-based and design-based learning have a rich history spanning from the Renaissance to modern day engineering education. More recently, design-based learning has emerged as a new curricular organization that is worthy of investigation. Chapter Two reviews and critiques the research into the actual efficacy of design processes and design-based science curricula on improving science learning. Chapter Three provides a summary of the findings, classroom implications, and suggestions for further research.

CHAPTER TWO: CRITICAL REVIEW OF THE LITERATURE

Introduction

As discussed in Chapter One, there has been a strong push in the United States over the last quarter century for a more effective science pedagogy to keep this country competitive or dominant. This push, mirroring the one following the Sputnik launch, has led to findings that science education does not reflect a modern understanding of the scientific process or of the learning process. Therefore, new pedagogies have been re-introduced, including project-based learning and design-oriented technology education. Design-based learning environments emerged from other constructivist learning environments such as project-based learning, problem-based learning, and some forms of technical education as a pedagogy with the potential to motivate students to improve their scientific, social, and technological literacy, as well as to motivate their learning more effectively. This is a pedagogy that takes very literally the idea that students construct their own knowledge. It is also an environment in which students have opportunities to work closely together as a design team, similar to the teams of professional adult designers such as engineers, architects, and artists. Design-based pedagogy is particularly close to other project-based pedagogies such as problem-based learning. However, the essential difference is that design-based courses are organized around the goal of creating artifacts, whereas in problem-based learning, the content is organized around the goal of answering an essential question, and the creation of artifacts is only a part of that process. While this paper will not argue that design-based instruction is better than problem-based instruction, it will argue that design-based instruction can also be effective for many of the same reasons, and that it offers its own unique affordances. The goal of this paper is also not to show that design-based education is an effective way to motivate students merely to participate, except where that motivation actually improves student learning. The paper also will not attempt to argue whether or not design-based learning is effective for teaching students to design or engineer. Instead, it seeks to answer if and how design-based science can be effective for improving science learning.

Chapter Two reviews research relevant to these topics. The first section includes three studies of project-based science with conclusions that are relevant to implementing design-based science, and four studies of technical, career, and engineering that show the use and limitations of these curricula for science learning. The next section reviews literature directly evaluating existing design-based science environments' efficacy for science education. The first subsection of that section reviews three articles showing comparatively favorable efficacy of design-based science, the next subsection reviews five articles that provide evidence for learning in design-based science without comparison to other environments, the third subsection reviews three articles providing evidence that design-based learning helps students learn non-traditional science, and the final subsection reviews three articles providing evidence of design-based environments that did not support student learning. The next major section reviews four studies investigating the particular features of design-based science environments that may support student learning. The final section reviews four studies providing evidence for the efficacy of particular techniques in design-based science teaching.

Insights From Research in Related Learning Environments

Many of the challenges of earlier learning environments such as project-based, problem-based, inquiry-based, and technical education should be expected in design-based learning, and successful efforts to overcome those challenges may be transferrable to design-based learning. The following two sections draw insight from the research on project-based science, and technical and engineering education, in order to anticipate these challenges and solutions.

Investigations of Project-Based Science

This sub-section draws insights from three studies investigating project-based science environments in order to anticipate likely strengths and weaknesses of design-based learning shared by the model's predecessors. Krajcik, Blumenfeld, Marx, Bass, Fredericks, and Soloway (1998) investigated how urban middle-school students work and learn in a project-based environment, and found that these students engaged in science-like observation, but not science-like analysis and presentation. The context that they describe is similar enough to design-based learning that their results may be transferrable. Marx, et al. (1994) found that middle school teachers attempting to implement a project-based science curriculum largely interpreted the new curriculum through the lenses of their existing pedagogical ideas and struggled to give up control to students and often poorly timed transitions to less scaffolding, limiting the efficacy of these curricula. The challenges that the teachers in their study encountered are basic challenges of moving from a traditional pedagogy to one centered on students as a learning and investigating community, and so these challenges should be expected in design-based classrooms. Finally, Panasan and Nuangchalerm (2010) compare the efficacy of project-based science with inquiry-based science, although they do not clearly define or explain the differences between these environments. This study illustrates two common challenges with research in innovative learning environment including the difficulty in differentiating definitions of curricula types, and the difficulty in differentiating the results when success is defined as performance on multiple-choice tests. On the other hand, their study does indicate that different pedagogies compatible with modern constructivist education theory may often produce similar results.

In a series of qualitative case studies, Krajcik, et al. (1998) investigated urban middle school students' first attempts at project based science using an analysis of several types of data and found that students were capable of high-level thinking in this context, but often did not effectively move from scientific observation to scientific analysis and presentation of results without further scaffolding.

The authors collected data in the form of observations, interviews and classroom artifacts. In addition, they video recorded each class at least three times per week throughout the duration of the study. Observations focused on target students doing inquiry. The authors conducted five 30 minute interviews of each target student, in which they asked students about the efficacy of their group work process.

The authors began their data analysis by summarizing evidence of the students' inquiry process with regards to thoughtfulness about content, group process, and affect. They also collected relevant observations of the teacher's actions related to inquiry processes. The authors then wrote narratives for each case study student, dividing inquiry into the processes of question asking, designing and planning investigations, building tools and investigating and analyzing data, and presenting findings. Finally, the authors compared the case studies by creating hypotheses, verifying evidence of these hypotheses from the different studies in tables, and then refining hypotheses in an iterative process.

The subjects of this study included the students in two classrooms, each taught by a different teacher in the same independent school in a small city in the United States. One teacher had five years of experience and a Bachelor of Science degree, the other had 21 years of experience and a Master of Science degree. Both teachers attended professional development trainings prior to and during this study, although they did not have outside help in the classroom. The authors more intensely observed two boys and two girls from each class selected by teachers as having low-middle level science achievement in the past. Two of these students were Black, one was Asian American, and the other students were White. The authors did not provide additional background information about these students, although did provide narrative descriptions of these students' classroom personalities and behaviors.

The authors found that although students often asked appropriate scientific questions, students also often asked questions that would not lend themselves to scientific or educative explorations. Students did not spend very much time developing these questions or discussing their merit without teacher help. Students were often able to discuss controlling variables while designing experiments, but some students lack of science content knowledge impeded their ability to design effective experiments. Students made observations using drawing or writing, but did not indicate how their observations would help answer their questions, or what questions were driving their observations. Students did not discuss possible questions or experiments that they might pursue based on information from their observations. Students often did not systemically carry out their experiments, and worked hastily due to a perceived lack of time. Although students often reported their data in detail, they did not describe their thinking behind the conclusions that they made related to the data in writing or in group presentations. The authors attribute this partly to the fact that the teachers did not effectively scaffold or model the process of data analysis and interpretation, and did not encourage students to work together during the data analysis phase of inquiry.

This study provided a detailed and clear account of strengths and limitations of a particular classroom engaging in project-based science. Many of the conclusions of the article are likely to be transferrable to other project-based science classes in the United States, especially ones in schools and contexts in which students have limited practice carefully reflecting on their own questions and building their own analysis, and with doing analysis collaboratively. On the other hand, many of the conclusions about the challenges in this context would likely be different if the classroom or context had provided more support for these processes, or less support for other processes. The authors did not provide any data on the economic situation at the school, although from the fact that it was an independent school, and from other descriptions in the article, it appears that these students were relatively privileged, and a class in a lower economic status community might have many more struggles. However, the authors did provide a clear account of their analysis procedures and clearly traced their conclusions to the case studies and their general observations.

In a series of qualitative case studies, Marx, Krajcik, Blunk, Crawford, Kelly, and Meyer (1994) found that middle school teachers in a variety of settings near Detroit attempting to implement a project-based science curriculum largely interpreted the new curriculum through the lenses of their existing pedagogical ideas. These teachers therefore struggled to give up control to students and often poorly timed transitions. They described their methods for the study that I describe below in a separate paper (Krajcik, Blumenfeld, Marx, & Soloway 1994).

This study followed teacher implementation of instruction in a project-based curriculum, “What's in Our Water?” in which students were required to work collaboratively with students in other schools via email. The researchers organized 13 full day “work-sessions” during which the researchers provided theoretical knowledge and teachers provided real-life teaching experience to develop an implementation of the material. The researchers collected data for the case studies via interviews, informal conversations with students and teachers, video and audio-taped work-sessions and video recorded lessons in the teachers’ classrooms. The researchers continuously reviewed a database of this data, and came to conclusions that they summarized in hypermedia documents linked to relevant data. As they did so, they sought out new information from the teachers to confirm or disconfirm their interpretations. Research assistants and primary investigators then discussed interpretations and wrote the narratives included in the article.

The authors selected four middle school teachers from the 11 teachers participating in the University of Michigan project. These 11 teachers volunteered for inclusion in the project and taught in “a variety of educational settings.” The author described four case studies that represented a range of challenges in implementing the curriculum. The authors provided narrative descriptions of these four teachers, their classrooms, and the communities in which they worked. All of the teachers taught in Michigan near Detroit, in both economically struggling and very wealthy communities. The authors did not provide specific or consistent information about student demographics in these classrooms, although the authors described one as having 25-35% minority enrollment, one as having many African American students, and one as having many mainstreamed special education students. Only one of the teachers had experience with project-based science. Many initially described their need or desire for control over the classroom and the curriculum.

The teachers largely interpreted the new curriculum through the lens of their established teaching ideas, and adopted a mix between their usual methods and those recommended in the new curriculum. The hybrid between methods was often ineffective. Teachers resisted allowing students to collaborate when that meant that the students worked with their own incorrect ideas. Teachers had a hard time using technology “as a cognitive tool”, instead only using it as “an instructional aid”. The teachers’ knowledge of content significantly impacted their implementation of these curricula. All teachers struggled to meet deadlines and to meet district curricula guidelines in their project-based curricula implementation. Teachers struggled to balance control and scaffolding with student autonomy, and failed to transition from structured to independent work. Teachers’ collaboration with each other was essential for their continued motivation and progress with implementing the project based curricula. The authors wrote that their conclusions about teacher challenges in implementing project based curricula are largely consistent with documented teacher challenges implementing other open ended collaborative curricula.

In conclusion, this study’s credibility benefited from the authors’ clear explication of their data analysis procedures. The confirmability of the study benefitted from the authors’ extensive narratives of each case study and triangulations that they made between the case-studies. The transferability also benefitted because authors described the subjects’ classroom contexts in detailed narratives. However, the authors did not provide concrete demographics for the contexts.

In addition to the above two studies investigating challenges in implementing project-based science curricula, the final study in this section compares the efficacy of project-based and inquiry-based science. When they compared the efficacy of inquiry-based and project-based curricula in a quasi-experimental non-equivalent control-group study of fifth grade students in Thailand, Panasan and Nuangchalerm (2010) found no significant difference in the efficacy of these curricula on science process skills.

The authors examined student learning using a multiple-choice test of achievement, analytical thinking, and science process skills with 30, 20, and 20 multiple choice items, respectively. The reliabilities of these tests were 0.86, 0.81, and 0.82, respectively, their discriminatory powers were between 0.28 and 0.81, their difficulty indices were between 0.36 and 0.67. However, the researchers did not define these measures or the processes for determining them. The authors gave this test as a pretest and a posttest to two groups of 44 students: one group who used an inquiry-based curriculum, and the other used a project-based curriculum. The authors designed the lesson plans for both curricula. The authors analyzed the tests by calculating an “effectiveness index,” a measure of the ratio between the students’ possible improvement and their real improvement.

The subjects of this study were 88 students divided into two groups of 44 students, selected via cluster random sampling techniques. The total population pool of the research included 365 fifth grade students in nine classrooms in an elementary school in a large city in northeast Thailand.

The authors found that the effectiveness indices of inquiry-based and project-based learning were 0.6774 and 0.6781, respectively, indicating that both curriculum types were effective, and had no significant difference in efficacy.

Although this research clearly shows data analysis methods and reliability indices, improving the reliability of the results, the internal validity of the results were limited by a small number of subjects, although the researchers did use statistical methods to locate representative subjects. Furthermore, the researchers did not define inquiry-based or project-based curricula or provide sample lesson plans, limiting the external validity of their findings. Also, the researchers provided no information about the questions in their testing instrument, so it is not possible to determine the usefulness of the instrument or therefore the results. In addition, the very different learning context in Thai schools limits the generalizability of the results with respect to American educational efforts.

The conclusions of Krajcik et al. (1998) about student challenges in project-based environments are likely to be ones that some students will face in design-based classrooms, such as the conclusion that students often ask scientific questions but not ones that they can realistically answer, and that students do not systematically pursue their investigations or analyze data without explicit teacher scaffolding. On the other hand, many of the conclusions by Marx et al. (1994) about teacher challenges in actually implementing project-based learning should also be transferrable to design-based curricula because these were challenges with aspects of project-based learning that are shared with design-based learning. These conclusions included that teachers struggled to give up control of classroom organization and content when students were to pursue their own investigations, and that teachers found transitions difficult between the different classroom organizations in project curricula. On the other hand, the conclusion that these teachers failed to use technology as a student learning tool is likely to be ameliorated in design-based science courses in which students' learning is focused on designing technology.

On the other hand, the study by Panasan and Nuangchalerm (2010) did not provide useful conclusions about science pedagogy, but nonetheless served as a useful example of potential problems with research on these learning environments. These authors provided detailed statistical support for their result, but failed to analyze, motivate, or even describe the curricula that they studied or the tests that they used to study the curricula.

Studies of Technology Education Environments Related to Design-Based Science

Design-based learning is related not only to older constructivist pedagogies such as project-based learning and problem-based learning, but is also related to various conceptions of Technology Education (TE) and engineering education. However, while in the past industrial education such as that promoted by Dewey was to be integrated with content including math and science, in the late half of the 20th century, Technology Education was primarily intended to teach technology-specific and vocational skills (Knoll, 1997). Engineering education was also introduced into secondary schools in order to prepare and motivate students for engineering as a career path (Brophy et al., 2008). While some proponents of these curricula promoted the fact that the curricula uses ideas from math and science, or provides a hands-on experience of these ideas, the curricula are not systematically designed to actually teach science content. Roth et al. (2001) describe the failure of purely “hands-on” or even “hands-on, minds-on” curricula that expects students to learn science content simply because they explore scientific objects.

The four studies in this sub-section investigated students' academic outcomes related to taking Technology Education, Career Technical Education (CTE), and Engineering Education courses.   These studies shed some light on the promises and shortcomings of including Technical Education in secondary school, as well their shortcomings when Technology Education is not sufficiently integrated with academic courses. Stone and Aliaga (2005) begin with an analysis of National Longitudinal Survey of Youth (NLSY97) data in which they conclude promisingly that students in technology education had a lower achievement gap vis a vis academically tracked students than did general tracked students. On the other hand, Tran and Nathan (2010) found that students in a widespread pre-engineering program performed significantly worse on mathematics after correcting for student and teacher variables and prior achievement, and did not perform significantly differently in science achievement. Childress (1996) investigated the efficacy of correlating technology education and science classroom's curricula in a small study and found no significant difference between those who received correlated science and math instruction and those who did not, although the level of integration may not have been sufficient. In a qualitative interview study of high school and adult designers, Crismond (2001) found that high school designers did not use or have to use their science skills and knowledge to complete non-scaffolded design tasks and were thus limited in their solutions and learning, although experts naturally used science skills and knowledge in their working processes.

In their analysis of data from the National Longitudinal Survey of Youth (NLSY97), Stone and Aliaga (2005) found that youth participating in Career and

Technical programs had a lower achievement gap ($p < 0.05$) with academic tracked students than did general tracked students with academic tracked students.

The authors analyzed data for this study from the first five rounds of student interviews from the NLSY97. The authors wrote that this survey is nationally representative according to the Bureau of Labor Statistics and includes data from 9,000 12-16 year old students from December of 1996. The data reported in this study are from students who had completed ninth grade. The authors analyzed data by first weighing observations from the survey following the Bureau of Labor Statistics guidelines to control for oversampling and to estimate population parameters in order to estimate the numbers represented by each response. They then used logistic regression to compare independent and dependent variables.

The authors found that General Tracked students, compared with Academic Tracked students, had a significantly lower GPA with a Beta coefficient of ß = -0.136. However, Career and Technical Education (CTE) students finished with GPAs that were significantly lower than Academic tracked students with a smaller gap, with ß = -0.076 for CTE students. In addition, dual CTE and Academic tracked students performed significantly lower than Academic only tracked students, but with ß = -0.061, again a smaller difference than that found with General tracked students. All significant results reported were significant at the $p < 0.05$ level. In general, CTE students had a smaller gap with Academic tracked students than did the General tracked students.

