Proceedings of the NARST 2007 Annual Meeting (New Orleans, LA, United States)

THE IMPACT OF AN ENGINEERING DESIGN CURRICULUM ON SCIENCE REASONING IN AN URBAN SETTING

This study examines use of engineering design as a means to facilitate students acquiring the knowledge and skills associated with science reasoning in high-needs, urban classrooms. The structure of a design-based curriculum combined with a focus on developing artifacts to satisfy student needs from their everyday lives was used to provide a meaningful context in which to reason scientifically. Using paper-based, multiple-choice assessments of science reasoning, students improved their performance from pre to post assessments, although that improvement was not as great for the lowest SES students. The results compare favorably with a much longer high-quality inquiry science unit, and students outperform those using a traditional textbook curriculum. Implications for the use of design-based curricula as a viable alternative for teaching science reasoning, particularly in urban communities with high concentrations of low socio-economic and minority students, are discussed.

Eli M. Silk, Learning Research & Development Center, University of Pittsburgh
Christian D. Schunn, Learning Research & Development Center, University of Pittsburgh
Mari Strand Cary, Department of Psychology, Carnegie Mellon University

Introduction

Although science teachers have been utilizing design projects in their classrooms for many years, only recently have these design projects begun to be the focus of formal science education research in the United States. The goal of this recent research is to clarify the theoretical foundations upon which the design-based curricula are based and provide empirical evidence of their impact on learning. In design-based science curricula, design is referred to in the sense of engineering or technological design, that is, the construction of an artifact in order to solve an identified need. Design-based learning is a special case of problem- or project-based learning (Barron et al., 1998), in which all the learning activities over an extended period of time are focused on the planning, construction, and evaluation of a design artifact. Inquiry-based curricula and other types of reform and traditional science curricula often incorporate design projects. In most of these curricula, however, the purpose of the design project is to serve as a culminating experience after all of the relevant science knowledge has been acquired. In design-based learning, on the other hand, the design project is the focal point of the classroom activities from the very first lesson of the unit to the last. In addition, only science content and processes that are relevant to the design are introduced, and only at the point in time when they are appropriate to help advance understanding for the design and improve its construction. The possibilities for convergence between design and inquiry in science classrooms are currently the focus of some questions in the education field, but more research is needed, especially in light of the priorities of raising achievement in science for all students (Lewis, 2006).

Background

The Connection Between Design and the Teaching of Science

A number of researchers advocate the use of engineering design as a promising context in which to learn science. Using engineering design as a basis for teaching science has a number of potential advantages, such as better connecting to the knowledge students bring to the classroom, while also providing a clearer sense of utility outside of the classroom. The presence of design concepts and principles in national science standards speaks to the emerging view of experts in the science, technology, and education fields that an understanding of design is complementary to, and supportive of, science literacy (Cajas, 2001). At an even more fundamental level, some of the big ideas that are central to an understanding of design are also central in science, such as an understanding of complex systems (Hmelo, Holton, & Kolodner, 2000) and the use of models (Penner, Lehrer, & Schauble, 1998). Therefore, the fundamental content of design may be mutually reinforcing with that of science, and possibly even shared directly.

In addition to the content, the instructional methods typical in design-based science curricula also have particular appeal to the learning of science. Even though technology educators justified much of its early use in classrooms with anecdotal evidence, typical forms of design-based learning have considerable alignment with contemporary theories of learning. The points of alignment include engaging students as active learners, encouraging students to use metacognitive strategies for self-monitoring and reflection, and supporting classroom communities in which knowledge is distributed such that interaction between members is essential (De Miranda, 2004). By taking a broader view of technology as being *designed*, rather than simply *used*, as it is conventionally taught in schools, design-based science curricula encourage students to solve their own everyday problems in real contexts. In doing so, students may be more likely to question and make sense of the data they collect, rather than distorting data or failing to accept contrary evidence as a result of wanting to confirm their initial beliefs or get the “right” answer (Benenson, 2001). The design artifact is not only the final outcome or product of design-based learning, but also provides opportunities to externalize ideas in concrete representations, which may have advantages for facilitating reasoning about science concepts. Three such advantages include freeing cognitive resources so an individual may more easily analyze her own ideas, forcing an individual to consider real-world physical constraints that may not be represented in mental images, and making the ideas of an individual accessible so that others may provide meaningful critiques (Roth, 2001). Taken together, there is considerable reason to think the development of design activities in which to teach science is something worthy of further consideration and empirical research.

A possible criticism of design-based learning is that, in many cases, the implementation of engineering design projects in classrooms can lead to a notable absence of science as students and teachers get absorbed in the process of construction (Roth, Tobin, & Ritchie, 2001). Without the appropriate amount and quality of reflection, the potential for learning abstract ideas and processes from the design experience is limited (Blumenfeld et al., 1991). Unless design projects are carefully planned to make the connections to science content and processes salient, students will most likely not be successful at making those connections on their own. Further concerns about the use of engineering design as a context in which to learn science have to do with the distinct epistemology of engineering design as compared to scientific inquiry. Many instances of engineering design in classroom settings obscure the authentic practices of design (Leonard, 2004). As a result, students may fail to appreciate engineering design as a discipline in and of itself, while also failing to acquire the ability to participate productively within either the discipline of engineering or science. But these concerns may be attributable to the quality of implementation, much more so than any fundamental issue with the theory advocating design for science. Successful design-based science curricula have addressed the problematic nature of implementation by using built-in supports and consistent structure in the development of materials and unit activities, as discussed below.

