

# The Impact of an Engineering Design Curriculum on Science Reasoning in an Urban Setting

Eli M. Silk · Christian D. Schunn · Mari Strand Cary

Published online: 21 January 2009  
© Springer Science+Business Media, LLC 2009

**Abstract** This study examines the use of engineering design to facilitate science reasoning in high-needs, urban classrooms. The Design for Science unit utilizes scaffolds consistent with reform science instruction to assist students in constructing a design solution to satisfy a need from their everyday lives. This provides a meaningful context in which students could reason scientifically. Eighth grade students from two urban schools participated in the unit. Both schools contained large percentages of racial/ethnic minority and economically disadvantaged students. Students demonstrated statistically significant improvement on a paper-and-pencil, multiple-choice pre and post assessment. The results compare favorably with both a high-quality inquiry science unit and a traditional textbook curriculum. Implications for the use of design-based curricula as a viable alternative for teaching science reasoning in high-needs, urban settings are discussed.

**Keywords** Science reasoning · Engineering design · Design for science · Urban education

## Introduction

Science teachers in the United States have been utilizing design projects in their classrooms for many years. And yet, science education researchers have just begun careful

consideration of design for science (DS)<sup>1</sup> curricula as an alternative to other reform science instructional approaches. A goal of this research is to clarify the theoretical foundations of DS curricula in terms of their potential impact on the development of science reasoning.

Design, in DS curricula, means the construction of an artifact to solve an identified need. Design-based learning is a special case of project-based learning (Barron et al. 1998). The learning activities are focused on the planning, construction, and evaluation of a device or process. Science content and processes are introduced only when they are directly relevant to the design. In contrast, when traditional and inquiry-based science curricula incorporate design projects, they do so as a culminating experience after all of the relevant science has been presented. Education researchers are now reexamining the alternative models for convergence between design and inquiry in science classrooms (Lewis 2006). More research is needed to determine the efficacy of the DS approach and examine whether it contributes to science learning for a wider range of students. This study provides evidence that DS curricula are a promising approach for reforming science classrooms to reach those traditionally underserved.

## Background and Rationale

The presence of design concepts and practices in national science standards speaks to the emerging view of science, technology, and education experts that design is complementary to and supportive of science literacy (Cajas 2001).

---

E. M. Silk (✉) · C. D. Schunn  
Learning Research & Development Center, University  
of Pittsburgh, 3939 O'Hara St, Pittsburgh, PA 15260, USA  
e-mail: esilk@pitt.edu

M. Strand Cary  
Department of Psychology, Carnegie Mellon University,  
Pittsburgh, PA, USA

<sup>1</sup> We make use of the label designated by Leonard (2005), Design for Science or DS, to refer to generic curricula of this form in order to avoid confusion with any particular instantiation (e.g., Design-Based Science [DBS], or Learning by Design<sup>TM</sup> [LBD<sup>TM</sup>]).

Science and technology is one of the eight main categories in the *National Science Education Standards* and includes technological design abilities as a content standard across grade levels (National Research Council 1996). Many of the unifying concepts identified in the *National Science Education Standards* as central to understanding science are also central in design. Those shared unifying concepts include systems and organization (Hmelo et al. 2000), form and function (Crismond 2001), and models (Penner et al. 1998). The prevalence of design in the national standards suggests that there may be a beneficial role for design in science classrooms.

### Science Reasoning in Traditional, Reform, and Design Science Curricula

In addition to learning science content, one of the central goals of science education is to help students learn to reason scientifically (Chinn and Malhotra 2002). Traditional science instruction focuses on students' acquisition of information and skills. Instruction of this type commonly presents knowledge through lectures, readings, and demonstrations in which the teacher or text is the sole authoritative source of knowledge. If scientific inquiry processes are taught at all, they are often taught separate from knowledge-rich contexts. For example, many textbooks may have a chapter at the beginning of the book on the scientific method or on measurement procedures. They are treated as abstract notions to be learned prior to engaging in any meaningful scientific practices. Traditional, textbook-based instruction is especially common in urban schools that are most in need of reform (Haberman 1991; Waxman et al. 1995).

When textbooks do include inquiry tasks, they often obscure authentic science reasoning by utilizing a “cook-book” approach that predetermines many of the cognitively challenging decisions (Germann et al. 1996). For instance, textbooks include simple controls implicit in the setup directions for an experiment, but rarely ask students to consider the issue of control themselves (Chinn and Malhotra 2002). Similarly, textbook inquiry tasks rarely ask students to conduct multiple studies of the same basic procedure in order to explore a wider number of hypotheses and experiments that might help to explain a phenomenon (Chinn and Malhotra 2002). As a consequence, the simple inquiry tasks commonly found in science textbooks are not likely to help students develop an understanding of science reasoning.

Reform science instruction puts more emphasis on knowledge in use. The *National Science Education Standards* (National Research Council 1996) includes a vision of classroom environments and instructional approaches that are consistent with science education reform. Consistent

with this vision are environments that provide opportunities for students and teachers to share responsibility and authority for constructing and making sense of extended investigations in knowledge-rich contexts. We argue that the DS approach can provide one such environment.

Even though technology educators justified much of the early use of design projects based on anecdotal evidence of student interest, typical forms of DS curricula have considerable alignment with contemporary ideas about reform science instruction (De Miranda 2004). DS curricula engage students as active learners responsible for generating and evaluating ideas and designs. DS curricula also rely on constructive interaction between members of the classroom community. Students are accountable to their peers to justify design choices and science explanations with sound reasoning and evidence. In these ways, DS curricula provide rich problem-solving contexts that engage students in scientific inquiry and reasoning.

There are also reasons to believe that DS curricula may have distinctive benefits for engaging students in science reasoning: (1) Using engineering design may help students better recognize the utility of knowledge learned in the classroom for solving real-world problems (Fortus et al. 2005). (2) When students are active problem-solvers in real contexts they are more likely to question the data they collect, rather than distorting or failing to accept contrary evidence as a result of what they believe to be the “right” answer (Benenson 2001). (3) Design artifacts also provide opportunities to externalize ideas in concrete representations. This has multiple advantages including forcing students to account for physical constraints that may not be represented in mental images and providing an accessible representation of one's ideas so that others may provide meaningful critiques (Roth 2001). Thus, the rich contexts of DS curricula may have a beneficial effect on science reasoning despite their primary focus being about design rather than science.

### Design for Science Curricula in Practice

A possible criticism of DS curricula is that the implementation of engineering design projects in classrooms can lead to a notable absence of science reasoning as students and teachers get absorbed in the process of construction (Roth et al. 2001). This may be the result of the different goal orientations that are characteristic of design versus science (Schauble et al. 1991). Students are likely to believe that sensible reasoning in design tasks is exclusively about one's ability to reproduce positive effects rather than to understand the causal structure underlying a phenomenon (Tschirgi 1980). This is not necessarily a negative aspect if this positive-effects orientation is highly motivating for students. Teachers can take advantage of the

high levels of task engagement as leverage for moving students to a point at which they value reasoning scientifically as directly relevant and useful (Schauble et al. 1995). To do this successfully, design projects require careful planning so that the connections are salient. As is true in project-based approaches generally (Blumenfeld et al. 1991), this depends on providing frequent opportunities for students to engage in quality reflection about their design experiences.