The authors provided rigorous statistical tests on a survey with a very large and statistically representative sample of all US students, therefore ensuring very good internal and external validity. However, it is difficult to draw strong conclusions from their findings, because GPA could be affected just as much by different course difficulties or grading scales as by any real difference in learning. On the other hand, the result that CTE and Dual major students tended to have a smaller achievement gap does merit further investigation into the usefulness of programs that include some CTE content for students who may otherwise do poorly in high school.

In a quantitative study comparing students enrolled in a pre-engineering secondary program versus those not in this program in a large Midwestern city, Tran and Nathan (2010) found that students in the pre-engineering program performed significantly worse ($p < 0.05$) on mathematics after correcting for student and teacher variables and prior achievement, and not significantly differently on science sections of the state standardized test.

The study investigated students in Project Lead the Way (PLTW), which provides a middle school program called Gateway to Technology consisting of five nine-week courses “showing students how engineering skills including those from math, science, and technology are used to solve everyday problems” (Tran & Nathan, 2010, p. 6). The high school Pathway to Engineering curriculum provides three one-year “foundation” pre-engineering project-based and problem-based courses and several specialization-specific courses.

The results of this study were based on state standardized science and math achievement tests given to all students in eighth and 10th grades. Scaled scores and categories of performance were provided by the test maker and included advanced, proficient, basic, and minimal performance levels. The eighth grade math scale was 350-730, the 10th grade math scale was 410-750 the eighth grade science scale was 230-560, and the tenth grade science scale was 240-610.

The authors constructed dummy variables were for student gender, free/reduced lunch status, PLTW enrollment and eighth grade math and science achievement for each student. They also collected teacher years-of-experience and educational level data. The authors analyzed the results using a multi-level (student variables were level one, teacher variables were level two) regression analysis.

The student samples were from a Midwestern city with an urban population of over a half-million. The district enrolled more than 87,000 K-12 students. District demographics were: 49% female, 51% male; 57% African American, 22% Hispanic, 12% White, 4% Asian, 4% other; 72% free/reduced lunch; 18% had special-education services; 8% were English Language Learners. The sampled students were from the five high schools in the district implementing the PLTW Gateway to Technology curriculum. All teachers involved were directed to adhere to the same state content standards in math and science. Schools implementing this curriculum had to pay $120,000 for start-up equipment per classroom. Teachers had to participate in a two-week training. Students in PLTW courses were chosen with available complete eighth and 10th grade standardized test data. Researchers used a chi-squared analysis to demonstrate that the subsample of students without missing data could be expected to provide an unbiased demographic sample except with regards to gender. Of this group, the 70 students who participated in at least one PLTW program were included in this study. A comparison group of 70 students was hand-picked based on prior achievement in science & math, gender and free & reduced lunch eligibility.

A paired t-test showed that students had significant overall gains ($p < 0.01$) between eighth and 10th grade math achievement and science achievement. Enrollment in at least one PLTW course was associated with a 10.76-point smaller increase in math scores. After controlling for prior achievement on eighth grade standardized tests, teacher and student variables, the authors found PLTW enrollment to be a significant predictor of student achievement in mathematics with $p < 0.05$. Enrollment in PLTW was associated with a 1-point decrease in science scores, which was determined to be not-statistically significant ($p = 0.911$). The results do not support the hypothesis that pre-engineering courses improve science and math scores

The authors wrote that a definite conclusion could not be reached because the sample size was relatively small, although at least relative to the other articles referenced in this paper, their sample page was not small. In addition, the authors do not provide a reliability coefficient of the testing instrument. However, the rigorous statistical measures that the authors took improved the internal validity of their results. It seems that their results might indicate that the PLTW courses did not sufficiently integrate math and science content into their curricula.

Childress (1996) investigated the efficacy of correlating technology education and science classroom planning using a quasi-experimental control group design with a 17 student experimental group and a 16 student control group, and found no significant difference between those who received correlated science and math instruction and those who did not.

Childress (1996) gave a pretest to both control and experimental group after first iteration of instruction and problem-solving and investigated variance of the groups. The pretests and posttests tested the efficacy of a design solution in terms of milliwatts of wind power generated. During the second iteration of problem solving, the author randomly selected six students to interview individually with questions to see if students attempted to apply science or math to their design processes. For each group, the author recorded responses to the question of why students thought that their second solution would perform better than their first and categorized responses into five categories.

The subjects were 17 eighth grade Technology Education students in the control group and 16 with a closely matched schedule for the experimental group, who had their Technology Education curriculum closely corresponding to their science curriculum. The samples were described as convenience samples. Each class was taught by a different teacher, and although the Technology Education class's curriculum was coordinated in time, the teachers did not plan curriculum together.

Childress (1996) found that both the experimental and control groups increased the power they generated on average from the pretest to the posttest. However, there were no significant differences at the $p < 0.05$ level between those who received correlated science and math instruction with those who did not. Interviews revealed that the experimental group tended to consciously apply science ideas to the wind collector problem, while control group depended on teacher information and their experiences.

The very small sample size of this study significantly limits the internal validity of their quantitative results. The Technology Education teacher was also not part of the science teacher's interdisciplinary planning team, indicating that the temporal correlation of curriculum may have been only at a surface level of content subject coordination. The lack of demographic data also limits generalizability of the study. However, the author's clear description of theoretical basis of the research and the limitations of the study increase its internal validity.

In a qualitative interview study of high school and adult designers, Crismond (2001) found that high school designers did not use or have to use their science skills and knowledge to complete design tasks and were thus limited in their solutions and learning, although experts naturally used science skills and knowledge in their working process.

The author interviewed pairs of subjects, giving each three examples of either jar openers or nutcrackers and asked the subjects to do several steps of "investigate and redesign," or I&R. The author asked subjects to try to identify what each device was in step one, to group two of the devices as "like" and explained why they were alike in step two, use the devices to become familiar with them in step three (steps one through three were phase one). The author probed for understanding of mechanical advantage in step four, asked for a wish list of ideal features of this kind of devices in step five, asked subjects to design a scientific comparison of devices in step six, (steps four through six were phase one), and finally to conceptually redesign their favored device for step seven, and reflect on that process in step eight.

The author collected videotapes, recordings and subjects' writing and sketches as data for analysis. They then described case studies for one male and one female team for each experience level that were interesting and representative of the work for that experience level, as determined by an analysis of data from each step. The author also presented overall data from steps four and seven, based on scores of data in several categories described in the article.

The subjects were 16 same-sex pairs of similarly skilled designers. Novice and naïve groups had equal numbers of male and female pairs. Six pairs of naïve designers were high school seniors or recent graduates who were recommended by teachers as scientifically literate, not adverse to mechanical devices, and who had little to no experience in design. Six pairs of novice designers were high school students who were recommended as having completed two years of design-oriented technology education courses. These courses were recommended as successful programs highlighting engineering and design thinking and were not vocational schools. The authors described the locations of each of these groups but did not describe any other demographics. The six experts were a pair of male professors and a pair of female professors from MIT and Stanford/Santa Clara SU respectively, a pair of male industrial engineers and a pair of male inventors. All experts had ten years of experience and either doctorate degrees in their field or ten or more patents.

Through analyzing these data and case studies, the author had three main findings. The first finding was that the Investigate and Redesign task was an engaging task for most subjects. The second finding was that naive and novice subject learning was context and device specific. In other words, the subjects learned more about the particular devices' features than general ideas about how devices work. The third finding was that naive designers did not have to confront or correct their scientific or technological misconceptions. In general, although there were many opportunities to use science skills in these tasks, non-expert designers did not use them unless directly asked to do so, although expert designers consistently did use science skills in the task.

Strengths of this study included well-structured interview activities and effective grouping of diads, improving credibility. Additionally, the authors explained their analysis of case studies in detail and provided convincing coding of compiled strategies for all groups.

On the other hand, the authors did not discuss subject demographics beyond sex and education/occupation. Although the author provided data showing that the case studies were representative, the author did not provide any more support that theirs were the best examples. Additionally, the task only provided an opportunity for one iteration of re-design, in which subjects conceived of, but did not physically realize, their plan. Pedagogical implications are not entirely clear from this study, although the study does provide support to the idea that students need scaffolding during design tasks in order to use or learn science during these tasks.

The studies in this section generally conclude that although students may be more motivated to succeed academically when they take hands-on courses including technology education, they do not necessarily do better in science or math. Crismond (2011) in fact concludes that even those students experienced in technological design do not necessarily use science when designing outside of those contexts. While these studies indicated that design-based instruction was likely to be a good way to motivate student learning, they also provided the important conclusion that teachers and curriculum designers need to very carefully and systematically scaffold students towards actually using and learning science while doing design activities because students are not likely to do this on their own.

Evidence For and Against Learning in Design-Based Science Classrooms

After reviewing relevant results from research into related learning environments, it is necessary to see the results of actual design-based science on student learning. As stated in prior sections, the goal of this paper is neither to show the efficacy of design-based learning on improving motivation to participate, except where that motivation directly impacts science learning, nor is the goal to show the efficacy of design-based learning on improving students' skills in designing. The overall goal of this paper is to investigate the efficacy of design-based science on improving students' skills and knowledge in science. The fourteen studies in the following four sub-sections investigate the efficacy of existing design-based curricula on improving students' skills and knowledge in science (mathematics is also included in some studies). The first sub-section presents evidence directly comparing the efficacy of design-based learning to traditional curricula and finding design-based environments to be favorable. The next section presents evidence that students improved their traditional science skills and content knowledge significantly in design-based science environments, but do not provide evidence that they improved relative to other environments. The third section presents evidence that students learn skills and content in design-based science environments that are not valued or supported by traditional pedagogy. The final section provides some evidence that design-based environments do not successfully support student learning.

Studies Showing Relatively Improved Learning

The three studies in this sub-section provide some of the strongest support for the efficacy of design-based science on improving science content and skills, because these studies directly compare design-based science environments to more well established environments and find favorable results. It is notable that all of the comparative studies reviewed did find favorable results. The first study is particularly strong because it has a relatively large sample size of 587 and compares design-based science to an inquiry-based constructivist curriculum that was already shown to be very effective (Mehalik et. al, 2008).

Mehalik et al. (2008) compared the efficacy of a design-based course and a scripted inquiry course on learning in 587 middle school students with more SES students and African American students in the design group and found about twice the possible gain in the design group on science knowledge tests, significantly higher gains for male although not female students in the design group, and gains that were eight times higher for African American students in the design group than in the inquiry group. Silk et al. (2009) investigated the efficacy of a Learning for Design unit in facilitating domain-general science reasoning in a diverse and high-poverty school using a one group pretest/posttest design study and a non-equivalent control group design study, and found a significant ($p < 0.001$) improvement from pretest to posttest with an effect size of 0.58, and found that the design group had a larger effect than a control group at the same significance level. Mooney and Laubach (2002) investigated the impacts of a design-based curricula on middle school students' mathematical understanding in a non-equivalent control group pretest posttest design, finding significant increase in conceptual understanding in the experimental but not control groups at the $p < 0.05$ level.

Mehalik, Doppelt, and Schunn (2008) compared the efficacy of a design-based course and a scripted inquiry course on learning in 587 middle school students with more SES students and African American students in the design group. They found about twice the possible gain in the design group on science knowledge tests, significantly higher gains for male although not female students in the design group, and gains that were eight times higher for African American students in the design group than in the inquiry group.

The authors created the design curricula for this study, while the scripted inquiry curricula was based on prior research had had been well tested by many US teachers. The researchers gave pretests and posttests to test content knowledge in electric circuits that the inquiry group would go over. They gave these immediately before the respective curricula, and immediately after the four to five week design or inquiry units about the relevant topics. They analyzed classroom means and gains, and compared differences in these means based on demographic groups using paired t-tests.

The subjects included 587 students from 26 classes, taught by ten teachers who implemented the system design unit, and 466 students in 20 classes taught by five teachers who implemented the scripted inquiry unit. The teachers were self-selected based on who wanted to attend professional training through their district, and were paid and given academic credit for attending these trainings. However, the inquiry group teachers volunteered after a request, and were not offered any compensation. In the design group, 53 percent of the 587 students were designated low socioeconomic status. In the inquiry group, 32 percent of the 466 students were designated low socioeconomic status. Fifty four and 51 percent of the members of the design and inquiry groups, respectively, were female, 66 and 33 percent of the design and inquiry groups, respectively were African American, and 64 percent and 58 percent, respectively, received free or reduced lunch.

The authors found that the design group's mean gain of 16 percent was significantly greater than the inquiry group's mean gain of seven percent, with $t = 2.02$, $p < 0.01$. However, the design group started at a lower mean score of 29% versus the inquiry group's initial mean of 38%. To correct for statistical floor effects, the authors found that the design group increased their mean score by 22.5% of their possible improvement, while the inquiry group increased their mean score by 11.3% of their possible improvement, a difference that was significant at the $p < 0.01$ level. African American students showed a mean gain in the design group of 16, while the mean gain in the inquiry group was 0.02. This is eight times the improvement, and was found to be significant with $t = 2.05$ and $p < 0.01$. Mean gains for non-African American students were about twice as high in the design group, at 0.21, versus 0.11 in the inquiry group, and significant with $t = 2.06$, $p = 0.07$. Male students gains were also significantly better in the design group, with a mean gain of 0.16 versus 0.05 in the inquiry group, with $t = 2.04$, $p < 0.01$. Mean female gains were not significantly higher in the design group. However, these students did show equally high mean gains as male students, but females in the inquiry group showed similar gains.

Overall, the systems design-based curriculum showed approximately twice the effect on learning as the scripted-inquiry curriculum. It was correlated with particularly better improvement with male and African American students and was as effective for female students. This study has a relatively very strong general result that design-based curricula can be more effective than scripted inquiry. It has a strong internal validity because it used both pretests and posttests, and because it had very large sample sizes, and compared design-based curricula with a well-tested inquiry based approach. It shows this effect across several population groups, improving its generalizability. However, the external validity of the study is limited because the authors did not provide any reliability measures of their testing instrument (they instead provided sample questions).

Silk et al. (2009) investigated the efficacy of a design-based science unit in facilitating domain-general science reasoning in a diverse and high-poverty school using a one group pretest/posttest design study and a non-equivalent control group design study, found a significant ($p < 0.001$) improvement from pretest to posttest with an effect size of 0.58, and found that the design group had a larger effect than the control group at the same significance level.

Silk et al. (2009) gave tests and collected other data at the beginning of the unit halfway through eighth grade, and at end of year. The tests were multiple choice and related to the curriculum of each group, and also included six questions about science reasoning common to both groups from Lawson's Classroom Test of Scientific Reasoning. In addition, seven items related to science reasoning were included in the test for the design group, two of which were validated by the Third International Mathematics and Science study, while the research team created the others. The data was analyzed with a Wilcoxon test because the data was not normally distributed. The authors also analyzed the following factors in a MANOVA test: TerraNova reading scores, gender, race/ethnicity, special education status, and free or reduced lunch.

The subjects included 170 students (seven percent of students were excluded from study due to excessive absences or suspensions). Of these, 70% were from of an "underrepresented minority racial/ethnic background", 80% were on free or reduced lunch. These were higher proportions than the district as a whole. There were no English Language Learner students in the sample, and almost all subjects were either White or African American. The control group used an established curriculum called "MARDS", and the group was selected because the teachers successfully implemented this curriculum as by its designers. The control group demographics included fewer minorities and lower income students: 39.3% racial minorities (mostly African American as well), and 36.2% on free or reduced lunch. The participating teachers were familiar with this unit, having taught it at least once. They had taught middle school for at least five years, and had a limited understanding of electronics content or engineering design outside of this unit. The participating teachers attended two professional development workshops as part of this study.

Silk et al. (2009) tested these instruments for reliability. All 13 of the items given to the design group only, referred as Full Test, had a pre-assessment Cronbach alpha of 0.57, and a post assess alpha of 0.72, and a pre-post correlation of 0.46. The six items on both design and inquiry groups were called the Reduced Test, which had an alpha of 0.49 for pre-assess, 0.68 alpha for post-assessment, and had a pretest-posttest correlation of 0.24.

The authors found significant improvement in the experimental group, from pretest to posttests with an effect size of $z = 0.58$, which is a moderate effect size. The initial regression model only fit fairly well with $R^2 = 29\%$, but was significant with $F(5,142) = 12.98$, $p < 0.001$. The authors found that minority racial/ethnic background, special education status were both negatively associated with post assessment scores with beta = -0.49, and -0.62 respectively. Gender and economic background not significant with other factors held constant.