Design-for-Science Curricula in Practice

It is a challenge to integrate the theoretical foundations of design-based learning in ways that can be implemented at a high level in science classrooms. Two programs of research have been particularly successful at developing, implementing, and evaluating design-based learning curricula in science classrooms. Learning by Design™ (LBD) is a program developed by researchers at the Georgia Institute of Technology, using problem-based learning and case-based reasoning as its foundations (Kolodner, Camp et al., 2003). The LBD framework includes the use of well-defined rituals and practices (e.g., whiteboarding, messing about, and gallery walks), and embeds those within interacting cycles of design and science for increasing complexity of a concept (Kolodner, Gray, & Fasse, 2003). Observational and quantitative performance data from LBD classrooms have demonstrated that students are able to transfer content knowledge and science process skills from their structured classroom activities to more open-ended collaborations with other students.

A second program, Design-Based Science (DBS) was developed by researchers at the University of Michigan (Fortus, Dershimer, Krajcik, Marx, & Mamlok-Naaman, 2004). In this program, as in LBD, a design cycle is used to structure the class activities, but in this case each successive cycle focuses on different content in the design problem. During consecutive enactments of three different DBS units, students were able to increase their content knowledge and were able to apply that knowledge toward the creation of their design artifact. In a related study, students who participated in one of the DBS units were able to transfer their knowledge gained from participating in the unit to an open-ended design task (Fortus, Krajcik, Dershimer, Marx, & Mamlok-Naaman, 2005). Thus, these enactments of DBS and LBD curricula demonstrate that when implemented with these instructional supports provided by the curricula, design is a viable alternative for students to learn science content and to transfer their science reasoning skills to open-ended tasks.

Limited Research on Design for Science in Low-SES, Urban Settings

One concern about the existing research on the use of design in the science classroom is that the bulk of the research has not been conducted in the most challenging, high-needs settings. The studies using DBS were conducted in a public high school located in an industrial town outside of Detroit. In one study, even though the students were described as being from “blue-collar families”, only 13% of the students were eligible for free or reduced price lunches, and only 13% were minorities (Fortus et al., 2005). No attempt was made in these studies to disaggregate by SES or race/ethnicity, presumably because it was not the focus of the research or possibly because there was not a big enough sample in the low-SES and minority categories.

Research on LBD has considered explicitly the community context as a factor in achievement. One study included contrasts between LBD and non-LBD classrooms from affluent communities and similar contrasts between LBD and non-LBD classrooms from middle-income communities, matching also for teacher expertise and student achievement levels (Holbrook, Gray, Fasse, Camp, & Kolodner, 2001). In these comparisons, the LBD classrooms consistently perform at least as well and often outperform their matched non-LBD classrooms on a variety of measures related to science reasoning and content knowledge. As might be expected, the LBD classrooms from affluent communities do considerably better than all other classrooms. Interestingly though, the LBD classrooms from the middle-income communities perform as well or better than the non-LBD classrooms from the affluent communities (Kolodner, Gray et al., 2003). Though not reported in depth, some data were also reported from an LBD classroom in a "lower-income community." Those data are contrasted with the data from the non-LBD classrooms in the middle-income communities. They indicate that the lower-income classrooms only perform better than their non-LBD comparison classrooms on the general science content items, not on the items targeting the specific content of the unit, nor any other category, such as the general science reasoning items (Holbrook et al., 2001). Thus, in the LBD classrooms, even though there is some evidence that classrooms from middle-income communities using LBD can benefit to the extent that their achievement level approximates traditional classrooms from more affluent communities, there is not enough evidence to make claims about the benefit of LBD in high-needs urban settings.

Another concern is that the measures used by previous studies have not been measures that are directly applicable to high-stakes accountability assessments. In some cases, paper-based, multiple-choice assessments are used for assessing science content only. Science reasoning has been assessed by other means. For instance, in the work using DBS, "designerly" skills are assessed through an open-ended, three-day novel design task done in student teams (Fortus et al., 2005). Although this approach certainly measures real-world problem solving, and it is certainly of value to understand what, if anything, students transfer to real-world settings, such a design task does not directly measure the individual's abstract, general knowledge. In the LBD research, performance tasks from the Performance Assessment Links in Science (PALS) database were also completed in groups, but were complemented with a 32-item paper-and-pencil test completed individually. Seven of the thirty-two items assessed "knowledge of general science practices such as drawing and interpreting graphs and understanding scientific procedure" (Holbrook et al., 2001). To the best of our knowledge, there have been no reports of the specific results from the paper-and-pencil test, only summary statements and item analyses from selected questions. It is not clear if the students were able to transfer the science reasoning knowledge they acquired to that type of assessment. To reiterate, proponents of design-based curricula are reporting favorable results on a number of alternative measures of science reasoning, but the evidence that students are able to transfer the knowledge they learn in these group settings to individual, paper-based, multiple-choice assessments is minimal. Given the pervasiveness of these types of assessments as the primary measure of school progress in the current high-stakes accountability environment, they represent an important reality for urban schools.

Although these implementations of science curricula using engineering design have been successful, they were not conducted in the most challenging urban contexts and not assessed with measures that best approximate accountability measures. Therefore, teaching science through design may not have the same level of success when these more challenging aspects are taken into account. Further exploration is warranted in these high-needs settings using measures that align more closely with accountability concerns.