The activities of a DS unit utilize built-in supports and consistent structure to help students to transition from their design ideas to reasoning scientifically. Two programs of research have been particularly successful at developing DS curricula to help students make this transition. The first program, design-based science (DBS) used consecutive enactments of its design cycle in three different content areas to help students generalize the process of applying science knowledge to the creation of a design artifact (Fortus et al. 2004). In a related study, students who participated in one of the DBS units were able to transfer general problem-solving skills gained from the unit to an open-ended design task (Fortus et al. 2005). They emphasized making the reasoning processes embodied in the design cycle visible through repeated reflection. The problem-solving skills assessed were more specifically focused on design skills rather than scientific reasoning. Despite this, the study provides evidence that a design context can be effective in helping students to use science ideas to reason through open-ended problems when transfer-promoting features are emphasized.

A second program, learning by design<sup>TM</sup> (LBD) has targeted science reasoning explicitly (Kolodner et al. 2003a). Through interacting cycles of design and inquiry, designs are improved and science concepts are examined at increasing levels of complexity (Kolodner et al. 2003b). Some of the particular curricular scaffolds that supported students' science reasoning included design diaries that encouraged students to justify design decisions using science ideas, pin-up sessions in which students articulated the reasoning behind their in-progress designs to the whole class, and teacher-led whole-class discussions to draw out science ideas that were common across the designs of individual groups (Puntambekar and Kolodner 2005). Observational and quantitative performance data from LBD classrooms demonstrated that students were able to transfer skills related to the design and evaluation of experiments to more open-ended performance assessments of experiment design and analysis (Kolodner et al. 2003b). LBD was able to effectively facilitate the learning of science reasoning in a design context by distinguishing the design practices from the science practices and encouraging an iterative and interactive connection between them.

### Limited Research on Design for Science in Economically Disadvantaged Urban Settings

Most existing research on the use of design in the science classroom has not been conducted in challenging, high-needs settings. This change in setting is likely to have a large impact on the effectiveness or the efficiency of any curriculum or instructional intervention. Researchers who have developed successful laboratory-based instruction targeting science reasoning skills have observed this effect repeatedly (Klahr and Li 2005; Kuhn and Dean 2008; Toth et al. 2000). Their instruction transfers well to privileged school settings but is much less successful in disadvantaged settings. For instance, when learning how to design valid experiments, students in high-achieving affluent schools achieve high levels of mastery in just 2 days. For those in low-achieving urban schools, mastery takes from 1 to 2 weeks depending on the particular conditions of the school or classroom (Li et al. 2006).

Urban settings are associated with many factors that might underlie such learning differences as found in Li et al. (2006), but a central characteristic is the high percentage of students from marginalized and traditionally underserved backgrounds. Understanding the effect of a curriculum on these different student groups is necessary for assessing the overall impact of a curriculum (e.g., Lee et al. 2005; Lynch et al. 2005). Relevant student factors include economic status, gender, race/ethnicity, and special education.

Socioeconomic status is often an important factor at the neighborhood, school, and classroom levels but there is not necessarily a clear effect at the individual student level (Sirin 2005). Most of the variance in learning outcomes that is observed within classrooms may be attributable to other student characteristics. Race/ethnicity is an important factor as the culture of minority students can often be in conflict or misunderstood in existing classroom environments (Norman et al. 2001). Incorporating students' social and cultural lives into the science classroom is one way that racial/ethnic minorities have become more engaged in inquiry-type tasks (Seiler 2001) and may lead to greater achievement.

Regarding gender, it is often assumed that girls are less likely than boys to perform well in science classes and are more likely to lose interest in science in the middle grades. In many cases, though, differences are not observed (e.g., Pine et al. 2006). When differences do exist, the empirical research is not definitive (Jones et al. 2000). Further complicating matters are interactions of race/ethnicity with gender. For example, African-American girls are *more* likely to succeed in science than African-American boys (Kahle et al. 2000). Peer influences that associate male racial/ethnic minorities with opposition to school culture (Graham et al. 1998) may lead students to engage in

classroom behaviors that negatively affect their achievement (Lundy 2003) and that may contribute to this result. When a classroom culture recognizes students' desire to be respected by their peers (Seiler et al. 2001) and maintains high academic expectations for all students (Buxton 2005), gender differences are less likely.

Another important student characteristic to consider are learning disabilities (McCarthy 2005). For students with disabilities, reading ability is one important determinant of success in traditional curricula (Cawley and Parmar 2001). Given that studies have shown that increases in reading ability are the result of increases in reasoning ability (Gleitman and Papafragou 2005; Renner et al. 1973), special education students with reading difficulties may be able to benefit from interventions that target reasoning.

Unfortunately, reported studies using DS curricula have not consistently examined student characteristics. The DBS curriculum units were conducted in a public high school located in an industrial town outside of Detroit in which only 13% of the students were eligible for free or reduced price lunches, and only 13% were minorities (Fortus et al. 2005). The analyses were not disaggregated by student characteristics. There is some research on LBD and socioeconomic status (Holbrook et al. 2001; Kolodner et al. 2003b). LBD classrooms consistently perform at least as well and often outperform non-LBD classrooms that are matched based on community socioeconomic status (Holbrook et al. 2001). 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 et al. 2003b). Though not reported in depth, data from an LBD classroom in a "lower-income community" indicate that the lower-income classrooms only perform better than their non-LBD comparison classrooms on the general science content items and not on any other category (Holbrook et al. 2001). There is not enough evidence from the findings of these studies to make claims about the benefit of LBD, or DS curricula in general, in high-needs urban settings.

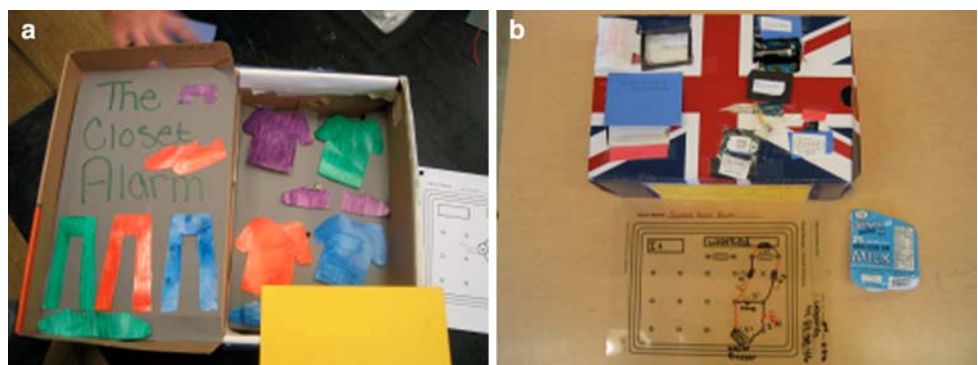
## A Design Unit to Teach Science Reasoning in Urban Settings

The *Electrical Alarm System* (Doppelt et al. 2005) was developed so that it was consistent with current science education reforms and could be effective in high-needs, urban classrooms. Student teams plan, construct, and evaluate an electrical alarm system to solve a personal, everyday need that they identify as a team. The DS unit aligned strongly with the local and state science content standards related to electronics and energy transfer in simple electrical circuits as well as local, state, and national standards relating to the design process and the function and behaviors of systems.