Both groups' gains were significant for the Reduced Test comparisons, although the design group had a larger effect size (0.49) than inquiry (0.39). Mean scores increased from 0.21 to 0.34 and from 0.28 to 0.40, respectively.

In conclusion, the author's provision of Cronbach alpha scores improved the reliability of the study, although only the posttest alpha is acceptable, which decreases the external validity. Detailed MANOVA statistics, along with detailed demographic information provides useful and possibly externally valid information about types of students that are may be most benefitted by these curricula. However, they did not reported a racial/ethnic minority and economic background correlation holding pre-assessment constant, which would have perhaps been more useful.

Mooney and Laubach (2002) investigated the impacts of a design-based curriculum on middle school students' mathematical understanding in a non-equivalent control group pretest posttest design, finding significant increase in conceptual understanding in the experimental but not control groups at the $p < 0.05$ level.

The authors assessed student understanding of curriculum-specific content through a quantitative analysis of open format conceptual questions in pretest and posttest. These ranked understanding at four levels: Sound Understanding (SU), Partial Understanding (PU), Misunderstanding (MU), and No Understanding (NU), and used a Wilcoxon signed-rank test to compare the distributions of two related samples by ranking the differences in pretest and posttest explanations. They found positive and negative ranks for the ninth grade physical science and eighth grade mathematics classrooms for experimental and control groups, and used a z-test to find effect size.

The ninth grade physical science curriculum was implemented in two experimental and two control classes in a suburban high school. While 47 and 53 students in the experimental and control groups, respectively took either the pretest or the posttest, 31 and 36, respectively, took both tests. The students in the experimental group included 12 males, 19 females, 23 White students, no Black students, one Native American student, one Hispanic student and six other students. The control group students included 15 males and 21 females, 29 White students, one Black student, two Native American students, one Hispanic student, and three other students. The eighth grade Pre-Algebra curriculum was implemented in one urban classroom for the experimental curriculum and in two control classrooms. Eighteen students in the Experimental class took the tests. Forty-two took one of the tests in the control classes and 28 took both tests. The students in the experimental class included eight males, ten females, five White students, seven African American Students, four Hispanic students, and two other students. The control group consisted of 11 male students, 17 female students, eight White Students, six Black students, one Native American student, eight Hispanic students, and five other students. The teacher was the same for each pair of experimental and control classes.

For the ninth grade physical science class, the experimental group had 19 positive ranks out of 31, while the control had eight positive ranks out of 36. The experimental group had two negative and six tied ranks out of 31, while the control group had nine negative and 13 tied ranks out of 36. The author found a statistically significant increase in conceptual understanding only for the experimental group ($p < 0.001$)

For eighth grade mathematics, the experimental group had seven positive ranks out of 18, while the control group had 13 positive ranks out of 27. The experimental group had one negative and 10 tied ranks out of 18, while the control had five negative and nine tied out of 27. The authors found that there was a statistically significant increase ($p < 0.05$) in conceptual understanding for the experimental, but not control groups, although both increased understanding.

Strengths of this study included a detailed description of demographics in all classes. Because the authors showed similar results in both urban and suburban schools, the results may be more generalizable. The pretest posttest control group design is effective for showing a difference in content understanding. However, the authors did not directly compare changes between control and experimental groups, instead only reporting that the experimental group improved significantly while the control group did not. Therefore, the conclusion that the experimental curriculum was effective may be valid, but no strong conclusion can be made about whether it was more or less effective than the control group. Although the total number of subjects was high, the results were divided by class so each conclusion did not draw from a large number of samples.

Although some of the studies appearing later in this paper provide evidence that some design-based science environments do not supports science learning, the studies in this section provided evidence that design-based science environments can support science learning, and can do so more effectively than other environments, even well-established research-based ones. The study by Mehalik et al. (2008) in particular provides excellent evidence for content learning because it studied a very large number of students including strong representations of low income and non-low income students and students of different racial groups, as well as because it used an effective inquiry-based curriculum as its contrast group. Finally, the study by Mooney and Laubach (2002) provided evidence of mathematical learning. Overall, these studies provide strong support for the hypothesis that design-based science can be particularly effective environments for learning in a wide range of areas important to science.

Non-Comparative Studies Showing Content Learning

While the previous sub-section reviewed studies that found design-based science learning to be relatively favorable, the five studies in this section did not have contrast groups so were only able to show significant improvement on tests, not relatively strong improvement. However, some of these studies did compare their results to those found by other researchers and found their results to be relatively strong. Most of these studies primarily tested student learning in science content using paper and pencil tests. For example, Apedoe et al. (2008) studied the effects of a design-based chemistry course using a paper and pencil chemistry test. However, Fortus, Krajcik, Dershimer, Marx, and Mamlok-Naaman (2005) also investigated whether some students could transfer their learning to the design of a new project. Doppelt, Mehalik, Schunn, Silk, and Krysinski (2008) also coded students' use of science concepts in their design projects in class. In addition, while the studies did not disaggregate for student demographics such as race or gender, Doppelt et al. and Fortus, Dershimer, Krajcik, Marx, and Mamlok-Naaman (2004) compared results between students who were traditionally "high-acheivers" and "low-achievers." On the other hand, Sullivan (2008) studied high-achieving, mostly male and about half-white 11 and 12 year olds, finding that they improved their systems understanding significantly after a three-week after school robotics program.

In a one group pretest posttest study, Apedoe, et al. (2008) investigated the impact of an engineering design-based science curriculum on students' chemistry concept knowledge in a variety of high school class types, finding that accuracy improved on all concepts ($p < 0.01$) based on factor analysis, with gains ranging from 14% to 21%.

The authors' assessment instrument used 24 questions from Mulford's Chemical Concept Inventory (CCI) and the American Chemical Society's Test Bank (ACS). The CCI questions were conceptually oriented multiple choice questions assessing scientific reasoning and common misconceptions. The ACS multiple choice questions were factually oriented and chosen because they were relevant to the curriculum focus of this study. The authors analyzed results using Factor Analysis with $p < 0.001$.

The subjects of this study were 380 high school students (grade 9-12) enrolled in spectrum science, general chemistry, or advanced chemistry courses. This was the subjects' first exposure to chemistry coursework in high school for all the students. Of the initial pool of subjects, 271 completed the unit and both the pretests and posttests. The authors did not provide any additional demographic information about the students. Five teachers taught the classes in this study, and attended professional development sessions related to teaching the unit. Three had prior experience with design-based learning and had previously implemented a version of the unit. Each teacher implemented the unit in at least two classes. The authors used a repeated measures ANOVA test, and used Cohen's d to determine effect size percentages.

The authors found that students had an overall 13% gain on the test with $F(1,270) = 27.65$, $p < 0.001$. Student gain on atomic interaction questions was 21% with $F(1,270) = 12.56$, on reactions questions was 12% with $F(1,270) = 7.60$, and on energy changes during reactions questions was 14% with $F(1,270) = 9.55$. The results by subject were all significant at the $p < 0.01$ level.

The internal validity of these results is bolstered by the large sample size, and the generalizability is improved by the variety of classes included in the sample. The fact that students showed significant gains at a high level of confidence also improves the conclusion of the study that the curricula was likely to be effective. However, the authors did not provide a reliability test for their instrument, limiting the external validity of their results. Additionally, they did not use a control group, so it is not possible to tell if the students performed better after this curriculum than they would have in a different curriculum.

In a quasi-experimental pretest/posttest design case study with an additional transfer task, Fortus et al. (2005) investigated if enactments of a Design-Based Science curriculum supported students' scientific content learning and whether they would transfer that learning to solving a realistic design problem. The researchers used ANOVA t-test with $p < 0.001$ to determine if learning occurred, and used a MANOVA regression analysis with $p < 0.05$ to determine transfer.

The researchers administered an identical pretest and posttest that consisted of 15 multiple choice items and three open-ended items with eight sub-items. They designed the tests to probe for understanding at low, medium and high levels of comprehension. The multiple choice items probed at low and medium levels, while the open ended items probed at the medium and high levels. Each multiple choice question and open ended sub-item was worth one point, with half points possible on open ended sub-items. The researchers compared the pretests with the posttests using a t-test.

Students worked in groups of four during the units' enactment, and worked in the same groups during the final transfer task. This task began by giving students design specifications, and required them to submit technical and concept drawings, a three-dimensional model, a justification of their solution, and a description of their steps. The teacher encouraged them to use human and textual resources inside and outside of school in creating their design and were given three days to work on it at the library. Coders rated these transfer task products along five criteria, and the score was then compared the score to the students' pretest and posttest scores using MANOVA regression.

The subjects of this study were 149 ninth and 10th graders in physical science classrooms in an industrial town near Detroit. Of these, 87% were white, 10% were Hispanic, 2% black, 1% Asian, and less than 1% Native American. Almost all these students were from blue-collar working families, and 13% were on free or reduced lunch. One of their teachers had no extra help in the classroom and had two years teaching experience and a Bachelor's of Science Education in Earth Science, while the other teacher had six years teaching experience, a Bachelor's of

Science in biology, and a Master's in educational administration. Neither teacher had previously taught using design-based learning or a related model.

Researchers ran a t-test to compare pretest scores of students who took the pretest and those who did not take the pretest to show that that it was unlikely that those groups differed. In this test, $t(116)=1.23$, $p = 0.22$. The mean pretest score was 7.9 ($\sigma = 2.9$), and the mean posttest score was 13.9 ($\sigma = 3.6$). A paired t-test revealed significant differences between these scores, with $t(101) = 16.8$, $p < 0.001$. The effect size was $z = 1.8$, indicating significant and noticeable gains in students' scientific knowledge.

Only the four classes led by the first teacher participated in the transfer task, and 66 students completed all three assessments, with only 49 clearly unique solutions (determined by teachers to not be copies) that were used in analysis. About two-thirds of students appeared to give full or partial responses indicating some transfer of knowledge. Scores on this transfer task were correlated with the posttest at $r^2 = 0.20$ and $p < 0.001$, while the transfer task was not well correlated with the pretest, with r $r^2 = 0.039$ and $p = 0.17$. A null-hypothesis test yielded a difference between correlations with $z = 2.06$, exceeding the 1.96 two-tailed significance level for $p < 0.05$, and indicating that the learning activities successfully supported the transfer task.

Overall, the authors created and described the statistics in careful detail, for example by controlling for mortality and providing significance levels before giving statistical findings. They described in detail the activities during the unit implementation and transfer task, and included an example solution and test items, improving verisimilitude. However, these authors gave little demographic description, reducing the generalizability and reader understanding of transferability of the study. The authors did not provide an alpha for their home-brewed testing instruments, which reduced the validity of the study.

In a mixed quasi-experimental pretest posttest designed and ethnographic case study, Doppelt et al. (2008) investigated the efficacy of a design unit on understanding of electricity concepts, and found that students in classes labeled “low achievers” did not improve significantly on test results, those labeled “high achievers” did improve significantly on test results, but the lower achievement group showed qualitatively better ideas, more generative thinking, and more detailed documentation, and that all students showed a high level of documentation.

The authors administered the pretest before any electronics instruction and administered the posttest after the last day of the 5-week unit. Two versions of these tests were assigned to students randomly, each of which consisted of seven multiple choice questions.

The researchers analyzed student portfolios for ethnographic study, and collected twelve sets of presentation transparencies as data. In particular, they analyzed the portfolios at nine required steps of the portfolio creation/design process for both classes. The researchers also observed 65% of class activities. Two researchers observed simultaneously and kept logs.

The subjects were from two classes, and were divided at the discretion of the school personnel to be “high achievers” and “low achievers”. There were 22 in the low achiever group, which was 41% minority, 55% male, and 50% low SES. There were 16 in the high achiever group, which was 25% minority, 38% male, 50% low SES.

The authors found that the high achievers gained significantly from the pretest to the posttest with $t = 2.24$, low achievers did not show significant improvement with $t = 1.49$, $p = 0.14$. The authors did not provide actual means except in bar graph form: I estimate from this graph that low achievers improved from 0.33 to 0.40, high achievers from 0.35 to 0.54. The authors also analyzed the results by demographic categories, finding that female improvement was not significant, with a change from 0.37 to 0.46, male improvement was significant from 0.32 to.43, and African American improvement was significant from 0.27 to 0.43. Non-African American improvement was not significant from 0.37 to 0.45. The authors set all significance levels at $p < 0.05$. Interestingly, from these bar graphs, it appears that every group did very similarly on the posttest, although they had quite varied initial scores. The difference in gains versus difference in final scores was especially striking with African American vs. non-African American students, indicating that the achievement gap was largely closed.

The authors included several findings from their portfolio analysis. They found that students often developed a more sophisticated understanding of parallel and series circuits through experimentation than their teachers had through professional development. They found that the lower achiever team actually showed better ideas, more generative thinking, and more detailed documentation. However, high-achieving team did more drawing and sketching of their process. The authors found that students were very involved in their ideas, to the point of having difficulty paying attention to peers during presentations because they kept working on their own projects. The teacher confirmed that non-engaged students became very engaged and attentive during this unit. An interesting finding that these students in the low-achieving group did better when assessed qualitatively, although did poorly on paper and pencil test, which is consistent with the results of other studies.

The teacher confirmation of some results improved the confirmability of the qualitative results, as did triangulation between portfolio evaluation and observation analysis. The confirmability of the study also benefitted from a fairly transparent and clear process of breaking down portfolio data and analyzing results. The qualitative analysis appears stronger than quantitative, but for final result, the comparison is reasonably believable, as is the result that certain students who do not do well on the final test did build significant knowledge. The qualitative results' validity was limited in particular by a low N, a small number of questions, and no Cronbach's alpha. A limitation of the ethnographic research was a lack of a coding or analysis scheme for observational data.

In a one group pretest posttest study, Fortus, et al. (2004) investigated student learning in ninth and tenth grade design-based classrooms in a small Midwest town using artifact analysis and pretest and posttest content data for each classes' content subject finding significant improvements ($p < 0.001$) for both high achieving and low achieving students that were unlikely to be caused by either mortality or test repetition.

The authors gave identical pretests and posttests for each of three consecutive design-based learning units. The tests consisted of multiple choice and open-ended questions at three levels of cognitive demand. The test for the Extreme structures unit, for example, had 15 multiple choice questions and three open-ended questions each with eight sub questions. The test had a maximum score of 23 points. The maximum score for Environmentally Safe Batteries and Safer Cell phones were 22 and 21 points, respectively. Each test scored by one of two raters, but five were scored by both. The authors compared the ratings and determined the interrater reliability to be greater than 99%. Seventy students completed both the pretest and posttest for Extreme Structures unit, 64 for the Environmentally Safe Batteries unit, and 56 for the Cellular Phones unit. Only those students' data were analyzed. Two tailed t-tests with $p < 0.05$ were computed for each unit to make sure that students who completed both tests did not score differently than those that did not.

The authors also collected or photographed artifacts such as posters and models collected or photographed during enactment. One of the authors was present for one third of the class sessions and observed. The authors developed scoring rubrics for each artifact consisting of checklists of which concepts students had considered in developing the artifact and whether science knowledge was used correctly. Artifacts were scored by either the teacher or both authors. Both groups scored two artifacts from each unit, and an exemplar model for each unit was analyzed in detail that had scores within one standard deviation above the mean so it could be considered a good but representative model.

The subjects of this study were students in ninth and tenth grade classes in the same school in a small industrial town near a large Midwest city. Almost all students were from working class backgrounds. Nineteen percent of the students in the school was on free or reduced lunch. There were a total of 92 participants with the following demographics: 86% white, 11% hispanic, 2% black, 1% Asian and 56% female. There was a 16% turnover including five students leaving and 10 joining the classes. The same teacher, who had three years full time experience and a bachelor’s degree in earth science education, taught all the classes studied here. He had no other experience teaching design-based learning or related curricula.

The authors found that students' content knowledge increased significantly during all three units. The effect size for the Extreme Structures unit was $z = 2.1$, for the Environmentally Safe Batteries unit was 1.9, and for the Safer Cell phones unit was 2.7, with $p < 0.001$. The pretest and posttest means were 7.9, 6.8, 4.0; and 14.7, 11.9, 11.2 respectively. The large change for the last unit was mostly attributable to students' nearly non-existent initial knowledge of wave content. Paired t-tests showed that it was unlikely that gains were due to test retake.