Methodology

Overview

In this study, we examine pre-post gains on paper-based, multiple-choice assessments of scientific reasoning over the course of a six-week design-based learning unit implemented in eighth grade science classrooms. Two teachers from a mid-size, urban public school district implemented the unit in their classrooms. The students in their classrooms consisted of primarily underrepresented minority, low-SES children. These gains are contrasted conservatively with pre-post gains on common assessment items from assessments of curricula spanning all three years of middle school science. The contrast curricula were implemented in comparable schools in the region and included a successful hands-on, inquiry curriculum as well as a traditional textbook curriculum. The engineering design-based curriculum was implemented starting in the second half of the eighth grade. Therefore, students in this group had already experienced half a year of instruction with these teachers and any large effects due to these teachers would be apparent in scores from the assessment administered at the start of the design-based unit.

Curriculum Context

The context for this research was a design-based learning unit called the *Electrical Alarm System* unit (Doppelt, Mehalik, & Schunn, 2005). The unit focuses on the teaching of core electricity concepts through the design of an electrical alarm system that meets an everyday need. This reform unit supplements the first four-to-six weeks of instruction in an established scripted-inquiry curriculum by incorporating the open-ended, engineering design project as a launching pad for the semester-long study of electronics (Schunn, Millar, Lauffer, & SCALE Immersion Design Team, 2004). The unit has been evaluated previously in terms of its impact on students' learning of the science content relative to students in the scripted-inquiry curriculum alone using a paper-based, multiple-choice assessment of basic electricity concepts. The design-based unit was found to significantly improve students' learning of the science content (i.e., qualitative physics concepts like voltage in series versus parallel circuits), especially for traditionally disadvantaged students (Mehalik, Doppelt, & Schunn, 2005). In this study, we take the evaluation of the design-based learning unit beyond our previous evaluations by assessing its impact on students' general science reasoning.

A number of features are included in the design of the *Electrical Alarm System* unit to improve students' learning of science. The content and problem that focus the unit are centered on students' everyday needs. That is, students begin the unit by evaluating existing electrical alarm systems and the needs those existing systems fulfill in their lives. They then identify places in which they have an existing need for an electrical alarm system that is not being fulfilled. In this way, students connect to their everyday lives and take ownership of a problem, which may increase overall motivation levels (O'Neill & Calabrese Barton, 2005). Having students determine the goal for their alarm system was intended to draw on urban students' cultural funds of knowledge and their voluntary interest in science when it is introduced and situated in non-traditional ways (Seiler, 2001). As a result, focusing the design task on an existing need may help students better understand the meaning of science concepts and appreciate their utility as they are used in context (Bransford, Brown, & Cocking, 2000).

Students work in teams of three to four students, creating the possibility for collaborative learning (Palincsar & Herrenkohl, 2002). They decide as a group what type of electrical alarm system to pursue (e.g., a locker alarm that rings a loud buzzer when someone is trying to break in and a bedroom alarm that calls the student on their cell phone if a sibling tries to enter without permission). Students then go about the process of designing and constructing a customized solution as a team.

Teachers assist as students analyze their problem as an authentic, engineering design task (Ulrich & Eppinger, 2004). This adds important structure to the design process and, as a consequence, better opportunities for learning science. Students generate requirements that their electrical alarm system must meet and evaluate alternative solutions to their problem in terms of their requirements—a process that is similar to proposing alternative hypotheses and choosing among them using evidence. They also break the electrical alarm system down into subsystems so that the complex task is decomposed into simpler tasks that are more manageable and that highlight particular science concepts in isolation. This decomposition strategy is similar to scientific experimentation strategies that systematically focus on the effects of particular variables. In addition, students are scaffolded in understanding the different subsystems as structures that work together to accomplish the task of the system as a whole, thus analyzing the relationship between the subsystems (Hmelo et al., 2000).

The entire design process is scaffolded, but it is not scripted. That is, there is not a single solution path for the electrical alarm system design. Instead, students are given tools to guide investigations. These include structured documentation, open-ended investigative worksheets and help preparing presentations to their peers. One possible danger of tightly scripted activities is that they greatly reduce the focus on process. Students think less about and, in turn, likely learn less about process. Indeed, previous research has detailed the ways in which process largely differentiates authentic scientific inquiry from simple scientific inquiry tasks outlined in science textbooks (Chinn & Malhotra, 2002; Germann, Haskins, & Auls, 1996). The use of scaffolding, rather than scripting, in the *Electrical Alarm System* unit is intended to provide many opportunities for students to engage in practices that are approximations of the complexity of authentic science practices.

Student teams spend a considerable amount of time constructing their alarms. Students are given springboards, batteries, wires, bulbs, LEDS, resistors, and a number of sensors, including a light and heat sensor. While the range of provided materials are wide enough to allow for many different system configurations, the choices are carefully selected so that students are challenged to resolve central conceptual ideas. For instance, students are given 1.5V batteries instead of 9V batteries, so that they are forced to consider different ways to combine batteries in order to provide enough voltage to sufficiently power their system. Throughout the design process, students often bump up against core content in electricity, such as parallel and series circuits, and the relationship between voltage, current, and resistance. When that happens, teachers highlight the problems and provide canonical science explanations to help students make sense of the phenomena and apply those ideas toward improvements of their designs.

Although teachers provided scaffolding for students as they attempted to answer questions about their suboptimal designs and also how electricity works more generally, the primary focus of the unit from the point of view of the students continued to be the goal of designing a working electrical alarm system to meet their chosen need. The help that teachers provided included how to design unconfounded experiments to test their ideas and develop representations and models to make sense of the data students collected. Therefore, science reasoning was an important outcome of the unit, but it was not the only outcome and was not the primary focus.