A number of features are included in the design of the *Electrical Alarm System* unit to motivate and support students' science reasoning. One feature differentiating the *Electrical Alarm System* unit from other DS curricula is that students begin by identifying places in which they have a personal need for an electrical alarm system. The range of everyday needs that students identify is remarkably varied and rich (see Fig. 1 for two examples). When students take personal ownership of a problem their overall motivation levels are increased (O'Neill and Calabrese Barton 2005). This approach draws on urban students' cultural funds of knowledge and their voluntary interest in science when it is introduced in non-traditional ways (Seiler 2001). In addition to highlighting personal relevance and stimulating engagement in the task, the needs-based approach provides a context in which students better recognize the utility of targeted science concepts for solving problems they care about (Bransford et al. 2000). This aspect of our DS unit may be especially important in a high-needs urban setting in which students often find little connection to science as it is traditionally presented.

Additional features of the *Electrical Alarm System* curriculum are consistent with other successful DS curricula. Each feature helps to facilitate students' bridge from design activities to science processes. Students work in teams of three to four students, creating the possibility for collaborative learning (Palincsar and Herrenkohl 2002).

**Fig. 1** Personal everyday needs identified by students in the *Electrical Alarm System* unit, the **a** *Closet Alarm* and the **b** *Spoiled Food Alarm*





Student teams analyze their problem as an authentic, engineering design task (Ulrich and Eppinger 2004), which adds important structure to the process. A number of parallels between design and science processes help provide this structure. (1) Students generate requirements for their electrical alarm system and evaluate alternative means of satisfying those requirements. This process is similar to proposing alternative hypotheses and choosing among them using evidence (Kanari and Millar 2004). (2) Students break down the full, complex electrical alarm system into subsystems that are more manageable to study. Each subsystem highlights particular science concepts. This decomposition strategy is similar to scientific experimentation strategies that systematically focus on one variable at a time (Toth et al. 2000). (3) Students then analyze the connections between subsystems in contributing to the function of the system as a whole (Hmelo et al. 2000). This systems approach helps students coordinate the many different variables and consider their additive and interactive effects (Kuhn 2007).

The entire design process is scaffolded, but it is not scripted. There are multiple valid and productive solution paths for the electrical alarm system design. Students are given tools to guide investigations including structured documentation in design portfolios, open-ended investigative worksheets, and help preparing presentations to their peers at multiple steps in the process (see Appendix 1 for an example). The use of scaffolding, rather than scripting, in the *Electrical Alarm System* unit is intended to provide many opportunities for students to engage in the complexity of authentic science reasoning without removing the cognitively challenging aspects.

Student teams construct their alarms with springboards, batteries, wires, bulbs, LEDs, resistors, and a number of sensors, including a light and heat sensor. The range of provided materials allow for many different system configurations, but the choices are carefully selected so that students are challenged to resolve central conceptual ideas (Sadler et al. 2000). For instance, students are given 1.5 V batteries instead of 9 V batteries, so that they are forced to consider different ways to combine batteries to sufficiently power their system. Throughout the design process, students bump up against core content in electricity as they try to explain the behavior of their system, especially when it fails. Teachers highlight the problems, encourage students to generate ideas that explain the behavior they observe, and provide guidance on how to design unconfounded experiments to test their ideas. As appropriate, they also provide canonical science explanations to help students make sense of their tests and then apply those ideas to improve their designs. The instruction is always motivated directly from students' designs. Though students' primary focus remains on designing an electrical alarm system to

meet their chosen need, many opportunities to apply scientific reasoning emerge from a need to understand and improve their designs in a systematic way.

At many stages during the process, teams are encouraged to share their designs and the knowledge that they have discovered with the whole class. As in the other DS curricula described earlier (Fortus et al. 2004; Kolodner et al. 2003a), the presentations are the primary opportunity for the teacher and peers to give feedback on ideas, to critique supporting evidence, to provide suggestions for improvement, and to gather ideas to inform redesigns and further investigations. It is also an opportunity for the teacher to highlight common important conceptual themes and reasoning processes teams used to justify their design decisions.

In sum, the unit follows an overall design process that includes problem identification, decomposition, multiple iterations, sharing information, and soliciting feedback from teachers and peers. For a more detailed discussion of this type of instructional approach, see Apedoe et al. (2008). The features of the DS unit help students learn important aspects of scientific reasoning including how to design tests that focus on the effect of one variable at a time on a system, proposing alternative hypotheses to explain the observed behavior of the system, and using evidence to choose among those alternatives explanations.

The *Electrical Alarm System* unit was evaluated previously using a multiple-choice test of electricity concepts and compared to students exposed only to a scripted-inquiry unit targeting the same concepts (Mehalik et al. 2008). The DS unit was found to significantly improve students' learning of the targeted science content, and was most helpful to traditionally disadvantaged students. In this study, we take the evaluation of the DS unit beyond our previous evaluations by assessing its impact on students' domain-general science reasoning.

## Method

In this two-part study, we examine the efficacy of a DS unit to facilitate learning of domain-general science reasoning in a high-needs urban setting. In the first part of the study, we examine an implementation of the DS unit by two-eighth grade science teachers, each in schools with high percentages of minority and economically disadvantaged students. We examine pre–post gains of scientific reasoning over the course of the 8 week DS unit. We disaggregate the pre–post gains by student demographic factors. In the second part of the study we contrast the magnitude of the gains observed in the DS unit to two other approaches to teaching science reasoning, a successful inquiry-based curriculum and a traditional textbook-based curriculum.

## Participants

The DS curriculum unit examined in this study was the *Electrical Alarm System*. We refer to the students who participated in the DS unit as the *Design* group. Two-eighth grade science teachers implemented the DS unit. Both teachers taught in separate high-needs schools located in a mid-size, urban public school district. They were selected because they were both very familiar with the DS unit from having taught it successfully at least one time prior to the study. They each had been teaching science at the middle school level for at least 5 years, but had limited understanding of electronics content or engineering design outside of what they learned from teaching the DS unit in previous years.

Of the 177 students in the eight sections taught by the DS unit teachers, data were collected from 170 (the reduction was due to excessive absences, suspensions, etc.). Over 70% of the students who participated were from an under-represented minority racial/ethnic background and over 80% were from an economically disadvantaged background (defined as qualifying for government-subsidized free or reduced price lunch). The demographics of the students in the particular classrooms were more challenging than the district as a whole, which contains approximately 63% underrepresented racial/ethnic minorities and 60% from economically disadvantaged backgrounds.<sup>2</sup> The classrooms did not contain any English Language Learners and very few students from racial/ethnic backgrounds other than Caucasian or African–American. This reflected the demographic make-up of the particular urban district. Table 1 summarizes the demographics of the students in the Design group. The high percentages of racial/ethnic minorities and economically disadvantaged students provided a suitable context for examining the efficacy of DS curriculum in a high-needs urban setting.