High achievers vs. low achievers' (those with pretests above and below the mean on each unit) scores were also analyzed separately, and both groups showed gains on all units. While mean gains are provided for these groups, no statistics are provided for these gains.

For the Extreme Structures Unit, the examplar model improved in virtually every category in ability to apply content knowledge to construction of the artifact.

Other artifacts also showed excellent ability to apply knowledge of science content and thinking. For example, the authors found that the exemplary poster for the Environmentally Safe Batteries unit demonstrated student understanding of central chemical and electrical concepts. Interrater reliability for the artifacts was determined to be above 99%.

The authors of this study were thorough in their discussion of demographics, and used thorough statistical methods to account for mortality, pretest influence on the posttest, and inter-rater reliability. However, the fact that they did not use a control group meant that it is not possible to know whether this curriculum was effective compared to other curricula, only that it did appear to impact student understanding.

In a mixed method quantitative one group pretest posttest case study, Sullivan (2008) found that 26 high-achieving, mostly male and about half-white 11 and 12 year olds improved their systems understanding significantly at the $p < 0.05$ level after a three week after school robotics program.

For the quantitative portion of this study, the authors administered pretests and posttests that consisted of five multi-part questions, including two fill-in the blank sub-questions that had correct answers, and three open ended questions with an associated rubric developed and published in the article to score these answers. Two raters scored the responses.

For ethnographic observations, the authors videotaped problem-solving sessions on days 8-13 of instruction and created descriptive logs. The unit of analysis was a shift in activity. The coding scheme that they developed was based on AAS report on scientific literacy and NRC's 1996 standards, consisting of thinking skills described by this report, most directly relating to computation, estimation, manipulation, and observation. The coding scheme also included science process skills such as hypothesis generation, control of variables, hypothesis testing, and evaluations of solutions.

The subjects of this study included twenty-six 11-12-year-olds (22 males and four females) at an intensive robotics course at a summer camp (Center for Talented Youth at John Hopkins) for academically advanced students. This camp accepted students who scored higher than 95th percentile on nationally-normed tests. Of those accepted, 14 students were White (12boys/2girls), and nine were Asian American. Two students were mixed White and Middle-Eastern, and one was White and Native.

The pretest to posttest data showed that course participants increased their systems understanding with $t = 22.46$, $p < 0.05$, $n = 21$. The tests were administered on the first and last days of the session. All students completed the pretest, four did not complete the posttest and were excluded from analysis. A graduate student also scored 25% of tests, showing an interrater reliability $k = 0.80$. The mean score on the pretest was 23.09, $\sigma = 4.14$, and on the on the posttest was $m = 25.82$, $\sigma = 4.04$, a significant improvement of $t = 30.04$ with $p < 0.05$.

Video analysis coding results showed interrater reliability of Cohen's kappa = 0.86. Of the eight skills evaluated, evaluation of solution and observation skills were used by all 26 students. In addition, 25 students used manipulation, hypothesis generation, control of variables, and hypothesis testing, 24 used estimation, 11 used computational skills, eight neglected to control variables, and eight utilized all coded skills.

Two main conclusions of this study were that students used scientific literacy skills in performing their tasks, and that students' scientific literacy increased. These conclusions have a higher internal validity given the analysis of interrater reliability analyzed and the description of tests based on established criteria. The authors provided clear coding methods and examples for ethnographic observations as well, improving confirmability. However, the authors did not provide a control group, so it is not possible to determine the relative efficacy of these methods. Also, the small sample size and narrow demographic profile of students reduces generalizability of the quantitative findings and transferability of the ethnographic findings to other group profiles. In particular, this study only included students with already high level of knowledge included, so it is not possible if this sort of curriculum would provide effective supports to less ready students.

The five studies in this sub-section continued to provide good evidence that students can learn science in design-based science classes. Of particular interest were the findings by Fortus et al. (2005) that students learned the science content and were able to apply it to new design projects (although ability in this familiar application cannot be a strong case for generally transferrable knowledge), the finding by Doppelt et al. (2008) that students used science content in their portfolios and that "low achieving" students showed better understanding and thinking through their artifacts than the "high achieving" group even when they did not improve significantly on paper and pencil test, and the finding by Fortus et al.

(2004) that both “low achieving” and “high achieving” students improved their content knowledge. Overall, these results indicate that design-based science often creates science content knowledge improvements that can be detected using traditional tests, but that traditional tests may not be sufficient for measuring learning in design-based environments for students labeled as “low achieving.”

Evidence of Non-Traditional Science Learning

In addition to the above studies that found student learning in traditional science content, the three studies in this sub-section found student learning in areas that are not strongly supported by traditional pedagogies yet are essential to science. For example, in a study with a very large sample size of students in diverse settings Kanter (2010) found overall significant improvements in biology content, but was surprised to find the highest effect sizes in the questions at the highest levels of cognitive demand. Modeling is a skill that many researchers now describe as an essential or even the essential part of the scientific process. Penner, Giles, Lehrer, and Schauble (1997) investigated a modeling-oriented design-based science curriculum for elementary students and found that students showed a similar level of model-based thinking as fourth and fifth grade students who had not experienced the design coursework. Resnick (1996) identified non-centralized systems thinking as an essential skill of twenty-first century scientists and found that students could improve their ability to understand these systems by designing computer simulations of them.

In a quantitative pretest posttest sample survey, Kanter (2010) investigated the efficacy of a performance Project-based Science curriculum (pPBSc) on student understanding of biology content in 37 diverse middle school classrooms, finding a significant ($p < 0.001$) overall content effect size and particularly large effect sizes at higher cognitive levels.

After designing the curriculum, the author selected 12 teachers based on their interest in the curriculum type and met with these teachers three hours per week during implementation. The author designed a pretest and posttest with several items targeting each of the first three levels of Bloom's taxonomy for each of two clusters of targeted standards-based content. Cluster one was about cells and body systems, and cluster two was about biological energy content. The test questions included 17 multiple choice and short response questions, which were equally weighted. The author analyzed gains using an Effect Size (ES) measure ($p < 0.001$), and found overall gains, gains by content cluster, and gains by cognitive level using paired t-tests. Effect Sizes were defined as the change in mean divided by the square root of the sum of the squares of the pretest and posttest standard deviations. The authors used a measure of reliability to delineate the cognitive levels, finding the reliabilities of the three content levels to be 0.4, 0.5, and 0.3, respectively.

The subjects of the study included 652 sixth, seventh and eighth grade students from 37 classrooms. They were 32.5% Latino, 32.5% white and 29.9% African American. In addition, 51.1% were in the state's low-income category.

Kanter (2010) found that the mean of the total score changed from 6.1 to 8.4 out of 17 points, with $t = 25.2$. Students exhibited the most correct prior knowledge at cognitive level 1, where the effect size was 0.85. The authors did not compare with a control group, but cited a study reporting an average effect size of 0.23 between any two middle-level years in science content. At the cognitive level three, the effect size was 0.51, at cognitive level two the effect size was 0.98, and at cognitive level one, the effect size was 0.35. The authors speculated that the difference in effect size between cognitive level one and cognitive level three may have been affected by ceiling effects at cognitive level one. The overall result of this study is that the pPBSc curriculum was effective for improving content understanding, especially at higher cognitive levels.

The biggest strength of this study was the large sample size and consistent curriculum, both improving its internal validity. The confidence level used in this study also strongly supported the significance of its conclusions. However, although the author provided split-half reliabilities for delineating the cognitive levels, he did not provide an overall reliability measurement of his test instrument, so it is difficult to tell if the results have external validity. In addition, the author did not provide a comparison with performance on a different or traditional curriculum, so while it appears that students learned from this curriculum and the author estimated that they learned more than in a "typical" class, it is not possible to tell if they learned more or less than they would have with a different curriculum.

In a mixed qualitative and quantitative control group study, Penner, Giles, Lehrer, and Schauble (1997) found that suburban first and second grade students in a design-based modeling course showed a similar level of model-based thinking as fourth and fifth grade students who had not experienced the design coursework, especially in an ability to discard perceptual qualities and analyze relevant functional qualities.

The first two authors assisted students in group work, taught one day of instruction, and videotaped all activity in the classroom. The third author participated in whole-class discussions and activities during a model revision process to encourage children to incorporate appropriate constraints into their models. The authors do not describe a method for their procedures analyzing the video footage. The modeling lessons took place over just three one-hour sessions.

For each of four models of an elbow, children were asked to rate on a Likert scale “How much do you like this model of how the elbow works?” from one to five. The students were assisted with representations of faces displaying the emotion associated with each likability response option. The authors also analyzed student’s justifications for the highest level of student justification, favoring functional justifications as higher than perceptual, and including several levels of perceptual justifications, with a possible score range from zero to eight. After evaluating the models, children were asked their opinions concerning three general questions about modeling: whether it is important to make one’s model as much like the real thing as possible, whether it is important to revise one’s model, and whether multiple models of the same thing might ever be useful.

The subjects were 48 children in two combined first- and second-grade classrooms in a suburban district near a small Midwestern city. The authors selected nineteen children representing a variety of levels of ability to be interviewed at the conclusion of the study, and also interviewed 13 second-grade and nine fifth-grade children who had no modeling experience for comparison.

During initial conversations with students, the researchers found that students believed that models that they had used in class, for example of the earth, did not have to be made of the same material as the real thing, and mostly just looked similar, or in some cases functioned similarly. The students believed that perceptual similarity was most important. The students also initially built models that looked like an elbow, and said that their similar appearance is why they were good models. Throughout the study of the curriculum, children created models that showed more functional similarity to elbows.

In analyzing Likert data of student opinions of different models during the post-modeling interview in a repeated-measures analysis, the authors found a significant group vs. type interaction with the groups of modelers vs. non-modelers with $F(6, 114) = 3.42$ and $p < 0.01$. The modelers rated a model without any constraints on motion, (not a good model) significantly lower ($p < 0.05$) than the grade two non-modelers, although grade four modelers rated it even lower than either at the same significance level. Modelers most highly rated the most functional model provided, although the authors do not provide a significance level for this.

The authors found that although the level of functional justifications for their answers largely corresponded to how functional the model they were analyzing were, there were aggregate differences in the highest level of response based on student group. When they applied a Kruskal-Wallis analysis of each student's highest level of response, the authors found that the groups provided different forms of justifications, with $H(2) = 6.84$ at the $p < 0.05$ level. Both the first and second grade modeling group and the fourth and fifth grade group made primarily functional justifications, however, the first and second grade group provided significantly ($p < 0.05$) more sophisticated justification with $H(2) = 6.67$.

In conclusion, the results indicated that the type of hands-on modeling activity that took place in this study is effective for improving young elementary students' understanding of scientific modeling, even after only three days of such activities and only two iterations of modeling. The study's credibility benefited from a detailed explanation of procedures and their motivations, and the internal validity of the quantitative results benefited from several forms of rigorous statistical analyses. In addition, the validity and credibility of the conclusions were bolstered by triangulation from several different forms of quantitative and qualitative data and analysis that all pointed to the same interpretation.

Resnick (1996) investigated how designing programming simulations in an after-school program affected high school students' understanding of positive feedback and centralized behavior in a qualitative participant-observer case study, finding that although students did not initially consider that behavior such as traffic jams could occur without centralized behavior, they developed several key non-centralized thinking skills after design investigations.

The subjects, 12 students from urban high schools near Boston, Massachusetts, came to the program at the Massachusetts Institute of Technology, where the author helped them select simulation projects, guided them in building these projects and observed their activity. The subjects were two thirds boys. The author wrote that they came from diverse ethnic background and that half were immigrants or first generation Americans, but did not elaborate more. The author saved all computer interactions and recorded all discussions on audiotape, although did not describe coding strategies or analysis strategies for interpreting these data.

The author explained that students indicated that they did not believe that centralized behavior, such as traffic jams, or termites creating wood-chip piles, could occur without some kind of external centralized action at the location of the behavior. Students were highly resistant at first to even trying to design decentralized behaviors to test. However, students found out through designing several simulations that their initial idea was incorrect. In fact, as they proceeded in designing and testing simulations, they became very uninterested in designing centralized mechanisms and were much more interested in investigating complex mechanisms for centralized behaviors. The author writes that the main changes in student thinking that he observed included the new ideas that: positive feedback can create useful structures, randomness can help create order, and a group can behave very differently from any of its parts.

In conclusion, the author’s use of audio-tapes and automatically recorded computer interactions improved the study’s confirmability, and his description of basic demographic information increased the study’s transferability. However, the lack of coding or analysis strategies decreased confirmability, and the lack of description of how the author selected the subjects decreased transferability. It was not clear how much of the design and programming work the author did, and how much the students did; how he taught them programming; or how they proceeded with their projects. This decreased the credibility of the research. In addition, the results from a small-group after school program may not be transferrable to a classroom environment with complex classroom management challenges and diverse learning needs.

Overall, the studies in this subsection provided strong evidence that students can not only improve their science skills and their understanding of traditional science concepts, but can also improve in areas of science that other pedagogies do not readily or traditionally support.

Evidence of Failure of Design Curricula to Teach Science

This sub-section summarizes studies that found a lack of efficacy of design-based science in engaging or teaching students in important science content or processes. The three studies in this section found some successes in the design-based classrooms that they investigated, but also found major areas of science in which the curricula failed to help students learn.

Barnett (2005) investigated if ninth grade inner city students learn meaningful science in design activities, and the logistical challenges of implementing design activities for these students. While Barnett did not find significant improvements in content knowledge for these students measured quantitatively, he did find that in many cases students provided evidence of the same knowledge while describing their artifacts that they failed to show on the paper and pencil test. This result is congruent with that of Doppelt et al. (2008), although Barnett did not pursue the qualitative results as systematically. Seiler, Tobin, and Sokolic (2001) found that although students were very engaged in discussing and constructing their models, teachers were not successful in guiding students towards methodical scientific thinking in a design-based science class in a school-within-a-school for students with historically low success. Resnick et al. (2000) investigated scientific investigations and found that students did not use a prescribed “scientific method” and mostly failed to follow through with data analysis, but they developed critical thinking about results, and were able to develop their driving question as they developed their investigation.

Barnett (2005) investigated if ninth grade inner city students learn meaningful science in design activities, and what were the logistical challenges of implementing design activities for these students. Barnett completed this investigation through a hybrid quasi-experimental control group comparison and pretest/posttest design with $p < 0.001$, and qualitative ethnographic description from classroom observations, student interviews, and attendance data.

To measure student understanding, the Barnett examined results of the district’s final physics exam for participating design students ($N = 25$) compared with non-design students ($N = 42$). Barnett also administered a pretest about physics concepts central to the unit before the design curriculum and an identical posttest afterward. The author also took field notes during classroom observations and coded conjectures about these notes that the author then formed into hypotheses about the investigation. At the end of the study, the authors retroactively analyzed the hypotheses and observations to generate final conclusions.

The design curriculum was implemented in a 350 student inner-city school in Boston with 85% of students from racial or ethnic minority backgrounds, 13% ESL students, 20% students with documented disabilities, 28% living in federally designated “empowerment zones”, 14% single parents. Typical attendance rates during classes in the time period investigated were about 50%. Although 38 students began the project, 25 completed it. One physics, one math, and one special education teacher with 30, 29, and 31 years teaching experience, respectively volunteered their classes for the project, although the author primarily investigated the physics teacher’s class for the study.

Although Barnett did not include statistical measures, he wrote that attendance rose dramatically and supported this with a graph of attendance of the design and comparison groups showing that both began at about 50% attendance, and the design group rose to 85% while the comparison group rose to 55%, indicating that the design curriculum improved student interest. The author also included a portion of an interview in which a student explains that he or she began going to class because it was fun, had interesting physics content, and was relevant to his or her goal as a future electrician. The author interpreted the increase in attendance as related to the student ownership and authenticity of the project.

Barnett did not show a significant difference in the final exam scores of the design group over the comparison group (and did not include data to support this null-hypothesis). However, the author described qualitative cases in which students showed understanding of a concept in their verbal descriptions of their robot, but got a related question wrong on the abstract and decontextualized final physics exam. This indicates that design activities may create content understanding even when this understanding is not evident on standardized test scores, and that may need to be assessed in alternative ways.

Finally, the author noted several important affective attributes of the students during these design activities. For one, when students were given the opportunity to paint their designs, they abandoned all efforts at improving the functionality of the designs. Second, while the students were highly resistant to Socratic methods of teaching, insisting that the teacher told them what to do precisely, their engagement and creativity improved significantly when the teacher gave a mini-lecture on how difficult he found the design process of this project himself, and how much he did wrong at first and how important those errors were to his learning.