At many stages during the process, teams are encouraged to share their work with the whole class through presentations. These presentations are an opportunity for teams to share knowledge that they have discovered, as well as get feedback on their ideas and how they can be improved. As in other design-based curricula (Fortus et al., 2004; Kolodner, Camp et al., 2003), the presentations are the primary opportunity for students to get feedback from their peers, and for the teacher to use the students' designs as sources of learning from which the whole class may benefit.

In sum, the unit follows an overall design process of problem identification, decomposition, multiple iterations, and soliciting feedback from experts and peers. Additionally, it contains many elements that could help students learn important aspects of scientific reasoning: proposing alternatives, using evidence to choose among alternatives, and focusing on one variable at a time.

Participants

This research was conducted in two schools in a mid-size, urban public school district, which had officially adopted the design-unit as a supplement to its existing hands-on inquiry curriculum years before. The first teacher was certified originally in special education for the elementary level and taught a self-contained special education class for fourteen years, but then was moved to middle school science and had been teaching that for the past fifteen years. This was her third time implementing the *Electrical Alarm System* unit. All five of her eighth grade science classes participated in the study. Having majored in accounting and earning her master's degree in special education, it was only through her teaching experiences that she learned about or participated in science or engineering. The second teacher had a general science certification for the secondary level and had taught middle school science for six years. This was her second time teaching the *Electrical Alarm System* unit and all three of her eighth grade science classes participated. Her undergraduate major was in Biology, so she did have more experience in science than the other teacher, but did not have any experiences related to engineering. Neither teacher had extensive experience with, or knowledge of, electricity concepts other than that gained from teaching the unit in the past.

Of the 177 students in these classes, data were collected from 170 (the reduction was due to absences, suspensions, etc.). Both schools were in challenging urban environments such that 73% of the students were from an underrepresented minority ethnic background and 83% were from a low socio-economic background (defined as qualifying for government-subsidized free or reduced lunch). The demographics of the students in these particular classrooms were more challenging than the district as a whole, which contains approximately 62% underrepresented minorities and 63% from low-SES backgrounds.[1]

Assessment Instrument

Students were given a paper-based, multiple-choice assessment before the beginning of the unit and the same assessment at the conclusion of the unit. The assessments consisted of commonly-used reliable and valid assessment items. In order to determine the efficacy of the unit in terms of science reasoning and relative to current concerns about accountability, the assessment measured science reasoning, including the control-of-variables strategy (CVS) and the drawing of appropriate conclusions given a set of data.

Of the thirteen items assessing scientific reasoning, eight items were drawn from two validated sources: two were released items from the Third International Mathematics and Science Study (International Association for the Evaluation of Educational Achievement, 1998) and six items were from the Classroom Test of Scientific Reasoning (Lawson, 1978). These six items (see Appendix A for an example) were also used in the contrast curricula assessments, thus facilitating direct comparisons. For use only with the design-based curriculum group, five new items were created to make a more powerful instrument. For example, items used to assess mastery of CVS (Toth, Klahr, & Chen, 2000) were transformed into multiple-choice items that were more characteristic of an item found on an high-stakes accountability assessment (see Appendix B for an example item). We refer to all thirteen of the scientific reasoning assessment items as the *Full Test*, and the subset consisting of the six items from Classroom Test of Scientific Reasoning as the *Reduced Test*. The *Full Test* provides a stronger statistical analysis of gains, and the *Reduced Test* provides a clean comparison to the contrast curriculum.

The assessment also included items that assess other skills and knowledge, but those are not relevant to the current research study. The reliability of the *Reduced Test* was adequate with a coefficient alpha of 0.49 for the pre assessment, 0.68 for the post assessment, and a pre-post correlation of 0.24. For the *Full Test*, the coefficient alpha of the pre assessment was 0.57, the post assessment coefficient alpha was 0.72, and the pre-post correlation was 0.46. Thus, the full assessment was a short, but reliable instrument, which included a subset of items that could be compared easily to the contrast curricula.

Results

Improvement from Pre to Post

First, to establish whether students' science reasoning improved over the course of the design-based unit, a paired means Wilcoxon signed-rank test was performed to examine the significance of mean differences from the pre to post assessment using the *Full Test*. The Wilcoxon test was selected because the data were not normally distributed. The test revealed significant improvement from pre to post ($p < .001$) with a moderate effect size of 0.67 (see Table 1). Although scientific reasoning was a secondary focus of the design unit, students still improved their science reasoning significantly over the course of the unit and were able to transfer those skills to the paper-based, multiple-choice items used in the assessment. The differences between schools were modest (i.e., effect sizes of 0.76 and 0.54).

Table 1: Proportion correct on items from the Full Test (13 items) in the design unit only. The pre assessment for the Electrical Alarm System unit was given in the middle of the 8th grade year, just before starting the unit. Post assessments were given at the end of the 8th grade year.

	Pre-test (SD)	*Post-test (SD)*	*Effect Size*[a]
Electrical Alarm System (N=170)	0.27 (0.17)	0.39 (0.23)	0.67

[a] The effect size was calculated by the difference in the mean between pre and post assessment scores divided by the standard deviation of the pre assessment scores.

Another way to evaluate the improvement of students from pre assessment to post assessment on science reasoning is by examining the advancement of students from very low performance to moderate or high performance. Analyzing the improvement in this way will allow us to better form an explanation for the overall improvement observed. We defined low-performing students as those who answered between 0 and 4 items correctly (i.e., were around or below a score that would be expected by chance alone on the *Full Test*). We distinguished medium-performing students as those who were above chance, but still with considerable room for improvement (i.e., answering between 5 and 9 items correctly). Finally, we distinguished high-performing students as those who were getting most of the items correct (i.e., answering between 10 and 13 items correctly). Table 2 summarizes the proportion of students in each category at the pre assessment and then at the post assessment. The largest advancement resulted for students improving from low performance to medium performance. Although students do not appear to be mastering all of the science reasoning ideas, it seems that many students are improving their reasoning abilities sufficiently enough to make sensible choices rather than simply guessing. This improvement is meaningful given the high-needs context in which the unit was implemented.