The contrast schools were determined subsequent to the implementation of the DS unit to contrast the gains from the Design group with both traditional and reform-based instruction. For the reform curriculum unit, we chose model assisted reasoning in science (MARS). One of its primary goals is to enhance science-reasoning skills among middle school students of diverse backgrounds. The curriculum focuses on helping middle school students reason about science ideas through the generation, use, and refinement of models (Zimmerman et al. 2003). A central aspect of the instruction in MARS is a computer software environment that allows students to generate a theoretical model of a situation, run their generated model in the

**Table 1** Demographic information of students participating in the design for science unit

	Design school 1 ( <i>n</i> = 101)	Design school 2 ( <i>n</i> = 69)	Design combined ( <i>N</i> = 170)
Female (%)	51.5	43.5	48.2
Racial/Ethnic minority (%)	74.3	71.0	72.9
African–American (%)	70.3	71.0	70.6
Other (%)	4.0	0.0	2.3
Economically disadvantaged (%) <sup>a</sup>	83.2	82.6	82.9
Special education (%)	16.8	20.3	18.2

<sup>a</sup> Students are classified as economically disadvantaged if they are eligible to receive free or reduced-priced lunches under the National School Lunch Program. Eligibility for the National School Lunch Program is based on household income status

software to view the outcome that their model predicts, and then compare that prediction to a run of the “real” model (i.e., the canonical science model). These computer tasks are coordinated with hands-on experiments using physical materials. The situations that students model are carefully chosen to elicit common misconceptions. An emphasis on model revision, over and above model generation alone, is the primary basis by which the MARS curriculum helps students to acquire domain-general science reasoning skills that result in the flexible coordination of experimentation, theory, and evidence. Although labeled explicitly as model-based, the MARS curriculum has the essential elements of inquiry-based instruction as students are challenged to think actively about the scientific reasoning that is required to make sense of knowledge-rich contexts. It is an exemplar in many ways for curriculum consistent with current science education reform.

The MARS curriculum has been successful in impacting students’ scientific reasoning on paper-and-pencil and alternative assessments of scientific reasoning (Raghavan et al. 2002). Students from the MARS curriculum have also been shown to perform better on hands-on and paper-and-pencil inquiry tasks when compared to students instructed with a nationally-recognized NSF-supported inquiry curriculum (Zimmerman et al. 2003). The MARS curriculum had been implemented in many schools throughout the region with various content modules for each middle school grade. We chose a MARS curriculum implementation that was considered to be of high quality by the MARS curriculum developers. It was implemented by two-eighth grade teachers in one middle school located just outside of the urban district of the design group. In the subsequent analyses, we refer to this group as the *Inquiry* group.

We also compared the DS unit performance to that found in a comparable school in the region using a more traditional textbook-based curriculum for teaching science.

<sup>2</sup> The school district data was retrieved November 2007 from Standard & Poor’s School Matters, <http://www.schoolmatters.com>. The data is from the 2005–2006 school year.

In this approach students were directly and explicitly presented with explanations of abstract science concepts. Students were not asked to generate their own explanatory models, experiments or hypotheses. The implementation of this textbook-based approach also occurred in one middle school located just outside of the urban district of the Design group and was implemented by three teachers in that school. In the subsequent analyses, we refer to this group as the *Textbook* group.

Student demographic information was not available at the individual level for these data. Table 2 summarizes the demographics at the school level of the two contrast groups, Inquiry and Textbook, as compared to the schools from the Design group.

### Procedure

The two Design teachers attended two 3 h sessions of professional development during which the research study was introduced and the teachers practiced ways to integrate as-needed science reasoning instruction into the DS unit. During the implementation of the DS unit by the Design teachers, researchers observed the classes and did not interact with the students or teacher during class time.

Students were given an identical paper-and-pencil assessment before and after instruction, although the time of year the assessment was administered and the items included on each assessment varied across groups. The curricula for the Design group and for the Inquiry group were both implemented only in the second half of eighth grade. As a result, both Design and Inquiry students took the pre assessment in the middle of eighth grade, just before beginning their respective units, and then at the conclusion of their respective units, close to the end of the eighth grade year. Thus, between the Design and Inquiry group, we have the ideal contrast of the reasoning test administered in the middle and end of eighth grade. In contrast, students in the Textbook group took the pre assessment test at the conclusion of the sixth grade year

and then the post assessment at the end of eighth grade. Thus, for the Textbook group, we have data from a longer gap: end of sixth grade to the end of eighth grade. Since textbooks place much less emphasis on reasoning and such content is often reserved for the beginning of the year, examining reasoning gains from this longer span may, in fact, be the more reasonable contrast.

### Assessment Instrument

A paper-and-pencil, multiple-choice assessment was utilized as the form of the assessment of science reasoning in all three groups. This ensured the classroom teacher could administer it to a whole class in a single period without specialized training. Although tasks that require use of real data in an interview setting may be the most authentic method for assessing science reasoning, such an approach is not always feasible or even necessary when the paper-and-pencil version is carefully-designed to capture the same underlying construct (Lawson 1978). In addition, researchers have observed a large amount of variability in student performance across particular performance assessment tasks (Pine et al. 2006). Paper-and-pencil allows one to sample across reasoning in more situations, creating a more curriculum-neutral measure (Zimmerman et al. 2003). In order to ensure the reliability and validity of the measure for domain-general science reasoning, we used validated sources to select the assessment items.

Each group was given an assessment created specifically for the particular curriculum used in that group. Common across all three assessments were six items that assessed science reasoning from Lawson's Classroom Test of Scientific Reasoning (Lawson 1987). The DS unit used in the Design group specifically targeted two aspects of science reasoning—controlling variables in the design of experiments and drawing conclusions about the relationships among variables from data—so only the items from the Classroom Test of Scientific Reasoning that measured those aspects were included (see Appendix 2 for an example).

**Table 2** Demographic information of students from the design and contrast schools

	Design school 1	Design school 2	Inquiry school	Textbook school
Urban status	Urban	Urban	Urban fringe	Urban fringe
Female (%)	44.7	43.3	49.8	44.7
Racial/Ethnic minority (%)	74.0	72.7	39.3	48.0
African-American (%)	72.8	71.9	36.2	46.9
Other (%)	1.2	0.8	3.1	1.1
Economically disadvantaged (%)	78.6	80.1	36.2	59.0 <sup>a</sup>

Note: Data retrieved from <http://www.schooldatairect.org> for the 2006 school year

<sup>a</sup> This data point was retrieved from the state's department of education website (<http://www.pde.state.pa.us/k12statistics/site/default.asp>), as the percentage provided in the national database (0%) was not realistic

The assessments used in all three groups also included items that assess other aspects of science reasoning or conceptual knowledge specific to each curriculum, but those were not comparable across curricula and were not relevant to the current research study. Seven other items were included in the assessment of science reasoning for the Design group to make a more powerful instrument in which to examine differences at the individual student level. These additional seven items targeted the same aspects of science reasoning as the six items from the Classroom Test of Scientific Reasoning (for a total of 13 scientific reasoning items on the Design group assessment). Of these additional items, two were validated items selected from the set of Third International Mathematics and Science Study released items (International Association for the Evaluation of Educational Achievement 1998). The others were new items created by the research team. For example, items used to assess mastery of the control of variables strategy (Toth et al. 2000) were transformed into multiple-choice items (see Appendix 3 for an example).