In conclusion, the triangulation between various quantitative and qualitative data helps support the qualitative conclusions that the author makes about the investigation. However, the author should have presented actual data for the null-hypothesis regarding state test results and should have provided statistical data for attendance results, because the graphs were convincing but not analytical. The small sample size decreases the internal validity of the quantitative results. Also, the author’s failure to describe his methods of analyzing qualitative data decreases confirmability.

In an ethnographic and interview-based classroom case study investigating the qualities of emerging culture and effective scaffolds in a design-based science class in a dominantly African American school-within-a-school for historically low success students in Philadelphia, Seiler et al. (2001) found that although students were very engaged in discussing and constructing their models, teachers were not successful guiding them towards methodical scientific thinking using verbal scaffolding techniques.

The authors used as classroom data audiotapes of verbal interactions created with a tape recorder near group tables as students worked, videotapes of the whole class during all periods, posters that some students made, and a student writing on a handout. Some 3D models were obtained as artifacts for study and used during follow up interviews. Teachers and researchers recorded and wrote down observations as well. Researchers and the teacher met after each class to discuss observations.

The authors did not describe a coding strategy for analyzing recordings and interviews, but instead repeatedly listed these data and used a situated analysis of the data, referring back to their initial questions. They investigated verbal utterances and physical motion and interactions for evidence of primary discourses and those of science and technology. In particular, they looked for patterns of coherence and contradiction, and of resistance. They analyzed the interviews after their classroom analysis and did so in a similar manner. They focused their interview analysis on data from two very verbal students.

The subjects for this study included 33 students in Incentive, a Small Learning Community (SLC) at City High School in Philadelphia. Almost all students reported to be African American, from poverty, and had been discharged from other SLCs or high schools in Philadelphia for nonattendance, academic failure, or violence. About 15 students on average attended each day. A primary teacher and a university science educator who had been teaching in the class for three months prior to the lessons taught the class.

Overall, the author found that students were very engaged in discussion and construction of their models, but that this engagement did not lead to practices readily identifiable as science or discernible as applying science concepts. Posters of the models focused entirely on accounting how the model was made, with no systematic record keeping, planning, modeling, or variable identifying or manipulating. Although teachers attempted to use verbal scaffolds to encourage inquiry discourse, this discourse did not develop in the classroom.

The authors also analyzed their results from a framework of “respect” as the primary currency of these students, and found that students frequently focused on using teacher attempts to work with them as opportunities to disrespect the teachers. They also found that student conversations about the cars centered on using other students carts to giv, and mostly to take away respect, for example after a car performed poorly or a student insulted it. They also found that students displayed an intense sense of ownership over their cars

The authors made two main conclusions about classroom practice. One was that many students rejected the idea that they should or could construct scientific understandings from their activity, and they also found that whole-class class discussions were the least helpful due to difficulty in keeping everyone's attention.

In interviews, the authors found that Kisha was able to elaborate a lot of meaning and knowledge about her model, but only when she had it in front of her to manipulate and gesture with. She could represent her knowledge to a much greater extent than would be possible without the model. In this interview, Kisha was also willing to work through predictions and hypotheses, something that she was not willing or able to do in class.

The authors recommended that these activities should incorporate the currency of respect in the classroom practice and use a more culturally appropriate model than drag-racing. I recommend trying non-verbal forms of scaffolding to help students generate scientific thinking.

In general, the authors explained very clearly the researcher and teacher experiences and roles and clearly described general student characteristics as well as the data sources, all of which improve confirmability and transferability. Their use of many data sources and two researchers to triangulate findings also improved confirmability. The authors' explications of the cultural context of the school was helpful for determining limitations of transferability, but indeed showed very limited transferability due to the unusual level of poverty and cultural conflict of this class setting.

Resnick, Berg & Eisenberg (2000) studied what scientific investigations look like with student-designed tools in an after-school program for fifth graders in participant-observer ethnographic case studies, finding that students did not use a prescribed "scientific method" and mostly failed to follow through with data analysis, but they developed more critical thinking about results, and were able to develop their driving question as they developed their investigation.

The author did not describe any method of taking or analyzing data. Instead, they only presented a narrative description of the investigation. The subjects that were the focus of this study, who the authors describe as chosen because they were interesting and representative cases, include Jenny, an 11 year old girl studying in a Boys and Girls club. The third subject was Alexandra, a fifth grade black girl. The third subject was Adrienne, a female Undergraduate at University of Colorado. The results from the second subject are not described here because they are not of a child and therefore not relevant here.

Jenny created a bird feeder that also monitored bird activity. Most of her feeder was physically and cognitively transparent so she could see what was going on. She continually re-evaluated and re-built her feeder, as well as the problem that she was attempting to solve, and extending its functionality.

Alexandra designed an interactive robotic art project involving balls rolling down ramps and a robotic controlled conveyor-belt. She did not use a self-prescribed “scientific method” involving a hypothesis or prior research. She tested different angles, timings, and programming methods to learn about acceleration on ramps. She found the science to investigate as she designed her projects.

In general, the authors found that students are able to do investigations by designing their own instruments that they could not with store-bought instruments. The authors also found that students develop more skepticism of instrument results when they built their own instruments. The authors found that project’s driving question evolves along with the artifact in these open-ended design-projects, allowing students to somatically investigate scientific principles as they encounter them. On the other hand, the authors found that students often succeed with one part of the project, but not in putting the parts together, and that many students did not follow through with data analysis after they built their projects and took data with them.

The authors did do some effective triangulation from multiple case studies, references to other cases by the same group, and reference to prior research, improving credibility. While the overall conclusions are believable, the authors were not transparent about their method, which decreased credibility. The author did not answer questions such as what was the environment of these students' investigations, and how much guidance were the students given. The study also provides no information about any gains in students' content knowledge.

In this subsection, Barnett (2005) and Seiler et al. (2001) found that while it can motivate students to participate, design-based curricula may not be effective for helping inner-city students learn science content and skills. The studies largely found that teachers were not able to scaffold students towards all important aspects of scientific thinking in these settings. However, studies in prior sections did show science learning for inner-city students, so it is possible for design-based learning to be effective for inner-city settings. The setting investigated by Barnett was so extreme and unique in its challenges that no curriculum would be expected to be successful without a carefully designed accompanying classroom management system that was lacking in this study. Resnick et al. (2000) found that more affluent students also failed to learn science process in a design-based class that did not explicitly scaffold such learning. It is notable that the teachers in none of these settings provided systematic scaffolding for developing science process and skills such as design or research journals.

Important Features of Design-Based Science Classrooms

The above sections analyzed whether and under what circumstances students learn science content and skills in design-based science classes. The two main goals of this paper are to determine whether and when students learn science in design-based science classrooms, and how these classrooms could be improved so that students learn more. In order to accomplish the latter goal, the four studies in this section go deeper in understanding the peculiar features of design-based science that can be used to improve student learning.

Roth (1996) investigated the elements of the learning environment of a fourth grade Canadian French immersion class that contributed to the design process and artifacts as well as the ontological status of identified elements, and found that children's design was most related to artifacts, tools, materials, teacher-set constrains, and trends in the setting, but that these elements' flexible ontology was key to the success of the design process. In a later study, Roth (2001) found that seventh and eighth grade students were able to employ gestures, collaborative visual representations, and design artifacts as critical components of a community cognitive system before they were able to verbalize the entirety of their understanding in a design-based class, and that it made no sense to differentiate between concrete and abstract thinking activities in this context. Doppelt & Schunn (2008) found that homework and instructional worksheets were the most influential activities for both inquiry-based and design-based science classes, and that other factors were rated similarly between the groups according to factor analysis of student survey data. Schauble, Klopfer, and Raghavan (1991) investigated how fifth and sixth graders change their models of inquiry during science and engineering oriented experimentation tasks and found that students developed scientific thinking more effectively when they started with engineering thinking tasks, as they would in design-based courses.

In a participant-observer ethnographic case study, Roth (1996) investigated how elements of a design-based learning environment contributed to the learning process and products, and what were the ontological status of those elements during the learning process. In doing so, this author investigating the entire design-based classroom as a single cognitive system and attempted to explain that cognitive system in terms of how its elements were used and modified during student learning and creating.

The author and graduate students engaged in direct observations during all class periods. Additionally, they recorded video of design-related activities from two continuously running video cameras, and took field notes and conducted video recorded interviews with students as well as regular and visiting teachers, and used students' engineering logbooks as well as sources of data. Researchers transcribed videos as soon as possible after they were recorded. The author generated assertions that attempted to answer the research questions during regular meetings with the research team, and the team tested all the data for each assertion's validity. The team then discarded, modified, or retained each assertion.

The subjects of this study were 23 fourth grade and five fifth grade French immersion students in Canada. They were mostly middle-class but represented a large range of socio-economic statuses. The class was team-taught by one female and one male teacher. The former had 12 years of experience while the other had three, and these teachers planned activities together.

The researchers observed that after students completed each artifact, specific groups or individuals received credit for it, however, evidence also showed that the design processes were situated and enabled just as much by other individuals and the context of the many elements of the environment as they were enabled by the individuals who created the objects. The author further studied the elements that were important in the development of these artifacts. They found that although the elements were generally present during the entire class, they were relevant for different parts of the design history for different people. For example, once one group effectively used a glue gun to make superior joints in their project, other groups adapted the technique and therefore included the glue gun as an integral part of their system of creating. At the same time, the students outside a particular design group who began a useful technique or made an influential suggestion or idea, became important actors in others' design project. As students created their designs, their thinking was embedded in their artifacts which were thus "reflections" of their thinking, and at the same time the students used the artifacts to create new and different ideas, and identify constraints on those ideas, so the artifacts served to foreshadow future thinking.

In general the authors found that children's designing was related to elements including artifacts, tools, materials, teacher constraints, and trends in the setting. The researchers found that the fact that the status of those elements were flexible, and that students frequently began using one element as a different element, to be essential to the design process.

Overall, the author described his method in great detail, including how data was used to create conclusions and how those conclusions were synthesized into a coherent theory of the design process. Also, the author cited very clear theoretical philosophical and cognitive foundations for this work. The author laid out the philosophical development of this new theory elegantly. However, the author presented very little demographic data for these students, indicating that the conclusions about ways of learning and working in a design-based learning environment were probably skewed towards cognitive patterns of the dominant cultural group.

In a qualitative classroom case study employing a wide variety of data including constant recording and strongly triangulated analysis, Roth (2001) found that seventh and eighth grade students were able to employ gestures, collaborative visual representations, and design artifacts as critical components of their cognitive system before they were able to verbalize the entirety of their understanding in a design-based class. He concluded that it made no sense to differentiate between concrete and abstract thinking activities in this context.

The students in this study were tested in written conceptual tests and in practical design activities before, during, and after the unit. The author taught the unit and used two video cameras simultaneously to record student and teacher activities throughout the unit. Two graduate students observed each class and engaged in post-lesson debriefing meetings. These graduate along with the author generated assertions about the curriculum, which were used to adjust the curriculum afterwards. The author also entered artifacts from teaching and from children's activities into the database.

A team of four documented and analyzed “knowing and learning” in the classroom in analysis sessions, which were themselves videotaped to capture emerging categories that would be useful for future analysis. The author wrote that more details about the enactment of the class and experiment are included in other papers, writing that his team published seven papers about this teaching experiment. This article includes three representative “episodes” of the author’s conclusions about learning and knowing.

The author taught this four month unit on simple machines in a sixth and seventh grade class. The subjects included ten sixth grader students including five boys and five girls, 16 seventh graders including seven boys and nine girls. One seventh grader left midway. The school guidance counselor described this as the hardest class he's worked with in his 21 year career. Four students were learning disabled, including one student with Attention-Deficit Hyperactive Disorder. One student had muscular dystrophy. Two seventh grade students were English as a Second Language students, three seventh graders had social difficulties and students refused to work with them. Three students had family problems that affected their academic work. A two by two (boys, girls; grade six, grade seven) MANOVA test did not find any statistically detectable effects for gender or grade using marks on final project, written examination, and discursive competence as dependent variables.

The author's findings represented the conclusion that students' cognitive and externalized representations of ideas were closely intertwined, and abstract and concrete relations were different parts of the same cognitive representations. They found that deictic gestures (pointing) played an important role in identifying elements and elements' location in systems, iconic gestures animated still drawings and showed the direction, for example, that different parties had to pull. The author noted that these gestures should also improve learning by creating verbal as well as sensorimotor representations. Drawings make students accountable for their ideas (exposing them to peer and teacher criticism and analysis). When a student could not convey his idea to anyone, he drew it on the board and everyone understood.

The author also investigated the cognitive role of design artifacts, finding that they allowed students another route to externalize ideas and make them concrete and accessible. Moreover, each time an idea was concretized and embodied, students made an epistemic "move", thus decreasing their mental demands. Not all of these artifacts were working models of a design, students sometimes used stand-in materials for future materials they intended to use and recognized that they were representations of the final thing. The technological design contexts also provide students with many communicative opportunities even if they had a more limited linguistic repertoire. This enabled students to do higher levels of analysis than they would be able to engage in with only verbal communication. Students found it difficult to articulate their ideas about the unit, but could show their ideas in the presence of their artifact. The unit provided opportunities to infuse science and design discourse because the students were engaged using their primary discourse talking about things in a way that was already similar to science discourse.

These findings' credibility are generally supported by a fairly detailed description of analysis methods, however the author's assertion that more can be found in other articles is not very helpful. The author did provide a detailed description of the study's subjects, also improving credibility and reader understandings of transferability. The analysis was based on well-explicated philosophical and cognitive theory. The author used many sources of data and several researchers for triangulation, also improving credibility, and referenced other relevant research for triangulation of concepts.

On the other hand, the author did not provide any information about the location or context of the study, and, except for counting three ESL students, did not describe the subjects' races, ethnicities, or cultures, decreasing potential transferability.

In a quasi-experimental pretest posttest comparison group study of eighth grade students in urban US schools in design-based and scripted-inquiry classes, Doppelt & Schunn (2008) found that homework and instructional worksheets were the most influential activities for both groups, and other factors were rated similarly between the groups according to factor analysis of student survey data. They found that homework and instructional worksheets were the most influential activities for both groups, and other factors were rated similarly between the groups.

The authors developed a questionnaire called the Design-Based Learning Environment Questionnaire (DBLEQ) as an instrument for this study. The questionnaire included 14 classroom features and 14 learning aspects, and asked students to rate the importance of each feature from one to five on each aspect of learning. They divided the survey into four versions each with seven features and learning aspects each, to access all 196 items. The authors used the same study for the scripted inquiry classes because both have been found to have similar classroom features, although these features may not be equally important in the two types of classes. The questionnaire was given at the beginning and at the end of the school year, and examined for stability and Cronbach's alpha. Students were also required to justify two of their response, with three assessors rating these justifications for reliability.

The subjects included 587 eighth graders from nine schools in design-based learning classes using the Electrical Alarm System module and 466 from six schools using a successful scripted inquiry curriculum. Of these, 464 and 248, respectively completed both the pretest and posttest. All the schools were mid-sized, urban public middle schools in the US. The authors did not provide additional demographic data.

The highest mean scores (most influential) reported for both experimental and control groups were for Homework (mean experimental = m_e = 2.79, mean control = m_c = 3.43) and Instructional Worksheets (m_e = 2.64, m_c = 3.38), while the experimental group also highly rated computer use (m_e = 2.50) even though this class rarely used computers in practice, and the control group highly rated independent learning activities (m_c = 3.37). The authors did not present statistical significance levels for these data, although used $p < 0.01$ for other parts of their paper, and presumably these means were higher than the others at the same significant level. They also wrote that they used factor analysis to reduce the number of learning environment characteristics and outcomes to seven groups each to find the most influential characteristics from pretest to posttest. They found that homework and instructional worksheets were the most influential characteristics for the experimental group, and most impacted outcomes were interest in science topics, teamwork skills, and independent learning activities. The authors found the Cronbach’s alpha of the DBLEQ to be 0.78, and its stability was $r = 0.4$ ($p < 0.01$). The mean quality of justifications was 3.54 ($\sigma = 1.33$) indicating good reliability and understanding of the questionnaire task. Interrater reliability of these justifications was 0.90, 0.81, and 0.78 between each possible pair of three assessors, indicating that these ratings were reliable.