Table 2: Proportion of students at high performance (10-13 items correct), medium performance (5-9 items correct), and low performance (0-4 items correct) from pre assessment to post assessment on the Full Test (13 items).

	Pre-Test (SE)	*Post-Test (SE)*	*Change*
High (10-13 correct)	3% (7%)	10% (7%)	+7%
Medium (5-9 correct)	22% (7%)	41% (6%)	+19%
Low (0-4 correct)	74% (4%)	49% (5%)	-25%

Examining the Impact of Student Factors

In order to evaluate the unit according to its impact on traditionally disadvantaged populations, it was necessary to disaggregate the observed improvement by a number of student factors that have been known to influence student achievement. The factors we considered included gender, ethnicity (Caucasian or minority), whether a student is classified as a special education student, and, as a proxy for socio-economic status, whether a student qualified for government-subsidized free or reduced lunch. Table 3 summarizes the improvements on the *Full Test* when the data are disaggregated by these four student factors.

Table 3: Proportion correct on items from the Full Test (13 items) in the design unit only, disaggregated by four student factors. The student factors include gender, ethnicity, classification of special education, and socio-economic status (SES) as measured by qualification for government-subsidized lunch.

	N	*Pre-test (SD)*	*Post-test (SD)*	*Effect Size*[a]
Total	170	0.27 (0.17)	0.39 (0.23)	0.67
Gender				
Male	88	0.28 (0.19)	0.40 (0.23)	0.62
Female	82	0.26 (0.16)	0.37 (0.22)	0.74
Special Ed				
Regular Ed	139	0.28 (0.18)	0.42 (0.23)	0.76
Special Ed	31	0.23 (0.13)	0.26 (0.16)	0.21
Ethnicity				
Caucasian	46	0.34 (0.21)	0.52 (0.23)	0.84
Minority	124	0.24 (0.15)	0.34 (0.21)	0.63
Socio-economic Status				
Non-Subsidized Lunch	29	0.31 (0.21)	0.47 (0.22)	0.80
Subsidized Lunch	141	0.26 (0.17)	0.37 (0.23)	0.64

Because many of the student variables are correlated with one another, it was necessary to use a multiple regression approach to understand which factors were most likely to independently modulate test performance. All four student factors, including gender, ethnicity, special education classification, and qualification for government-subsidized lunch were used as predictors of post assessment scientific reasoning scores on the *Full Test*. In addition, pre assessment scores were used as a controlling predictor. The regression model was only a fair fit to the data ($R^2_{adj} = 31\%$), but the overall relationship was significant ($F_{5,164} = 16.24$, $p < 0.001$). Pre assessment scores ($\beta = .36$, $t_{164} = 5.14$, $p < 0.001$) were positively associated with post assessment scores, and ethnicity ($\beta = -.26$, $t_{164} = -3.71$, $p < 0.001$) and special education ($\beta = -.23$, $t_{164} = -3.62$, $p < 0.001$) were negatively associated with post assessment scores. With the other variables held constant, gender and SES were not significant predictors of post assessment scores. It was notable that gender was not a significant factor, even though some may argue that design-based activities bias against females. No effect of gender on achievement is consistent with other research using design-based learning activities, although it is possible that effects were present in other outcomes, such as confidence, but those were not assessed in this study (Klahr, Triona, & Williams, 2007). Therefore, the analysis suggests that some student factors—ethnicity and special education in particular—did have an impact on student achievement above and beyond other factors.

It may be that these observed differences in post assessment scores for Caucasian versus minority students and regular education versus special education students are due to differences in reading abilities. In order to test this hypothesis we added an additional predictor to the regression analysis, consisting of a standardized reading score from the seventh grade TerraNova assessment. These scores were not available for all the students, so the number of students included in this analysis is 149.[2] With the addition of the standardized reading scores to the first regression model, the fit of the regression model improved significantly ($F_{1,142} = 15.50$, $p <$

0.001), while the model was still only a fair fit overall ($R^2_{adj} = 35\%$, $F_{6,142} = 14.43$, $p < 0.001$). In this second model, the pre assessment ($\beta = .29$, $t_{142} = 3.94$, $p < 0.001$), ethnicity ($\beta = -.15$, $t_{142} = -2.01$, $p < 0.05$), and standardized reading scores ($\beta = .34$, $t_{142} = 3.94$, $p < 0.001$) were significant predictors when controlling for all the other predictors in the model. Hence, the lower performance of special education students may be better explained by differences in prior reading achievement. In other words, special education students are disadvantaged when compared to the regular education students only to the extent that their special needs are related to lower levels of reading ability.

Comparison to an Established Inquiry Curriculum

To investigate whether the improvement in scientific reasoning results that we observed in the design-based unit were strong relative to those found in other middle school curricula, we contrasted our data with data from another inquiry-based science curriculum that has as one of its primary goals to enhance reasoning skills among middle school students of diverse backgrounds. The Model Assisted Reasoning in Science (MARS) curriculum was designed to provide a bridge between the concrete science ideas taught in elementary school and the more abstract science ideas taught in middle school by focusing on the use of models to depict and test ideas (Zimmerman, Raghavan, & Sartoris, 2003). It is a three-year curriculum that spans the middle school grades. The MARS curriculum has shown to be successful in impacting students' scientific reasoning on alternative assessments, such as interviews, and on the Classroom Test of Scientific Reasoning (Raghavan, Sartoris, & Zimmerman, 2002). For that reason, the MARS curriculum provided a strong contrast against which we could evaluate the impact of our short design-based unit.