In summary, we refer to all 13 of the scientific reasoning assessment items used only with the Design group as the *Full Test*, and the subset consisting of the six items from Classroom Test of Scientific Reasoning that were used in all three groups as the *Reduced Test*. The *Full Test* provides a stronger statistical analysis of gains within the Design group at the individual student level, and the *Reduced Test* provides a clean comparison to the contrast curricular approaches. 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 (See Fig. 2

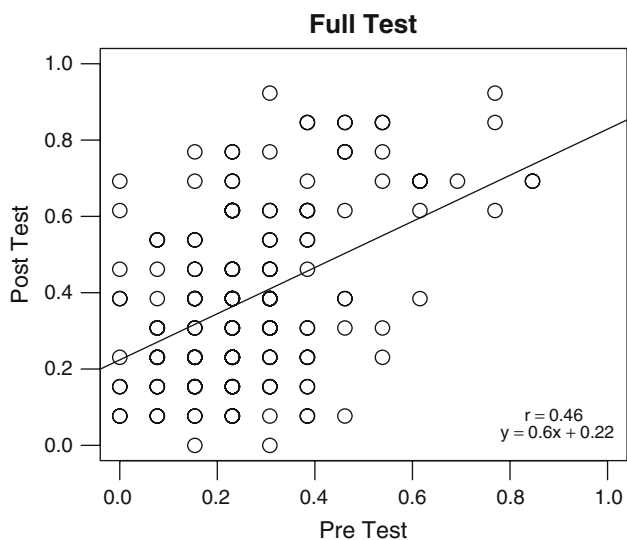
for a scatter plot of the pre-test versus post-test on the *Full Test*). Thus, the full assessment was a short, but reliable instrument, which included a subset of items that were used for comparisons to the contrast curricula.

**Results**

**Improvement From Pre to Post in the Design Group**

First, to establish whether students’ science reasoning improved for the Design group over the course of the DS unit, a paired means Wilcoxon signed-rank test was performed to examine the statistical significance of gains 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 statistically significant improvement from pre to post ( $p < 0.001$ ) with an effect size of 0.58 (see Table 3), which is a moderate effect size (Cohen 1992). Although the primary activities in the unit involved design (i.e., scientific reasoning was a secondary focus of the DS unit from the students’ perspective), the Design students still improved their science reasoning over the course of the unit. At the pre assessment there was no difference between the Design schools, and the difference at the post assessment was modest, but not statistically significant. Because we did not observe a statistically significant difference between Design schools, we collapsed across the schools for all further analyses (although we confirmed that the general patterns for all analyses were similar for both schools).

Another way to evaluate the improvement of Design students from pre assessment to post assessment on science reasoning was by examining the advancement of students from very low performance to moderate or high performance. Analyzing the improvement in this way allowed us to better characterize 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 defined medium-performing students as those



**Fig. 2** Scatter plot of pre-test versus post-test scores on the full test

**Table 3** Mean proportion correct for design students on the full test (13 items) of science reasoning

	Pre-test mean (SD)	Post-test mean (SD)	Effect size
Design schools combined (N = 170)	0.27 (0.17)	0.39 (0.23)	0.58***
Design school 1 (n = 101)	0.27 (0.17)	0.37 (0.22)	0.49***
Design school 2 (n = 69)	0.27 (0.19)	0.41 (0.24)	0.69***

\*\*\*  $p < 0.001$



**Table 4** Change in performance levels of design students ( $N = 170$ ) on the full test (13 items) of science reasoning

Performance level	Pre-test (SE)	Post-test (SE)	Change (%)
High (10–13 correct)	3% (1%)	9% (2%)	+6
Medium (5–9 correct)	23% (3%)	41% (4%)	+18
Low (0–4 correct)	74% (3%)	50% (4%)	–24

who were above chance, but still with considerable room for improvement (i.e., answering between 5 and 9 items correctly). Finally, we defined high-performing students as those who were getting most of the items correct (i.e., answering between 10 and 13 items correctly). Table 4 summarizes the percentage of Design students in each category at the pre assessment and then at the post assessment. It is important to note that a large majority of the Design students (74%) were at chance before beginning the DS unit, despite already participating in half of a year of instruction in science with their teachers in addition to their science classes from previous years. At the conclusion of the DS unit, the largest advancement resulted for Design students improving from low performance to medium performance. Although the Design students did not appear to be mastering all of the science reasoning ideas, it seems that many students did improve their reasoning abilities sufficiently enough to make sensible choices rather than simply guessing. This improvement was meaningful given the high-needs context in which the unit was implemented and the failure of their previous classroom experiences to improve their science reasoning.

#### Examining the Impact of Student Factors in the Design Group

In order to evaluate the DS unit according to its impact on traditionally disadvantaged populations, it was necessary to disaggregate the observed improvement in the Design group by a number of student factors that have been known to influence student achievement. The factors we considered included gender, race/ethnicity (Caucasian or minority), whether a student is classified as a special education student, and, as a proxy for economic status, whether a student qualified for government-subsidized free or reduced lunch. Table 5 summarizes the improvements on the *Full Test* when the data from the Design group are disaggregated by these four student factors. In all cases, the students in the Design group improved from pre to post, except the special education students.

Because many of the student variables were 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, race/ethnicity, special

**Table 5** Mean proportion correct for design students on the full test (13 items) of science reasoning disaggregated by demographic factors

	$n$	Pre-test mean (SD)	Post-test mean (SD)	Effect size
Total	170	0.27 (0.17)	0.39 (0.23)	0.58***
Gender				
Male	88	0.28 (0.19)	0.40 (0.23)	0.55***
Female	82	0.26 (0.16)	0.37 (0.22)	0.61***
Race/Ethnicity				
Caucasian	46	0.34 (0.21)	0.52 (0.23)	0.81***
Minority	124	0.24 (0.15)	0.34 (0.21)	0.52***
Economically disadvantaged				
Non-subsidized lunch	29	0.31 (0.21)	0.47 (0.22)	0.79**
Subsidized lunch	141	0.26 (0.17)	0.37 (0.23)	0.54***
Special education				
Regular education	139	0.28 (0.18)	0.42 (0.23)	0.66***
Special education	31	0.23 (0.13)	0.26 (0.16)	0.19

\*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

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. Because reading ability is an important factor predicting success in a number of the student groups that we considered, students' prior reading ability was added as an additional predictor in a second model. Students' prior reading ability was assessed using 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 both regression analyses is 148.<sup>3</sup> We analyzed the two regression models using a step-wise method, so that we could examine the added predictive value of including prior reading ability. A summary of the results from both models is provided in Table 6.