The researchers of this study used a straightforward and consistent instrument and provided detailed reliability statistics. They surveyed a very large number of students for design-based curricula studies, increasing the reliability and validity of the study. The justifications provided further indication of the reliability of the instrument in their study. However, the authors do not provide detailed demographics that could impact the outcome, which reduces generalizability. It is also unclear if these students had been using inquiry or design-based learning in their schools prior to their pretest. If they had not, it doesn’t make sense to rate their opinions of the features of these classrooms before they started working on them.

In a quantitative interview study, Schauble et al. (1991) investigated how fifth and sixth graders change their model of inquiry during science and engineering oriented experimentation tasks and found that students developed scientific thinking more effectively when they started with engineering thinking tasks.

For this study, each subject worked on a set of experimentation tasks designed to mirror an engineering model of inquiry, in which a desirable outcome appears to exist, as well as on a set of tasks designed to mirror a science model of experimentation, in which no desirable outcome is apparent, and variables are more independent. Both were related to objects moving through fluids. The engineering task involved determining what could make a boat move faster in a canal with a constant amount of force, and the science task involved determining what variables affected how much a spring stretched that suspended an object in water. Half of the students worked on the engineering task first, then on the science task, and the other half worked in the other order.

Each task had a set number of experimental variables to change and a set number of options for each variable, so there existed a fixed number of possible experiments that the subjects could try. Subjects spent three 40-minute sessions working on each of the two tasks, with a total time of two hours on each task, completed within one week. Students described levels to investigate for each variable on a data card to keep track of experiments they ran.

Investigators used these completed data cards as well as audio recordings of predictions, judgments of predictions, judgments about the effects of variables, and justifications for inferences and conclusions. The authors primarily analyzed the inquiry process by determining the percentage of inferences that students made that were valid according to researchers' criteria: the inference had to be based on correct content and based on sufficient and unambiguous experimental data. The authors used different critical p values for different results.

The subjects of this study were nine girls and seven boys from four fifth and sixth grade classrooms (four students from each) in the same middle school in a middle-class suburban neighborhood near a mid-sized city. Subjects were chosen by their teachers to assure that each class' representatives were diverse in terms of race, culture, gender, and science ability and achievement. Eight subjects were white, six were black, two were Hispanic. Ages ranged from nine years one month to 12 years 11 months, with a mean age of 11 years, two months.

The authors found that the percentage of valid inferences rose steadily over the first three sessions (first task). The mean percentage rose from 37.6% on first task to 54.2% ($t = 2.94$, $p < 0.005$), which was statistically significant. While mean valid inferences increased from the first task regardless of which task was first, with $F[2,14] = 5.93$, $p < 0.007$, students who began with the engineering task showed a jump in correct inferences in the science task and continued improvement in that task, while students beginning in the science task decreased correct inferences during the course of the engineering task. This result was consistent with the hypothesis that because students are more comfortable manipulating variables to create a desired effect, they will learn best if they begin with activities that allow them to do this and only then engage in activities that require more strategic variable manipulation as in canonical scientific inquiry.

These authors clearly described their coding techniques, improving credibility. They also thoroughly described the diverse student demographics, improving transferability. On the other hand, a lack of teacher demographics or experience levels decreases potential transferability. Small sample sizes weaken the study’s credibility, and a lack of statistics for the author’s conclusion of a larger increase on the second task for students beginning with engineering decreases confirmability of the results. Similarly, the authors did not determine a critical p-value before analyzing data.

The studies in this section by Roth (1996; 2001) provide some of the most detailed analysis of the culture and environment of design-based courses. His key result is that transformation of objects and actors in design-based classrooms provide the key opportunity for students to access, build and share their understanding. If it is true that ontological flexibility and transformation is key to the efficacy of these environments, Doppelt and Schunn’s (2008) investigation into the relative importance and utility of the different ontological elements of design-based classrooms becomes rather unhelpful. However, their result that the most important elements of the classroom for students include worksheets and homework, elements just as present and important in scripted inquiry courses, does not motivate the extra effort to design and implement design-based learning. Schauble et al. provided evidence that design-based inquiry may scaffold students towards scientific inquiry particularly effectively.

Evidence for the Efficacy of Particular Techniques for Design-Based Science

The above section provided some evidence for the features of design-based classrooms that could enable them to be particularly effective environments for teaching science. However, some of the previous studies showed that design-based science does not always scaffold all important aspects of science learning effectively. Certainly, technical education that only incidentally overlaps science content does not effectively teach science content. From the above studies, it should be clear that students will not learn science from hands on design activities unless the curriculum is carefully designed to scaffold students towards doing so. Some of the previous studies in particular showed that verbal scaffolding is not enough to get students to organize their thoughts into a more scientific understanding. The four studies in this section investigate what kinds of scaffolding and implementations of design-based science could be used to improve these curricula.

Sadler, Coyle, and Schwartz (2000) investigated what effective engineering competitions look like at the middle school level and found that design competitions were more effective when students were competing against nature instead of against each other, when it was possible to improve results by several orders of magnitude, and when repeated, small design iterations were developed. Penner, Lehrer and Schauble (1998) investigated how designing functional models developed a suburban class of third-grade children’s understanding of science concepts and found that students were able to emulate many scientific thinking skills in designing and presenting their investigation when the teacher provided significant scaffolding support including class evaluations between design iterations and utilizing modeling as the design goal. Puntambekar and Kolodner (2005) investigated appropriate scaffolds for design-based and problem-based curricula in a mid-SES suburban Atlanta middle school, finding no improvement in scientific thinking during the first enactment of the curriculum, but significant improvement in scientific thinking during a second enactment of the curriculum that provided more social and cognitive scaffolds. Barak and Doppelt (2000) investigated how students interacted with a portfolio in design-oriented science inquiry and how this interaction could support their higher order cognitive skills, finding that although students strongly resisted documenting their work, the portfolios were an effective scaffolding context for teaching higher order cognition and metacognition in science design. Hmelo, Holton, and Kolodner (2000) found that a teacher implemented effective learning scaffolds using verbal questions and white-board recording, but did not provide scaffolding to help students use their last results to make future design plans, and students did not iterate their designs effectively after making a design that worked.

In a mixed methods study, Sadler et al. (2000) investigated what effective engineering competitions look like at the middle school level and found that mean scores increased significantly at the $p < 0.05$ significance level. From an analysis of their qualitative data, they found that design competitions were more effective when students were competing against nature instead of against each other, when it was possible to improve results by several orders of magnitude, and when repeated, small design iterations were developed.

For the quantitative part of the study, the authors administered open-ended pretests and posttests to measure changes of students' science process knowledge. The authors described their scoring rubric, which quantifies correctly testable predictions and manipulated and controlled variables, and the quality of an experimental design required in this test. The authors administered this test twice during the school year to a treatment group of N = 457, in 22 classes (12 teachers) and ran t tests on this data. They also gave a subject matter test to measure common preconceptions versus more scientific views, but gave these tests in only one of the units, the electromagnet module, and gave it to each of the 17 students who worked in groups of two and three.

The authors obtained qualitative data via student interviews centered on tracking students' explanations for the results of a classroom challenge. They also conducted classroom observations, although did not provide specifics, and analyzed student storyboards of student progress on class projects.

In addition, the authors compared 10 dyads of sixth grade students in their ability to improve designs by only changing one variable at a time. One group was required to build and test a prototype to begin their design process, while the other group was allowed to start any way they wished. The dyads were split into two groups of 10 students each and compared based on the percentage of successful single variable changes that they made.

The authors stated that the 22 middle school classes were from a variety of locations in the United States, but did not give any further demographic information.

In analyzing quantitative data, the students' mean pretest scores increased on all items, and increased significantly ($p < 0.05$) on 8/11 items and the mean score, which was initially 0.437, rose to 0.553, $z = 0.363$. For the electromagnet subject test, the authors recorded gains for 11/12 items, and significant gains ($p < =$

0.05) for three. Mean scores rose from 0.556 to 0.689, with $p < 0.05$, effect size $z = 0.833$.

Interviews tracked engagement levels of students in this course. The authors fFound that sixth graders all engaged at the sensori-motor level. The authors reported interview quotes showing development of the representational level of engagement such as improved or more scientific models of their electromagnet that were indicated by the students' drawings. The authors found that many of the male students and an even larger fraction of the female students did not feel competent to compete if doing so required them to build from scratch.

The authors reported several general conclusions, ostensibly based on some broad impression or analysis of their data. They found that design competitions in middle school were more effective when students were "competing against nature" or against prior iterations versus competing against other teams. They found that it was important that tests of students designs could show objective improvement in efficacy, such as by using maximum wind turbine power or battery output. They also found that it was important for designs to potentially include large dynamic ranges of results (10-100X the original performance).

For the dyad test, students starting by building and testing a prototype (using a cookbook method) were able to change only one variable at time 80% of the time, while the other group succeeded 50% of the time with $p < 0.05$.

In analyzing student records, the authors found that few students saw the purpose in making detailed laboratory notebooks, except in the few cases where these records were used repeatedly during reflective activities. The authors successfully implemented and recommend storyboards instead of lab reports to document student progress towards improvement. Devices were either attached to each storyboard box or drawn in. Teachers found the storyboards useful partly because other students doubted when their peers changed more than one variable at a time.

Finally, the authors found that male students who started were no longer ahead after many iterations once everyone had learned the requisite skills.

In conclusion, the internal validity of this study's experimental results benefitted from a relatively very large sample size. The study's reliability also benefitted from extensive discussion of theoretical basis of project, and triangulation with others doing similar work, and as similar findings from other researchers. On the other hand, this study lacked a comparison group so the authors could not successfully argue if the curriculum was comparably effective. The authors also failed to provide a Cronbach alpha, weakening the validity of the quantitative results. The authors' failure to describe their procedure for gathering or analyzing qualitative data weakened their qualitative conclusions and made it difficult to decide if their recommendations were simply based on the fact that the classrooms were effective and followed those recommendations, or if they were based on a deeper analysis of the data. The strongest findings of this study are that students tend to perform better when they work in many iterations, and that students perform better when they begin with a prototype.

In an ethnographic case study, Penner et al. (1998) investigated how designing functional models developed a suburban class of third-grade children's understanding of science concepts and although they didn’t describe data gathering techniques or analysis methods, and found that students were able to emulate many scientific thinking skills in designing and presenting their investigation when the teacher provided significant scaffolding support.

Participants in this study included 17 third-graders in one suburban district near a small Midwestern city. Children worked in dyads or triads. The first author created the curriculum along with the classroom teacher, which was centered on an iterative design goal of making something that worked like an elbow. The teacher taught this curriculum over five 45-90 minute classes. The authors did not describe any method for taking or analyzing their data.

The authors found that in the iterative design process, students naturally evaluated their models by comparing their models to other students’ models, and to real elbows. The authors also found that the design challenges in this curriculum were effective in fostering collaboration and debate about scientific models.

The authors found that the teacher was effective in guiding students through the following issues, and that these problems successfully contributed to student learning: students had to decide what to investigate, and how to organize and present their data visually in order to make their arguments. The authors concluded that the teacher had provided significant scaffolding for these science processes by facilitating group data manipulation decisions in a computer spreadsheet program. However, the author found that students were not successful analyzing or describing the scientific principles underlying the data patterns that they easily observed.

In presenting this study, the authors provided a detailed narrative account of the classroom and their conclusions about it. They included a wide range of exemplary student quotes, which improved the study's confirmability. However, the authors did not provide any description of how they collected or analyzed their data to create their narrative or analyses. This lack of method information decreases the confirmability of the study. Nevertheless, the detailed narrative and conclusions that were reasonably consistent with other studies made credible the conclusions that iterative design activities can motivate students to engage in scientific thinking, and that young students in design curricula may be able to effectively find patterns in their data but not be able to analyze those patterns using scientific principles.

In an observer-participant case study with two iterations, Puntambekar and Kolodner (2005) investigated appropriate scaffolds for design-based learning curricula (although they called it a project-based curriculum) in a middle socioeconomic status suburban Atlanta middle school by coding and analyzing responses in student journals and recording and analyzing observations. They found that students exhibited no improvement in scientific thinking during the first enactment, but significant improvement in scientific thinking ($p < 0.001$) during the second iteration that provided more social and cognitive scaffolds.

The initial scaffolds that the researchers provided were based on "Design Diaries", in which each page had prompts intended to assist students with the related design step. It also included ideal and non-ideal examples for responses. For the first enactment, the authors analyzed student diaries using a four-point coding scheme. For each, a score of 0 was given for no scientific thought, one for some science that was general, two for some scientific explanation without significant depth, three for a well-justified scientific explanation. Seven coding categories were analyzed representing progressive stages in the design process: student understanding of the challenge (U), raising questions or learning issues to clarify the problem (Q), initial set of ideas (I1), learning issues to clarify the topic (LI), second set of ideas before learning about erosion (I2) after research, third set of ideas that were proposed solutions (S), and criteria given for preferring solution (C). The authors used a Wilcoxon signed rank tests to analyze progression of student's individual scores. To elucidate further, the authors recoded whether students concentrated their responses on structure, function, or behavior of their designs.

The authors described classroom observations to elucidate their numeric findings. The first author collected classroom observations from these classes. Every class was observed at least three times per week. Observations focused on noting student uses of design diaries and observing group discussions. The study did not describe how observation notes were analyzed.

In a second iteration, the authors redesigned their scaffolding framework to include more visible specifics of each phase, make more visible the connections between the phases, and provide scaffolding for thinking about function. The diaries now provided prompts at several cognitive levels and included pages for students to write a specification for their design. The new iteration also included social scaffolds such as pin-up sessions in which students would show off their ideas, models and designs; and whole-class discussions inserted at appropriate points in the sequence. The authors used the same coding scheme as in the first study.

The authors also compared student responses across the two studies – focusing on design issues in the phase (I1), (I2), and (S) phases using the Mann-Whitney U test, a non-parametric version of an independent samples t-test.

These studies were carried out in a mid-SES suburban Atlanta middle school with eighth grade students. Each of the four Earth Science classes included 30 students with a range of abilities. The teacher was a seasoned project-based learning facilitator, but this was his first time doing the Jekyll Island problem. The second study was carried out a year later by the same teacher in the same school with another set of eighth grade Earth Science classes.

In the first study, the data showed that 75% of students simply repeated the design problem verbatim, so this 75% scored a zero or one on initial design ideas (I1). Additionally, 65% scored zero or one on second set of questions (L1), 70% scored 0 or one in second set of design ideas (I2). The Wilcoxon signed rank revealed an insignificant improvement in number of students with a better understanding of scientific ideas from the I1 to I2 design phases and a significant decrease in number of scientific answers to solutions from I2 to S ($z = -2.085$, $p = 0.037$). Further analysis revealed that students concentrated heavily on structure of their solutions, neglecting the more central qualities of function and behavior. The data revealed only small gains in function responses recorded near the end.

An analysis of observation data indicated that students interacted well together and did their journals as homework. Most students refined their design by trial and error without systematically reasoning about the changes they might make to their designs and what might result. Students did not usually refer to their criteria they had set out. They were not encouraged to go back and annotate their diaries after group work, and questions and ideas explored independently in diaries were not explored in group or class discussions. The Teacher didn't check the diaries daily, so wasn't aware of the lack of depth therein. Students had a particularly hard time using available resources to gather specific information needed to move forward, and were unsystematic.

In the second study, students began with a much more accurate identification of the problem and showed improvement in their justifications of designs. On this study, 22% of students scored a two or a three on the U phase and 20% on Q. They also more scientifically elaborated solution ideas. Student explanations were coded as 20% 2s and 3s on I1, 50% on I2, 50% on S. The difference between I1 and I2 was now significant ($z = -0.308$, $p < 0.001$). The students also showed improvements also for coding of structure, function and behavior. Now, the number of function-related responses out of 100 students from increased from 31 during U, to 34 during I1, to 38 during I2, to 42 during S. Observational data indicated that student spent more time trying different ideas before choosing a solution and asked more function-related questions than in the first study.

In comparing the first study (group one) to the second study (group two) comparison, the authors found insignificant differences between groups during the I1 phase. However, group two significantly outperformed group one on I2 a week later with twice as many twos and threes ($z = -3.86$, $p <.001$). For the S phase, 21% of responses were coded as twos and threes for group one versus 50% for group two ($z = -4.84$, $p < 0.001$). For the C phase, 49% were coded as twos and threes for group one, and 67% in group two.