For this analysis, we examined the significance of mean differences from the pre to post assessment using the *Reduced Test* (6 items that were common in the assessments from both curricula). Overall, the students from the MARS curriculum outperformed the students from our design-based unit on the post assessment (Wilcoxon rank sum test, $p < .001$). Table 4 provides a summary of the mean scores on the pre and post assessment for these items and the associated effect sizes. On these items only, the effect size for the design students from pre to post was 0.58. The MARS curriculum used data from a pre assessment administered at the beginning of the sixth grade year and post assessment administered at the end of the eighth grade year and showed an effect size during that time of 0.81. Therefore, over a time period that was more than two years longer, the inquiry-based curriculum outperformed the design-based curriculum. On the other hand, the effect size was only 25% larger for 6 times as much instructional time, and it is not unlikely that multiple years working with the design-based unit would have similar or possibly even greater impact on students' science reasoning.

Table 4: Proportion correct on Reduced Test items from the Classroom Test of Scientific Reasoning (6 items). The pre assessment for the Electrical Alarm System unit was given in the middle of the 8th grade year, whereas the pre assessments for the MARS and Textbook curricula were given at the beginning of the 6th grade year. Post assessments for all curricula were given at the end of the 8th grade year.

	Pre-Test (SD)	*Post-Test (SD)*	*Effect Size*[a]
Electrical Alarm System (N=170)	0.21 (0.22)	0.34 (0.29)	0.58
MARS (N=614)	0.25 (0.23)	0.43 (0.25)	0.81
Textbook (N=414)	0.21 (0.22)	0.28 (0.33)	0.34

[a] The effect size was calculated by the difference in the mean between pre and post assessment scores divided by the standard deviation of the pre assessment scores.

Further, the MARS curriculum data were mainly collected in two districts, one of which was comparable to the district in which the design-based unit data were collected, but the other district was considerably more affluent, with less than 1% minority and only 8% from a low-SES background.[3] This setting difference may mean that socio-economic differences explain some of the performance advantages of the MARS curriculum. This hypothesis is supported by the eighth grade pre assessment data from the design-based unit being lower than the sixth grade pre assessment data from the MARS curriculum, although only at a marginally significant level (Wilcoxon rank sum test, $p = 0.052$).

Comparison to An Established Textbook Science Curriculum

We also compared the design-based unit to the control classrooms from the MARS dataset, classrooms that taught in the same districts but used a traditional textbook curriculum. Students from the design-based unit outperformed students from these control classrooms on the eighth grade post assessment (Wilcoxon rank sum test, $p < 0.001$). This result suggests that the design-based unit, although possibly not as powerful as a full three-year curriculum focused on inquiry, does significantly improve students' science reasoning scores relative to a full three-year curriculum using a more traditional textbook.

Discussion

The findings reported here add support to the claim that a design-based unit is an effective method by which to teach science reasoning skills. That the students gained in science reasoning is surprising for several reasons. First, scientific inquiry was only a secondary focus of this unit, which was primarily focused on an engineering-design project. Prior to the design-based curricula, the students' overall performance was at chance, suggesting that almost two and half years of an existing hands-on middle school curriculum did not have any observable effect on this measure of students' science reasoning skills. Second, the learning was observed in a low-SES, urban setting. The challenges in classrooms comprised mostly of students from a low socio-economic background are substantial, so that the effect from the design-based unit is particularly meaningful. Third, the learning was measured using a test of far transfer, made up of validated assessment items from paper-based, multiple-choice assessments. Although paper-based, multiple-choice assessments are not the only possible measure of students' science reasoning, and may not detect important skills or knowledge that students may transfer as a result of their learning experiences (Bransford & Schwartz, 1999), these types of measures are particularly relevant for the high-needs, high-accountability context in which the study was conducted (Li, Klahr, & Siler, 2006).

Further results from a regression model revealed that students pre-assessment and reading scores were the largest predictors of the post science reasoning score, although students who did not qualify for government-subsidized free or reduced lunch also performed better than the lower-SES students. In comparison to another much longer research-based inquiry curriculum, students from the design-based unit did not perform quite as well, but they did outperform students in the same dataset from a much longer traditional textbook curriculum. Overall, students from this high-needs urban context improved their ability to perform well on basic items of science reasoning.

There are some limitations to the data that we have presented. One issue is our use of an informal comparison group rather than a carefully matched or randomly assigned control group. We deliberately chose a comparison curriculum that was biased against the achievement of students from the design-based unit, but a cleaner comparison would provide a better estimate of the effect due to the unit itself. Second, our implementation took place with only two teachers, and they were each given considerable support from us during the process. We cannot determine from these data whether bringing the design-based unit to scale, especially in high-needs urban contexts, would limit its effectiveness. Further research must be conducted that demonstrates the reliability of the findings.