The initial regression model was only a fair fit to the data ( $R_{adj}^2 = 29\%$ ), but the overall relationship was statistically significant ( $F_{5,142} = 12.98$ ,  $p < 0.001$ ). Pre assessment scores ( $\beta = 0.38$ ,  $t_{142} = 5.14$ ,  $p < 0.001$ ) were positively associated with post assessment scores. Being from a minority racial/ethnic background ( $\beta = -0.49$ ,  $t_{142} = -2.94$ ,  $p < 0.01$ ) and being designated as a special education student ( $\beta = -0.62$ ,  $t_{142} = -3.36$ ,  $p < 0.001$ ) were negatively associated with post assessment scores. With the other variables held constant, gender and students from economically disadvantaged backgrounds were not statistically significant predictors of post assessment

<sup>3</sup> The results of the initial regression model (i.e., without standardized reading scores) were rerun with the larger sample of 170 students and the same results were obtained with respect to the statistical significance of the predictors.

**Table 6** Summary of multiple regression analysis for variables predicting post test science reasoning score of design students ( $N = 148$ )

Variable	$B$	SE ( $B$ )	$\beta$
Step 1			
Pre science reasoning	0.50	0.10	0.38***
Female	-0.02	0.03	-0.09
Racial/Ethnic minority	-0.11	0.04	-0.49**
Economically disadvantaged	-0.02	0.04	-0.08
Special education	-0.14	0.04	-0.62***
Step 2			
Pre science reasoning	0.38	0.10	0.29***
Female	-0.04	0.03	-0.17
Racial/Ethnic minority	-0.08	0.04	-0.33*
Economically disadvantaged	-0.00	0.04	-0.01
Special education	-0.04	0.05	-0.19
Seventh grade TerraNova reading	<0.01	0.00	0.34***

Note:  $R_{adj}^2 = 29\%$  for Step 1;  $\Delta R_{adj}^2 = 6\%$  for Step 2 ( $p < 0.001$ )

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

scores. Overall, the analysis suggests that some student factors—racial/ethnic minority and special education in particular—did have an impact on student achievement above and beyond other factors in the Design group.

With the addition of the standardized reading scores to the regression model, the fit of the revised regression model improved a statistically significant amount ( $\Delta R_{adj}^2 = 6\%$ ,  $F_{1,141} = 14.96$ ,  $p < 0.001$ ), while the model was still only a fair fit overall ( $R_{adj}^2 = 35\%$ ,  $F_{6,141} = 14.38$ ,  $p < 0.001$ ). In this second model, the pre assessment ( $\beta = 0.29$ ,  $t_{141} = 3.95$ ,  $p < 0.001$ ), being from a racial/ethnic minority background ( $\beta = -0.33$ ,  $t_{142} = -2.02$ ,  $p < 0.05$ ), and standardized reading scores ( $\beta = 0.34$ ,  $t_{142} = 3.87$ ,  $p < 0.001$ ) were statistically significant predictors when controlling for all the other predictors in the model. In this revised model, being designated as a special education student was no longer a statistically significant predictor. Hence, the lower performance of special education students in the Design group may have been partially explained by differences in prior reading ability. The effects of race/ethnicity could not be entirely attributed to reading ability, although the strength of the relationship between race/ethnicity and reasoning was reduced when reading ability was included ( $\beta = -0.49$ ,  $p < 0.01$  in the initial model compared to  $\beta = -0.33$ ,  $p < 0.05$  in the revised model).

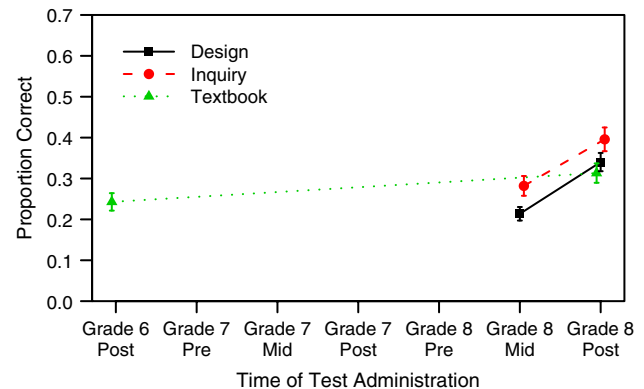
#### Comparison of Effectiveness of Design, Inquiry and Textbook Curricula

For this analysis, we examined the statistical significance of gains from the pre to post assessment using the *Reduced Test* (i.e., 6 items that were common in the assessments

**Table 7** Mean proportion correct on the reduced test (6 items) of science reasoning

	End of 6th grade mean (SD)	Middle of 8th grade mean (SD)	End of 8th grade mean (SD)	Effect size
Design ( $n = 170$ )	—	0.21 (0.22)	0.34 (0.29)	0.49***
Inquiry ( $n = 120$ )	—	0.28 (0.26)	0.40 (0.32)	0.39***
Textbook ( $n = 116$ )	0.24 (0.23)	—	0.31 (0.26)	0.29*

\*  $p < 0.05$ , \*\*\*  $p < 0.001$

**Fig. 3** Mean performance for science reasoning (+SE)

from all three curricula). In all three groups, there was a statistically significant increase from pre to post (Table 7). The Design group had the largest effect size, followed by the Inquiry group, even though those two interventions occurred over a much shorter time period than the Textbook group, which is graphically represented in Fig. 3. Note that the pattern across groups is identical whether one uses means or medians.

To test whether the relative increases were different at a statistically significant level, we used the assessment administered to the Textbook group at the end of sixth grade and the assessments administered to the Design and Inquiry groups at the end of eighth grade as the covariate in an ANCOVA. In this analysis, there was not a difference between groups ( $F_{2,402} = 1.79$ ,  $p = 0.17$ ).

## Discussion

### Summary of the Results

Students' science reasoning test scores increased a statistically significant amount after participating in the DS unit. This gain in science reasoning is educationally important for several reasons. First, the learning was observed in an urban setting with large proportions of economically

disadvantaged and racial/ethnic minority students. Prior to participating in the DS unit, the students' overall performance was nearly at chance, suggesting that almost two and half years of an existing hands-on middle school curriculum did not have much observable effect on their science reasoning abilities. The challenges in classrooms comprised mostly of students from traditionally marginalized and underserved backgrounds are substantial, so that the positive effect from the DS unit is particularly meaningful. Although students finished with far from perfect performance, it is important to keep in mind that it is unreasonable to expect that one middle school unit will produce students with an entirely coherent and flexible ability to reason scientifically. On the other hand, successful experiences in this unit can serve as the foundation for continued development of the ability to do scientific inquiry that would result from multiple, frequent opportunities to engage in similar sorts of activities (National Research Council 1996).