Overall, the results indicated that design-based science requires scaffolding at several different levels and domains, which can be largely implemented through well-designed portfolios. The authors' clear explanation and justification of their coding methods improved confirmability. The qualitative observations and analysis of the journals provided triangulation, improving confirmability. The basic conclusion that scaffolding needs to target many domains and levels and be well structured towards functional thinking is essentially believable. On the other hand, the confirmability of these qualitative results was weakened by a failure to describe how classroom observation notes were analyzed. In addition, it is possible that the teacher's natural improvement in implementing this curriculum after one year affected the results. The lack of discussion of the demographics of the subjects reduces the external validity and transferability of this study's quantitative and qualitative results.

In an ethnographic case study, Barak and Doppelt (2000) investigated how students interacted with a portfolio in design-based science inquiry and how this could support their higher order cognitive skills, finding that although students strongly resisted documenting their work, the portfolios were an effective scaffolding context for teaching higher order cognition and metacognition in science design.

The authors examined portfolios that were used in a five-year program in which students designed and programmed robots using LEGO-Logo. Students produced a total of 35 portfolios. Researchers documented student difficulties during problem solving, designing and programming, and documented teacher difficulties. They provided little explanation for how they accumulated or analyzed their data. Presumably some information is available in other studies by these authors about the same 5-year program. The authors did describe how they developed a four level system for analyzing the cognitive levels evidenced in drawings and descriptions of designs and processes, but described four cognitive levels of responses that they used to code portfolios.

The subjects included 56 10th grade Israeli students, all described as having low academic achievement and most described as at risk for not being able to graduate due to academic standing. The author did not provide any additional information about the subjects.

The authors found that every year, students were highly resistant to documenting all stages of the design and construction process, and explained that perhaps they were used to being graded primarily on final turned-in work rather than on their process, so did not see the value in work towards process goals. Students assessed their own and peers work within the journals, and showed a high consistency between peer and teacher scores (reliability alpha = 0.987), indicating that these portfolios were a good context for self and peer evaluation. The authors found that the portfolio provided opportunities for many levels of thinking, from copying standard schematics to describing how mental models changed as a result of tests. Most portfolio elements at the end of 10th grade were at the second and third cognitive levels, although students created more third and fourth level elements when they continued their projects into 11th and 12th grades.

Overall, the researchers concluded that portfolios provided a good opportunity for students to document the development of their thinking processes as well as their design and construction process. Therefore, creative portfolios could be a useful scaffolding context for teachers to create lessons on higher order metacognition in science design.

Strengths of this study were that it was conducted over several years, and triangulated its findings with similar results each year, improving credibility. On the other hand, a weakness of this study was very little description of method, so it was difficult to establish which of its findings were transferrable. I'd like to know how the teachers taught students to use their portfolios so that they could be effective. The authors established that portfolios are useful tools for scaffolding, but not how they can be used for scaffolding within a day-to-day learning environment. Finally, the Israeli context of this study may limit transferability to settings in the United States.

In a mixed and pretest posttest quasi-experimental, and descriptive ethnographic comparison of traditional and design-based diverse sixth grade Atlanta classrooms, Hmelo, Holton, and Kolodner (2000) found that the teacher implemented effective learning scaffolds using verbal questions and white-board recording, but did not provide scaffolding to help students use their last results to make future design plans, and students did not iterate their designs effectively after making a design that worked.

The authors randomly selected 20 students in the experimental group to draw a model of the respiratory system. One independent rater coded all transcripts and diagrams as corresponding to five levels of models (5 indicating most sophisticated understanding) while a second rater coded six models to test for interrater reliability, which was determined to be better than 95% for diagrams and greater than 90% for the interviews. The researchers also recorded LBD classroom activities via daily observations, videos, and field notes. They analyzed these to create a rich ethnographic description, included in the article, from which to base findings.

The subjects of the study were 42 sixth grade students in two design-based learning classes in Atlanta taught by a teacher who had participated in a summer workshop with the researchers. The teacher had only briefly experimented with problem based learning previously and did not have experience with design-based learning. The students were described as being of mixed genders, abilities and socioeconomic groups, but the authors did not provide more demographic information.

This study found that students initially had level one or two models in both groups (68% for comparison and 61% for experimental group). At posttest, both group had mostly models of levels two and three, but design-based learning students reliably improved with $p < 0.005$, while the comparison group did not with $p > 0.25$. Several students in the design-based learning group achieved level five models while only one on the comparison group did. Design-based learning students described more relations in the posttest, with $F(1,17) = 6.16$, $p < 0.05$.

The researchers also analyzed their rich picture description of the classroom enactment for how that enactment varied from how they intended it to work. They found that the teacher effectively used computer whiteboards designed for PBL to scaffold students in generating questions and planning and investigating answers. However, scaffolding to help students plan, design, debug, or explain their models was missing and would have been useful. They found that the teacher effectively used these whiteboards as external memory to orient students to what students had learned the day before. The teacher did not ensure that students could create and analyze multiple iterations of their models, which was problematic because the students could not physically try out their ideas about how the lungs really worked as they improved their understanding. Without iteration, students were not able to take into account the mechanisms underlying how lungs work, and how the different features were connected to each other. The authors also concluded that the design challenge was too complex to allow students to apply what they learned about the content, because they only had time to make one iteration.

This study provided very detailed ethnographic account and analysis, and tied the results and conclusions explicitly to prior understanding of the research. However, the missing comparison group interviews means that the study sheds little light on how student thinking in a design-based class is different from in other classes. The authors' statement that the classroom was heterogeneous provides little information about the demographics that could be used to determine transferability of the study's results.

In conclusion, the most important scaffolding technique in effective design-based science curricula is iteration. The key then is to use this iteration as effectively as possible, which seems to occur when students use scientific techniques to evaluate their previous iterations, and construct scientific understandings of those iterations that they then apply to improve their next iteration. Additionally, teachers need to get different groups involved in analyzing and evaluating each other's work so that they can learn from each other and teach each other what they have learned. Throughout this process, teachers should also provide students with a scaffolded system to record their progress in constructing, designing, thinking, and learning, although this record system need not include the artificial linearity, rigor, and objectivity of traditional lab reports.

Summary

The research reviewed in this chapter provided insight into the challenges in implementing project-based science for students and teachers that are likely to be present in design-based science, and showed that technology and engineering education classes not systematically designed to teach science are unlikely to do so. Studies from this section also showed that design-based learning is likely to be effective for teaching science in a variety of settings, and can in fact be more effective than previously well-established and effective criteria when it carefully implements appropriate scaffolding. It can do so because of the opportunities for transformation and exchange of ideas that it provides, and because students may more naturally pursue investigations as design improvement than as problem solving to answer essential questions. Research also shows that design-based science curricula are most effective when they scaffold student development and sharing of ideas through evaluating and improving iterations of design.

CHAPTER THREE: CONCLUSION

Introduction

Chapter One reviewed the rationale, historical background, purpose and limitations of this paper. That chapter described design-based learning as a learning model that emphasizes the creation of technological artifacts as a way to learn science and math skills. This learning model originated with the design challenges in early architecture and engineering schools, and also has its roots in the constructivist learning theory of the 20th century that shows that students construct new knowledge when they need it. Like project-based learning and problem-based learning, design-based learning would ideally provide both the need and the opportunity for students to develop a scientific way of thinking and working. Design-based learning is also related to technology and pre-engineering education, and has been similarly lauded as a solution to the United States' supposed loss of technological competitiveness. However, findings described in Chapter Two indicated that a key difference between design-based science learning and technology education is that technology and pre-engineering education may not systematically or successfully teach science content and skills, while design-based science can do so. Chapter One also indicated that the purpose of this paper is to determine if, when, and how design-based learning can be an effective way to teach science.

Chapter Two detailed 30 peer-reviewed studies that answered the question of if, when, and how design-based learning can be an effective way to teach science. This chapter began by describing relevant studies of project-based science that indicated various difficulties students and teachers may be expected to face with these kinds of learning models. It then described studies of technology education classes indicating that the inclusion of these classes is unlikely to improve student learning in science. The next section reviewed studies that provided evidence for student learning in design-based science classrooms, finding it comparable with other curriculum types. The only design-based science implementations in which students did not learn were those that provided weak scaffolding or classroom management. In addition, studies described in that section showed design-based learning to be particularly effective at teaching skills at higher-cognitive levels, and provided an opportunity for students to express their learning when they could not on a traditional test. Studies in the next section showed that design-based learning classrooms were able to help students learn by offering them many different ways to represent and express ideas, and by allowing students to transform elements of their environment as part of the learning process. The studies described in the final section of Chapter Two provided evidence of good practice in design-based science classrooms, carefully scaffolded science learning while students improve their designs over several iterations.

Chapter Three summarizes the research findings of each of the sections and sub-sections of Chapter Two, then describes implications for classroom teaching that can be gleaned from these findings. Finally, this chapter discusses limitations of the extant research and suggest improvements and new ideas for future research of design-based science, and provides concluding remarks.

Summary of Findings

Insights From Research in Related Learning Environments

This first section of Chapter 2 reviewed research on learning environments related to design-based science, including project-based learning and various forms of technical education. Overall, this section concluded that several challenges in implementing project-based science are likely to be encountered as well in design-based science, such as the difficulty in getting students to analyze and present their scientific observations, and the difficulty for teachers to let go of control and fost transitions to student-directed investigations. Furthermore, technical education that does not systematically teach science is unlikely to incidentally do so, although it may motivate students to be more academically successful in general.

Investigations of Project-Based Science. The studies in this section particularly investigated challenges encountered in enacting project-based science that are likely to also be encountered in design-based science. Krajcik et al. (1998) found that students engaged in science-like observation but not analysis or presentation, a tendency that is found in some research in design-based learning (Resnick et al., 2000). Marx et al. (1994) found that teachers often struggle or fail to give up control to student groups as students do their own investigations and summarize their conclusions to peers. Panasan and Nuangchalerm (2010) attempted to compare inquiry-based and project-based science, and although they failed to clearly define these overlapping pedagogies, they hinted at the often similar results of different modern constructivist curricula.

Studies of Technology Education Environments. The studies in this section investigated whether or not different technology education curricula that do not set out to systematically teach science content nonetheless do teach science content. If these studies had found that students learned science content just by taking technology education courses that involve science, the work of developing design-based science would be somewhat unnecessary. This was not the result. Stone and Aliaga (2005) found that students tracked in technical education or school to work programs do better than their counterparts tracked in general education, there is no evidence that this success is due to anything but a generally improved motivation to engage in school. Tran and Nathan (2010) in fact found that although engineering education might be expected to improve science and mathematics learning more than other forms of technical education, students taking popular engineering education courses did worse on a state mathematics test and only as well as their peers on a science test. Childress (2011) found that students in technology education classes that were timed to correlate with the content in their science classes did not do better on a science test. Finally, Crismond (2011) found that unlike professional designers, students do not tend to use science knowledge in their designing, indicating that teachers need to provide structures for students to do so. Overall, the findings of these studies indicate that including or correlating technology design and education in the curriculum is not enough to improve science learning as it is traditionally understood.

Evidence For and Against Learning in Design-Based Learning Environments

Studies showing relatively improved learning. The three studies in this section provided the strongest evidence that design-based science is an effective way to teach science because these studies not only show significant learning, but significantly better learning than in other types of curricula. The study by Mehalik et al. (2008) in particular provided strong evidence for design-based learning because this study included a very large sample size of 587 diverse students in a variety of classroom types, and found that design-based science compared favorably to a well-established research-based and validated inquiry-based curriculum. Silk et al. (2009) also used a large sample size for the experimental group (although not a large comparison group) and found improved science process skills for students in a high poverty school. Mooney and Laubach (2002) found better mathematics learning in a design-based course. Overall, the studies show that design-based education can produce superior results in a variety of settings for student learning in science content, science skills, and mathematics.

Non-comparative studies showing content learning. Unfortunately, only the three studies in the previous section directly compared student learning in design-based classes with student learning in other settings. As a result, it is necessary to review the more abundant studies that simply provided evidence that students learn or don't learn in design-based environments. The five studies in this section provided quantitative evidence that students learned in these environments (some in the next sections provided evidence that students did not not learn). In a large study, Apedoe et al. (2008) found that students learned chemistry successfully in a design-based chemistry curriculum in a variety of settings. In another large study, Fortus et al. (2005) found that students improved their science content knowledge and successfully transferred that knowledge to a new design task. However, these authors' implications are dubious because a transfer of knowledge to a task similar to the ones students were used to not is a particularly useful way of measuring transferability of knowledge. In another large study, Doppelt et al. (2008) found that traditionally "high achieving" but not "low achieving" students improved their scores on a traditional test, but that the "low achieving" students actually showed better scientific thinking as seen in their documentation of artifacts. On the other hand, Fortus et al. (2004) found that both categories of students improved significantly on a traditional test. Finally, Sullivan (2008) found that traditionally high-achieving students improved their systems design skills during an after school robotics program. Overall, a wide variety of students can certainly learn traditional science content and skills in design-based classes, but for some, traditional assessment is insufficient to identify this learning.

Evidence of non-traditional science learning. While design-based science may be an effective model for teaching traditional science content, the studies in this section identified non-traditional areas of science for which it is particularly effective. The strongest and most general conclusion in this section was provided by Kanter (2010), who found in a very large study that while students showed overall improvement on a science test after design-based curricula, they improved most at the highest levels of cognitive demand. Penner et al. (1997) investigated the effects of a design-based curriculum on elementary school students' modeling thinking, finding excellent improvement. Resnick (1996) found that a simulation design curriculum was effective for improving students understanding of non-centralized systems that are increasingly important to science.

Evidence of failure of design curriculum in teaching science. The previous three subsections generally supported the conclusion that design-based science can be effective in helping students learn science. The three studies in this section, however, provided evidence that design-based classrooms were not effective. Barnett (2005) and Seiler et al. (2001) both investigated design-based science in high needs inner-city schools where they did not find it to be effective as measured by those authors' primary test instruments. However, Seiler provided some qualitative evidence that students knew the content tested, and Barnett's team did not appear prepared for the classroom management challenges of their rather extreme setting. Resnick et al. (2000) also found that a more exploratory and less structured design-based class did not promote student learning of science process. In fact, in all three of these studies, the curricula investigated did not include strong scaffolding.

Important Features of Design-Based Science Classrooms

The above sections showed that design-based science curricula can support student learning in science, especially when effective classroom management and scaffolding are present, although sometimes learning cannot be tested effectively using traditional testing measures for traditionally underserved students. The four studies in this section attempted to identify the important features of design-based classrooms and their roles in facilitating such learning. The first study by Roth

(1996) investigated the roles of different classroom elements in particular detail, finding that the great affordance of design-based science lay in the fact that students constantly rearrange, rebuild, and reimagine those elements as they pursue their design projects. In his second study, Roth (2001) described how these elements contributed to the classroom environment as a unified cognitive system, finding that students were able to express and exchange ideas using the artifacts and other representations before they could express those ideas using organized language. On the other hand, Doppelt & Schunn (2008) attempted to identify the most important and influential elements in design-based classes, and found that they were largely the same as in more traditional inquiry-based classes. Finally, Schauble et al. (1991) found that young students were more successful in conducting science-like experimentation after they had conducted engineering-like experimentation, providing support for the theoretical use of scaffolding students in systematic science processes through introducing them first to design processes.

Evidence for the Efficacy of Particular Techniques for Design-Based Science

After determining that design-based curricula can be an effective way to teach science, the most important goal of this paper is to determine how design-based science can be most effective. The four studies in this section directly investigated which teaching and scaffolding techniques are most important for making design-base science effective. All of these studies emphasized the importance of design iterations throughout which students learn science content. Sadler et al. (2000) emphasized the usefulness of choosing design challenges that allow students' iterations to improve greatly. Puntambekar and Kolodner (2005) found that including detailed structures for students to reflect regularly on their design, thinking, and learning processes, as well as social structures for student exchange of ideas, significantly improved a design-based science curriculum. Barak and Doppelt (2000) found that design portfolios were especially useful for scaffolding higher order thinking including metacognition. Hmelo et al. (2000) found that verbal scaffolds and white-board recording were insufficient for helping students learn from one iteration to the next. It is noteworthy that no study provides evidence that students should follow a strict process or procedure for designing, nor is it helpful for them to attempt to describe their process as linear in records such as in traditional science lab reports. Instead, students need to have opportunities to record, reflect, on and exchange the progress of their science ideas as they design so they can build upon their progress in future thinking, designing, and physical constructing.