Given these limitations, a further concern from the findings in general is that although we observed improvements, gaps still remain between low-SES and high-SES students. Over 80% of the students in our sample qualified for government-subsidized free or reduced lunch, and their post-test scores were 10 percentage points lower on average than the students who did not qualify for subsidized lunches. This suggests that we have not done enough to identify and target the particular needs of low-SES students in ways that would compensate for their disadvantages. Although this result is troubling, indicating that the barriers to achievement for low-SES students are substantial and persistent, it is worth noting that all the students were from a low-SES community and therefore made significant learning gains despite their difficult community context. It could be argued that the design-based unit that we implemented took advantage of cultural funds of knowledge (Seiler, 2001) by connecting with students' personal needs in the design activity, but there are clearly still more challenges in these settings and for these disadvantaged students that are still to be understood (Seiler, Tobin, & Sokolic, 2001). Further instructional innovation is required to capitalize on the resources that students bring to the classroom and help them transition to more formal ways of knowing characteristic of science reasoning. Nevertheless, our work using design-based learning suggests that design could have a powerful effect for the most at-risk students in the most challenging schools.

Curriculum developers need to justify the increased time and resources spent on participating in open-ended science activities when that almost always implies that less time will be spent covering the content in the traditional sense. Further, design-based units must facilitate learning not only of engineering and technology skills, but science content and inquiry skills as well in order to be considered a practical alternative to inquiry-based units. Our findings further those efforts. By articulating and documenting the advantage of design-based units for low-SES populations in particular, these findings will also serve to bolster design as a viable alternative for helping to create systemic change in science education.

References

Barron, B. J. S., Schwartz, D. L., Vye, N. J., Moore, A., Petrosino, A., Zech, L., Bransford, J. D., & The Cognition and Technology Group at Vanderbilt. (1998). Doing with understanding: Lessons from research on problem-and project-based learning. The Journal of the Learning Sciences, 7(3/4), 271-311.

Benenson, G. (2001). The unrealized potential of everyday technology as a context for learning. Journal of Research in Science Teaching, 38(7), 730-745.

Blumenfeld, P. C., Soloway, E., Marx, R. W., Krajcik, J. S., Guzdial, M., & Palincsar, A. S. (1991). Motivating project-based learning: Sustaining the doing, supporting the learning. Educational Psychologist, 26(3), 369-398.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). How People Learn: Brain, Mind, Experience, and School. Washington, D.C.: National Academy Press.

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. Review of Research in Education, 24, 61-100.

Cajas, F. (2001). The science/technology interaction: Implications for science literacy. Journal of Research in Science Teaching, 38(7), 715-729.

Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. Science Education, 86(2), 175-218.

De Miranda, M. A. (2004). The grounding of a discipline: Cognition and instruction in technology education. International Journal of Technology and Design Education, 14, 61-77.

Doppelt, Y., Mehalik, M. M., & Schunn, C. D. (2005). A close-knit collaboration between researchers and teachers for developing and implementing a design-based science module, Annual Meeting of the National Association of Research in Science Teaching (NARST). Dallas, TX.

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

Germann, P. J., Haskins, S., & Auls, S. (1996). Analysis of nine high school biology laboratory manuals: Promoting scientific inquiry. Journal of Research in Science Teaching, 33(5), 475-499.

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.

Holbrook, J. K., Gray, J. T., Fasse, B. B., Camp, P. J., & Kolodner, J. L. (2001). Assessment and evaluation of the Learning by Design™ physical science unit, 1999–2000: A document in progress. Retrieved February, 2007 from http://www.cc.gatech.edu/projects/lbd/pubtopic.html#assess.

International Association for the Evaluation of Educational Achievement. (1998). TIMSS science items: Released set for population 2 (seventh and eighth grades). Retrieved from http://timss.bc.edu/timss1995i/TIMSSPDF/BSItems.pdf.

Klahr, D., Triona, L. M., & Williams, C. (2007). Hands on what? The relative effectiveness of physical versus virtual materials in an engineering design project by middle school children. Journal of Research in Science Teaching, 44(1), 183-203.

Kolodner, J. L., Camp, P. J., Crismond, D., Fasse, B., Gray, J., Holbrook, J., Puntambekar, S., & Ryan, M. (2003). Problem-based learning meets case-based reasoning in the middle-school science classroom: Putting Learning by Design™ into practice. The Journal of the Learning Sciences, 12(4), 495-547.

Kolodner, J. L., Gray, J. T., & Fasse, B. B. (2003). Promoting transfer through case-based reasoning: Rituals and practices in Learning by Design™ classrooms. Cognitive Science Quarterly, 3(2), 183-232.

Lawson, A. E. (1978). The development and validation of a classroom test of formal reasoning. Journal of Research in Science Teaching, 15(1), 11-24.

Lawson, A. E. (1987). Classroom Test of Scientific Reasoning: Revised Paper and Pencil Version. Tempe: Arizona State University.

Leonard, M. J. (2004). Toward epistemologically authentic engineering design activities in the science classroom, Annual Meeting of the National Association for Research in Science Teaching. Vancouver, B.C.

Lewis, T. (2006). Design and inquiry: Bases for an accommodation between science and technology education in the curriculum? Journal of Research in Science Teaching, 43(3), 255-281.

Li, J., Klahr, D., & Siler, S. (2006). What lies beneath the science achievement gap: The challenges of aligning science instruction with standards and tests. Science Educator, 15(1), 1-12.

Mehalik, M. M., Doppelt, Y., & Schunn, C. D. (2005). Addressing performance of a design-based, systems approach for teaching science in eighth grade, Annual Meeting of the National Association of Research in Science Teaching (NARST). Dallas, TX.

O'Neill, T., & Calabrese Barton, A. (2005). Uncovering student ownership in science learning: The making of a student created mini-documentary. School Science and Mathematics, 105(6), 292-301.

Palincsar, A. S., & Herrenkohl, L. R. (2002). Designing collaborative learning contexts. Theory Into Practice, 41(1), 26-32.

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.