Second, the DS unit did incorporate instruction in formal science reasoning, but the primary focus from the students' point of view was the engineering design project. All of the unit activities were centered on the design project first and foremost. Whenever there were opportunities to incorporate formal science reasoning in ways that were directly relevant to the design project, the DS unit was organized such that they would be emphasized. Therefore, the design project provided a rich context in which students could appreciate the value of reasoning scientifically.

Third, economic disadvantage and gender were not statistically significant predictors at the student level when controlling for other factors, but race/ethnicity and special education status were. Both of these factors seem to be at least partially explained by students' reading ability. It was notable that gender was not a statistically significant factor, even though some may argue that engineering design activities bias against females.

To evaluate the magnitude of the gains made in the DS unit with the Design group, we compared them to gains made in the Inquiry and Textbook curricula. All three curricula showed statistically significant gains from pre to post, but the Design group showed the largest gain. The success of the DS unit is important for several reasons. First, the DS unit and the Inquiry curriculum took place over a much shorter time frame than the Textbook curriculum, but still outperformed it. This suggests that students' knowledge of science reasoning improves if it is taught explicitly in a rich context, and that traditional curricula that do not explicitly teach science reasoning are not as effective, even over longer time periods.

Second, the DS unit outperformed the other curricula despite the fact that it took place in the most challenging

setting of the three. Although we chose high-needs contexts for all three groups in an attempt to make an appropriate comparison, the schools in the contrast groups were still not disadvantaged to the same extent that the schools in the Design group were. This setting difference may explain some of the performance advantages of the Inquiry curriculum at the pre-test. Using ANCOVA, we found no statistically significant differences among the gains made by the three curricula. However, this comparison is imperfect due to the fact that the control group was post hoc rather than part of the study design. To more definitively determine how design compares to inquiry and textbook curricula, the design of future studies should employ a more controlled design in which the participating schools are more closely matched.

### Study Limitations

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 DS 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 moderate professional development from us during the process. We cannot determine from these data whether bringing the DS unit to scale would limit its effectiveness. This may be an especially relevant concern in high-needs urban contexts that often provide very minimal professional development. Further research must be conducted that demonstrates the robustness of the findings.

Another concern may be related to the choice of measure that we used to assess science reasoning. A case could be made that many science inquiry standards—such as formulating scientific explanations or communicating scientific understanding—cannot be adequately measured using a multiple-choice format because of the oversimplification of the complexity of inquiry tasks and the rigidity of only recognizing one correct answer when multiple are possible (Wilson and Bertenthal 2006). Paper-and-pencil, multiple-choice assessments are certainly not the only possible measure of students' science reasoning, but assessments of this form have been developed successfully to measure complex cognitive processes (e.g., Hestenes et al. 1992). We did choose a measure that aligned with the particular types of science reasoning that we targeted in the DS unit. In addition, the Classroom Test of Scientific Reasoning has been validated as a measure of formal science reasoning (Lawson 1978). In some cases, the test

predicts success in learning science content as a result of inquiry-based instruction better than prior knowledge of the content (Coletta and Phillips 2005; Johnson and Lawson 1998). Furthermore, if anything, the use of a paper-and-pencil, multiple-choice assessment format may have biased against our population of normally low-achieving and racial/ethnic minority students (Lawrenz et al. 2001), effectively underestimating the gains that they made. Nevertheless, our assessment approach could be enhanced by using multiple types of measures (Wilson and Bertenthal 2006), or by using measures that are sensitive to dynamic forms of transfer (Bransford and Schwartz 1999). Further research would certainly benefit from an attempt to assess in more detail the types of understanding that students develop in this DS context.

### Implications and Future Work

On the whole all students improved after participation in the DS unit despite being in a high-needs, economically disadvantaged community. But traditional achievement gaps still remained. Students from traditionally underserved populations started at a lower level of achievement than their more mainstream peers and also ended at a lower level (e.g., the post-test scores for racial/ethnic minorities were 18% points lower on average than the Caucasian students). This troubling result indicates that the barriers to achievement for traditionally underserved students are substantial and persistent. We argued that this DS unit took advantage of cultural funds of knowledge (Seiler 2001) by connecting with students' personal needs in the design activity. But there is clearly more that needs to be done to understand and respond to the challenges in these settings for disadvantaged students (Seiler et al. 2001). Further instructional innovation is required to capitalize more effectively on the resources that students bring to the classroom. Our work using design-based learning in science classrooms suggests that, with further improvement, design could have a powerful effect for the most at-risk students in the most challenging schools.

Since including open-ended science activities almost always implies that less time will be spent covering the content in the traditional sense, curriculum developers need to justify the increased time and resources spent on such activities. DS curriculum units must facilitate learning not only of engineering and technology skills, but also target science content and inquiry skills to be considered a practical alternative to inquiry-based units. Our findings further those efforts. By articulating and documenting the advantage of DS units for high-needs urban populations, these findings will also serve to bolster design as a viable

alternative for helping to create systemic change in science education.

**Acknowledgments** We would like to acknowledge Anton Lawson for allowing us to use the Classroom Test of Scientific Reasoning, Kalyani Raghavan from the MARS curriculum for her help with the data collection and useful feedback on the writing, and the teachers and students who invited us into their classrooms. This work was supported by the National Science Foundation under Grant EHR-0227016. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.

### Appendix 1: Design Documentation Sheet

#### Open exploration: Discovery Documentation

By answering the following questions carefully, your team will be well prepared to continue exploring ideas important for building your alarm system.

#### *Your Team's Idea*

Write ONE idea that your team discovered during open exploration today. Example: yellow wires make things turn on more often than blue wires.

#### *Circuit Sheets*

Using the circuit sharing sheets provided, draw 1, 2, or 3 circuits that would best help you show your idea to someone else. Be careful to label what each component is and the color of the wires.

#### *Circuit Design*

Describe the most important feature(s) of the way you built your circuit(s) that allows you to test your idea.

#### *Observations*

Describe what you observed for each of the circuits that show your idea. Things you may want to include are: what turned on and what did not? How loud or bright did they get? What was the voltage reading across them? How do they compare to each other?

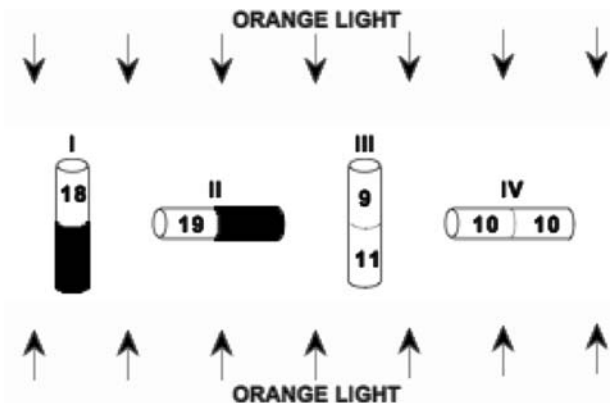
#### *Connecting to Your Alarm System*

How might this idea be important for building your team's alarm?