Classroom Implications

The above studies provide evidence that design-based science can be a powerful environment for teaching science content and skills in most school contexts. While these curricula can more effectively improve student understanding of basic science concepts than can guided inquiry (Mehalik et al., 2008) or traditional curricula (Mooney & Laubach, 2002; Silk et al., 2009), they are especially effective at improving student learning at higher cognitive levels (Kanter, 2010), in improving non-traditional science understanding such as modeling (Penner et al., 1997; Penner et al., 1998) and systems thinking (Resnick, 1996; Sullivan, 2008). They can also provide improved contexts for assessment and evaluation when traditional contexts are insufficient, especially for traditionally underserved students (Doppelt et al., 2008). On the other hand, when inappropriately or insufficiently scaffolded, design-based science is not effective for teaching science content and process (Barnett, 2005; Resnick et al., 2000; Seiler et al., 2001), nor is technology-oriented curricula that does not systematically teach science content (Childress, 1996; Crismond, 2001; Tran & Nathan, 2010).

The question, then, for the teacher interested in implementing design-based science, is how to do so effectively, and what scaffolding to use to enable students to learn the science content through their science investigations. While design activities have a long history in science education, they were traditionally done as a culminating project “applying” what students had learned previously (Silk et al., 2009). The findings of the research reviewed in this paper shows that this sort of design project is unlikely to be effective. Instead, the research emphasizes the need for design-based science to teach the science between and within several iterations of design projects, not outside of these projects (Barak & Doppelt, 2000; Hmelo et al., 2000; Penner et al., 1998; Puntambekar & Kolodner, 2005; Sadler et al., 2000).

The improvement of design iterations can motivate students to systematically investigate relevant science principals by first trying experimenting to get better results in the vein of engineering research, an experimental process that may be more natural for children (Schauble et al., 1991). Roth (1996; 2001) further motivated why students learn while they attempt to improve their creations. He wrote that the fact that students are constantly reinvisioning their artifacts, sometimes as symbols for ideas, sometimes as tools, and the broader fact that they can constantly reconstruct the elements of their environments, means that they can use their entire classroom environment more effectively as a cognitive system to scaffold their construction of ideas and processes before they can fully comprehend or enact these ideas and processes on their own. Furthermore, when students are systematically provided opportunities to record and share their ideas, such as in presentations and portfolios, students can build on each others' knowledge and teachers can help students synthesize their thoughts into more coherent and scientific ideas (Hmelo et al., 2000; Putambekar and Kolodner, 2005).

Similarly, design-based learning can potentially provide opportunities for assessing students not well-served by traditional tests. Roth (2001) showed that students in design-based learning classrooms were able to express their thinking using words, pictures, gestures, and their physical artifacts. He writes that students were in fact able to explain ideas using gestures and drawings before they could clearly explicate them verbally, so teachers should consider using gestures and student drawings as data for formative assessment. Doppelt et al. (2008) wrote that traditionally "low-achieving" students, who did not show improvement on paper and pencil tests, actual showed more exemplary thinking through their artifacts than did their "high achieving" peers. Seiler et al. (2001) wrote that students who were not able to explain scientific thinking in the usual class setting were able to do so in a one-on-one interview setting if they were allowed to hold and gesture to their design artifacts. Overall, teachers should employ a variety of assessment techniques for design-based learning, including traditional pen and paper tests, a portfolio assessment that includes drawings and written descriptions, an assessment of archetypes, and oral interviews or presentations.

One of the areas that students should be expected to struggle the most in design-based curricula is in systematically analyzing their results. A variety of studies showed that students are naturally very good at asking questions and in finding patterns for the data that they take to help answer those questions, but that they will not tend to analyze the patterns in terms of their scientific knowledge, nor use their results in future iterations without additional support (Hmelo et al., 2000; Krajcik et al., 1998; Resnick et al., 2000; Seiler et al., 2001). It is therefore advisable for teachers to find ways to guide students in scientifically analyzing their results, and provide a way so that students can see the efficacy of doing so. It is therefore absolutely essential that students record the progress of their thinking and designing (Barak and Doppelt, 2000; Puntambekar and Kolodner, 2005). Putambekar and Kolodner showed the particular efficacy in requiring students to keep journals of their thinking at several different cognitive levels. In particular, these design journals should require students to reflect on how they used their previous experiences and results to improve their future iterations. In addition, there is no reason why teachers could not introduce information from text as a resource for learning that students could use to improve their designs (and explain how they used it). Teachers should resist, though, using these journals to require students to follow a prescribed designing process, or as an overly formalized (and unrealistic) representation of their process, doing so is not necessary or effective (Putambekar and Kolodner).

Finally, many teachers should expect that they will struggle to give up control of class time and content to students as students pursue their projects (Marx et al., 1994). Teachers also, should be prepared to implement an open-inquiry environment in which students respect them and each other, or their environment is likely to fail (Seiler et al., 2001). Furthermore, as much of the scaffolding processes in design-based learning are new to students, individual teachers, and education as a discipline, teachers should expect the successful implementation of design-based science to be a slow process, and one that teachers themselves will need to iterate carefully.

Suggestions for Further Research

While design-based science has emerged in the last two decades as a viable pedagogy and focus of education research, the field of research is still very limited. Much of the research in this field cited by others was from conference proceedings, or reviewed conclusions generated in the design of new curricula without support from evidence, and thus could not be included in this paper. Few large scale studies have compared design-based science to other pedagogies, with the notable exceptions of Mehalik et al. (2008) and to a lesser degree Silk et al. (2009). While several other large scale studies have shown statistically significant learning, they were not able to show that this learning was equivalent to or an improvement upon other pedagogies (Apedoe et al., 2008; Fortus, 2005; Kanter, 2010). In addition, few authors have successfully disaggregated their results for student characteristics, although there is now evidence of student learning in several populations. The disaggregation between “low achieving” and “high achieving” students is not sufficient for determining for which students design-based learning should be expected to be effective, especially given the mixed results when students are grouped that way (Doppelt et al., 2008; Fortus et al. 2004). The conclusion by Doppelt et al. that “low achieving” students may do better when assessed using artifact analysis is important because it indicates that design-based learning may afford effective assessment measures for otherwise underserved students. Overall, in order to determine if design-based science is really a viable pedagogy on which new science education should be built, researchers need to produce several more large scale studies directly comparing the results of design-based science to more established pedagogies, and they need to more seriously attend to student characteristics in doing so.

However, even if researchers more thoroughly establish design-based science as an effective pedagogy, they still need to more rigorously show how it should best be accomplished. Almost every study reviewed in this paper included a vast section on how design-based science should be done but most failed to provide evidence from the research that these were in fact good practices. Even the studies reviewed in the last section of Chapter Two mostly drew their conclusions based on what appeared to work in the classrooms they investigated. The notable exception is the work of Puntambekar and Kolodner (2005), who described an initial enactment of their curriculum that failed to meet their goals, and then the process of changing their curriculum and their results showing that the changes were effective. This sort of iterative research process, called design-based research, provides good evidence of what works and is in fact analogous to the iterative approach to learning in design-based science classrooms (Wang & Hannafin, 2005). Therefore, researchers should more systematically show what strategies work for design-based science either by showing improvement in their own curricula, or by directly comparing similar curricula and analyzing qualitative differences in the enactment and results. In addition, researchers need to be more transparent about why they argue for their pedagogical decisions, and carefully avoid making conclusions about what works and what does not work without basing this on their actual research, especially when reported in their findings and conclusions.

Finally, while this paper reviewed a number of relevant studies of related curricula in the first section of Chapter Two, the studies on design-based science did not explicitly address which conclusions should be relevant from the larger body of research on project-based and problem-based learning, as well as the much larger body of research on constructivist education and collaboration. It is not advisable to re-invent the wheel in designing a new pedagogy, yet conclusions from other pedagogies may not always be transferrable to new ones. For this reason, future researchers should explicitly investigate to what degree lessons from other research in implementing collaborative group work, managing student projects, and interjecting canonical science in open-ended pedagogy is applicable to design-based learning, and to what degree these lessons need to be modified in that setting.

Conclusion

Design-based learning is a young curricular organization that emerged as a formal pedagogy from problem-based learning, project-based learning, and technology education in the early 1990s, although teachers have utilized it in different forms for centuries. This paper reviewed research on the efficacy of design-based courses on science and math learning, as well as relevant research on related pedagogies, and research investigating how design-based science could be most effectively enacted.

This research found that, as in project-based learning, teachers should anticipate challenges in giving over control of content organization and class process to students. At the same time, students are likely to resist pursuing investigation systematically. While research on technology education shows that students are not likely to learn science just from designing technological projects, they are likely to learn it if they experience systematic and careful scaffolding to help them develop more scientific ideas that they can then use to improve their designs. Such effective scaffolding should ideally include a system for students to share and build upon their ideas and evaluate each other's' work in terms of their use of prior learning, and should also include a record keeping device that requires students to reflect on their work and learning, and continue to use what they have learned in their future design work as well as identify science that they learn from their experience in designing.

While some of these conclusions were strong, and the research as a whole strongly suggests that design-based science can teach science, there is a paucity of research investigating if design-based science is actually as effective or more effective than other curricular organizations. In addition, while researchers made claims about what makes effective design-based science, they rarely supported their conclusions systematically, so future research needs to use either iteration of a curricula, or detailed comparison between similar curricula.

While research and curricula development in design-based education is still in its infancy, results have been very promising that it could serve as a powerful pedagogy to motivate and organize student learning in science. In particular, when enacted with care, design-based science can teach students not just the skills and content central to an past mechanistic conception of science, but those central to a future science that is collaborative, creative, and seeks to understand systems that cannot be reduced to the sum of their parts.

REFERENCES

Apedoe, X. S., Reynolds, B., Ellefson, M. R., & Schunn, C. D. (2008). Bringing engineering design into high school science classrooms: The heating/cooling unit. *Journal of Science Education and Technology*, *17*(5), 454–465.

Arends, R. (1997). *Classroom instruction and management*. Boston: McGraw Hill.

Barak, M., & Doppelt, Y. (2000). Using Portfolios to Enhance Creative Thinking. *The Journal of Technology Studies*, *26*(2), 16-25.

Barnett, M. (2005). Engaging inner city students in learning through designing remote operated vehicles. *Journal of Science Education and Technology*, *14*(1), 87–100.

Bishop, J. H., & Mane, F. (2004). The impacts of career-technical education on high school labor market success. *Economics of Education Review*, *23*(4), 381–402.

Brophy, S., Klein, S., Portsmore, M., and Rogers, C. (2008). Advancing engineering education in P-12 classrooms. *Journal of Engineering Education*, *97*(3), 369–387.

Childress, V. W. (1996). Does integrating technology, science, and mathematics improve technological problem solving? A quasi-experiment. *Journal of Technology Education*, *8*(1), 16–26.

Conner, C. D. (2005). *A People’s History of Science*. New York, NY: Nation Books.

Crismond, D. (2001). Learning and using science ideas when doing investigate-and-redesign tasks: A study of naive, novice, and expert designers doing constrained and scaffolded design work. *Journal of Research in Science Teaching*, *38*(7), 791-820.

Dewey, J. (1938). *Experience & Education*. New York, NY: Simon & Schuster.

Doppelt, Y., & Schunn, C. D. (2008). Identifying students' perceptions of the important classroom features affecting learning aspects of a design-based learning environment. *Learning Environments Research*, *11*(3), 195–209.

Doppelt, Y., Mehalik, M. M., Schunn, C. D., Silk, E., & Krysinski, D. (2008). Engagement and achievements: A case study of design-based learning in a science context. *Journal of Science Education*, *19*(2), 22-39.

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.

Fortus, D., Krajcik, J., Dershimer, R. C., Marx, R. W., & Mamlok-Naaman, R. (2005). Design-based science and real-world problem-solving. *International Journal of Science Education*, *27*(7), 855–879.

Hmelo, C. E., Holton, D. L., & Kolodner, J. L. (2000). Designing to learn about complex systems. *The Journal of the Learning Sciences*, *9*(3), 247–298.

Hmelo-Silver, C.E., Duncan, R. G., & Chinn, A. C. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, *42*(2), 99–107.

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.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational psychologist*, *41*(2), 75–86.

Knoll, M. (1997). The project method: Its vocational education origin and international development. *Journal of Industrial Teacher Education*, *34*(3). Retrieved from http://scholar.lib.vt.edu/ejournals/JITE/v34n3/Knoll.html

Krajcik, J. S., Blumenfeld, P. C., Marx, R. W., & Soloway, E. (1994). A collaborative model for helping middle grade science teachers learn project-based instruction. *The Elementary School Journal*, *94*(5), 483-497.

Krajcik, J., Blumenfeld, P. C., Marx, R. W., Bass, K. M., Fredricks, J., & Soloway, E. (1998). Inquiry in project-based science Classrooms: Initial attempts by middle school students. *The Journal of the Learning Sciences*, *7*(3/4), 313-350.

Lewis T. A. (2004). Turn to Engineering: The Continuing Struggle of Technology Education for Legitimization as a School Subject. *Journal of Technology Education*. *16*(1), 21–36.

Lynch, R. L. (1997). *Designing vocational and technical teacher education for the 21st century: Implications from the reform literature* (Contract No. RR93002001). Columbus, OH: Eric Clearinghouse on Adult, Career, and

Vocational Education. Retrieved Dec. 8, 2011 from http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED405499

Marx, R. W., Blumenfeld, P. C., Krajcik, J. S., Blunk, M., Crawford, B., Kelly, B., & Meyer, K. M. (1994). Enacting project-based science: Experiences of four middle grade teachers. *The Elementary School Journal*, *94*(5), 517-538.

Mehalik, M. M., Doppelt, Y., & Schuun, C. D. (2008). Middle-school science through design-based learning versus scripted inquiry: better overall science concept learning and equity gap reduction. *Journal Of Engineering Education*, *97*(1), 71-85.

Mooney, M. A., & Laubach, T. A. (2002). Adventure engineering: A design centered, inquiry based approach to middle grade science and mathematics education. *Journal of Engineering Education*, *91*(3), 309–318.

National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas.* Washington, DC: The National Academies Press.

Panasan, M., & Nuangchalerm, P. (2010). Learning outcomes of project-based and inquiry-based learning activities. *Journal of Social Sciences*, *6*(2), 252-255.

Penner, D. E., Giles, N. D., Lehrer, R., & Schauble, L. (1997). Building functional models: Designing an elbow. *Journal of Research in Science Teaching*, *34*(2), 125–143.

Penner, D. E., Lehrer, R., & Schauble, L. (1998). From physical models to biomechanics: A design-based modeling approach. *The Journal of the Learning Sciences*, *7*(3/4), 429-449.

Project-based learning. (n.d.). In *EducTech Wiki*. Retrieved December 8, 2011, http://edutechwiki.unige.ch/en/Project-based_learning.

Puntambekar, S., & Kolodner, J. L. (2005). Toward implementing distributed scaffolding: Helping students learn science from design. *Journal of Research in*

Science Teaching, *42(*2), 185–217.

Resnick, M. (1996). Beyond the centralized mindset. *The Journal of the Learning Sciences*, *5*(1), 1–22.

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

Roth, W. M. (1995) *Authentic School Science: Knowing and Learning in Open-Inquiry Science Laboratories.* New York, NY: Peter Lang Publishing.

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

Roth, W. M. (2001). Learning science through technological design. *Journal of Research in Science Teaching*, *38*(7), 768–790.

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

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

Seiler, G., Tobin, K., & Sokolic, J. (2001). Design, technology, and science: Sites for learning, resistance, and social reproduction in urban schools. *Journal of Research in Science Teaching*, *38*(7), 746–767.

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

Stone III, J. R., & Aliaga, O. A. (2005). Career & technical education and school-to-work at the end of the 20th century: Participation and outcomes. *Career and Technical Education Research*, *30*(2), 125–144

Stone, J.R., Alfeld, C., & Pearson, D. (2008). Rigor and relevance: enhancing high school students' math skills through career and technical education. *American Educational Research Journal*, *45*(5), 767–795.

Sullivan, F. R. (2008). Robotics and science literacy: Thinking skills, science process skills and systems understanding. *Journal Of Research In Science Teaching*, *45*(3), 373–394.

Tran, N. A., & Nathan, M. J. (2010). Pre-college engineering studies: An investigation of the relationship between pre-college engineering studies and student achievement in science and mathematics. *Journal of Engineering Education*, *99*(2), 143-157.

Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology Research and Development*, *53*(4), 5–23.