Raghavan, K., Sartoris, M. L., & Zimmerman, C. (2002). Impact of model-centered instruction on student learning, Annual Meeting of the National Association of Research in Science Teaching (NARST). New Orleans, LA.

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

Roth, W.-M., Tobin, K., & Ritchie, S. (2001). Re/Constructing Elementary Science. New York: Peter Lang Publishing.

Schunn, C. D., Millar, T., Lauffer, D., & SCALE Immersion Design Team. (2004). Immersing all K-12 Students in Extended Inquiry Science and Design (SCALE Goal 2 Concept Paper). Madison, WI: Wisconsin Center for Education Research.

Seiler, G. (2001). Reversing the "standard" direction: Science emerging from the lives of African American students. Journal of Research in Science Teaching, 38(9), 1000-1014.

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.

Toth, E. E., Klahr, D., & Chen, Z. (2000). Bridging research and practice: A cognitively based classroom intervention for teaching experimentation skills to elementary school children. Cognition and Instruction, 18(4), 423-459.

Ulrich, K. T., & Eppinger, S. D. (2004). Product Design and Development (Third ed.). New York: McGraw-Hill.

Zimmerman, C., Raghavan, K., & Sartoris, M. L. (2003). The impact of the MARS curriculum on students' ability to coordinate theory and evidence. International Journal of Science Education, 25(10), 1247-1271.

Footnotes

[1] The school district data was retrieved August 2006 from Standard & Poor’s SchoolMatters, http://www.schoolmatters.com.

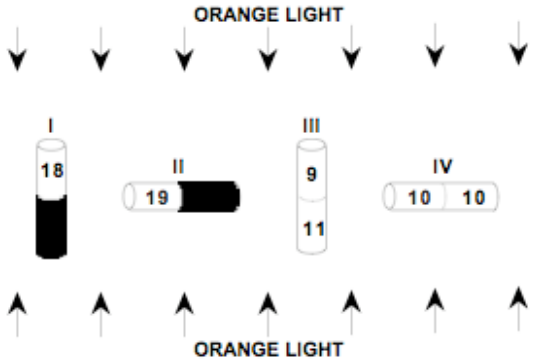
[2] The results of the first regression model (i.e., without reading and math scores) were rerun with the smaller sample of 148 students and the same results were obtained with respect to the significance of the predictors.

[3] The school district data was retrieved August 2006 from Standard & Poor’s SchoolMatters, http://www.schoolmatters.com.

Appendix A

Classroom Test of Scientific Reasoning Example Assessment Item (Lawson, 1987)

Twenty fruit flies are placed in each of four glass tubes. The tubes are sealed. Tubes I and II are partially covered with black paper; Tubes III and IV are not covered. The tubes are placed as shown. Then they are exposed to orange light for five minutes. The number of flies in the uncovered part of each tube is shown in the drawing.



A. These data show that these flies respond to (respond means move to or away from):

a) Orange light but not gravity

b) Gravity but not orange light

c) Both orange light and gravity

d) Neither orange light nor gravity

B. Because

a) Some flies are in both ends of each tube

b) The majority of flies are in the lighted ends and the lower ends of the tubes

c) Most flies went to the bottom of Tubes I and III

d) The flies need light to see and must fly against gravity

e) Most flies are in the lighted end of Tube II but spread about evenly in Tube III

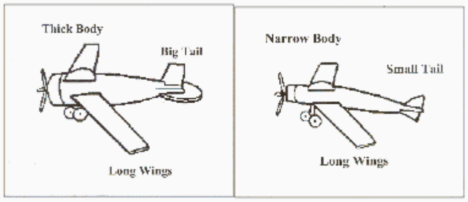
Appendix B

Newly Created Example Assessment Item Based on Prior Research (Toth et al., 2000)

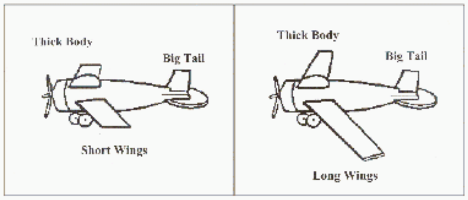
A group of engineers wants to design a model airplane that can fly as fast as possible. They can change the BODY (narrow or thick), the WINGS (long or short), and the TAIL (big or small).

A. If they want to find out whether the length of the WINGS makes a difference, which set of planes should they build?

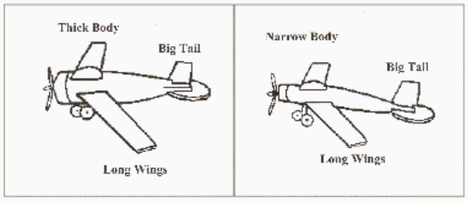
(a)



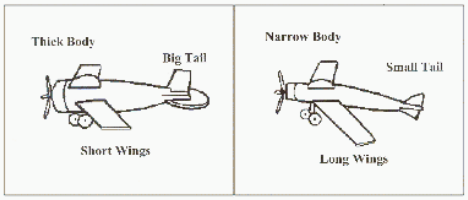
(b)



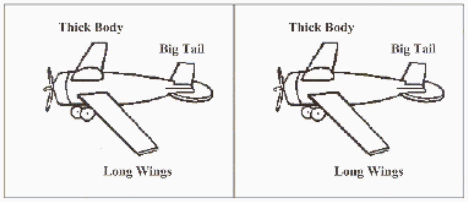
(c)



(d)



(e)



B. Why did you choose that set of planes?

(a) The planes are different in every way

(b) The planes are different in every way except wing length

(c) The planes are the same in every way except wing length

(d) For each plane, wing length and tail shape fit well together

(e) The bodies are big enough to hold the wings