**Appendix 2: Sample Assessment Item From Classroom Test of Scientific Reasoning (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 5 min. The number of flies in the uncovered part of each tube is shown in the drawing.

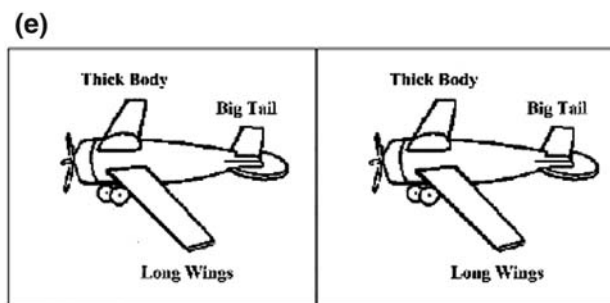
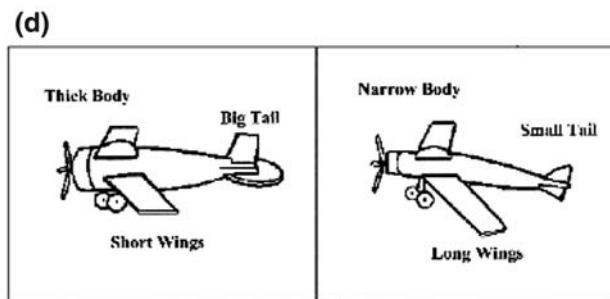
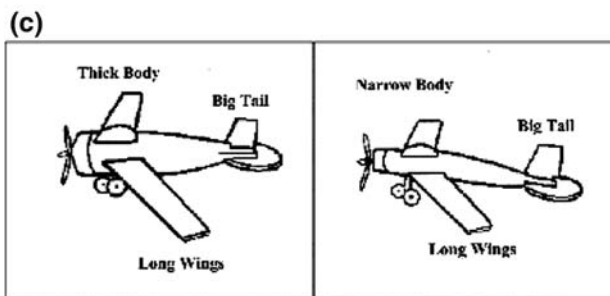
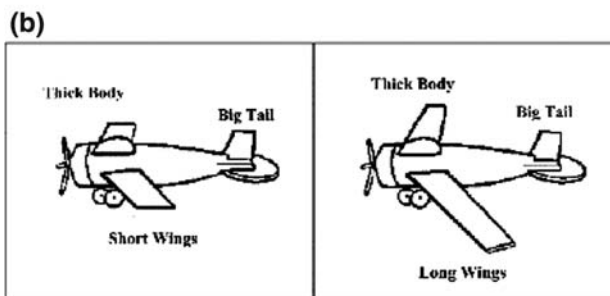
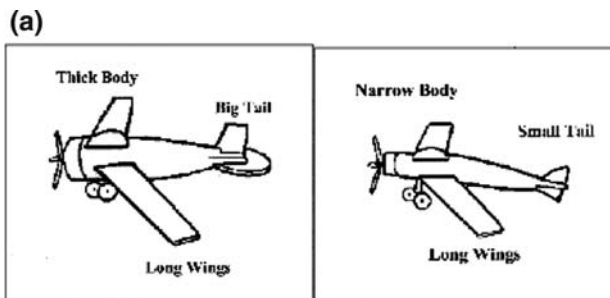


1. These data show that these flies respond to (respond means move to or away from):
  - (a) Orange light but not gravity
  - (b) Gravity but not orange light
  - (c) Both orange light and gravity
  - (d) Neither orange light nor gravity
2. Because
  - (a) Some flies are in both ends of each tube
  - (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 3: Sample Assessment Item Adapted From 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).

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



2. 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

## References

- Apedoe XS, Reynolds B, Ellefson MR, Schunn CD (2008) Bringing engineering design into high school science classrooms: the heating/cooling unit. *J Sci Educ Technol* 17:454–465. doi: [10.1007/s10956-008-9114-6](https://doi.org/10.1007/s10956-008-9114-6)
- Barron BJS, Schwartz DL, Vye NJ, Moore A, Petrosino A, Zech L, Bransford JD, The Cognition Technology Group at Vanderbilt (1998) Doing with understanding: lessons from research on problem-and project-based learning. *J Learn Sci* 7(3/4):271–311. doi: [10.1207/s15327809jls0703&4\\_2](https://doi.org/10.1207/s15327809jls0703&4_2)
- Benenson G (2001) The unrealized potential of everyday technology as a context for learning. *J Res Sci Teach* 38(7):730–745. doi: [10.1002/tea.1029](https://doi.org/10.1002/tea.1029)
- Blumenfeld PC, Soloway E, Marx RW, Krajcik JS, Guzdial M, Palincsar AS (1991) Motivating project-based learning: sustaining the doing, supporting the learning. *Educ Psychol* 26(3):369–398. doi: [10.1207/s15326985ep2603&4\\_8](https://doi.org/10.1207/s15326985ep2603&4_8)
- Bransford JD, Schwartz DL (1999) Rethinking transfer: a simple proposal with multiple implications. *Rev Res Educ* 24:61–100
- Bransford JD, Brown AL, Cocking RR (2000) How people learn: brain, mind, experience, and school. National Academy Press, Washington, DC
- Buxton CA (2005) Creating a culture of academic success in an urban science and math magnet high school. *Sci Educ* 89:392–417. doi: [10.1002/sce.20057](https://doi.org/10.1002/sce.20057)
- Cajas F (2001) The science/technology interaction: implications for science literacy. *J Res Sci Teach* 38(7):715–729. doi: [10.1002/tea.1028](https://doi.org/10.1002/tea.1028)
- Cawley JF, Parmar RS (2001) Literacy proficiency and science for students with learning disabilities. *Read Writ Q* 17:105–125. doi: [10.1080/105735601300007589](https://doi.org/10.1080/105735601300007589)
- Chinn CA, Malhotra BA (2002) Epistemologically authentic inquiry in schools: a theoretical framework for evaluating inquiry tasks. *Sci Educ* 86(2):175–218. doi: [10.1002/sce.10001](https://doi.org/10.1002/sce.10001)
- Cohen J (1992) A power primer. *Psychol Bull* 112(1):155–159. doi: [10.1037/0033-2909.112.1.155](https://doi.org/10.1037/0033-2909.112.1.155)
- Coletta VP, Phillips JA (2005) Interpreting FCI scores: normalized gain, preinstruction scores, and scientific reasoning ability. *Am J Phys* 73(12):1172–1182. doi: [10.1119/1.2117109](https://doi.org/10.1119/1.2117109)
- 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. *J Res Sci Teach* 38(7):791–820. doi: [10.1002/tea.1032](https://doi.org/10.1002/tea.1032)
- De Miranda MA (2004) The grounding of a discipline: cognition and instruction in technology education. *Int J Technol Des Educ* 14:61–77. doi: [10.1023/B:ITDE.0000007363.44114.3b](https://doi.org/10.1023/B:ITDE.0000007363.44114.3b)
- Doppelt Y, Mehalik MM, Schunn CD (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 RC, Krajcik JS, Marx RW, Mamlok-Naaman R (2004) Design-based science and student learning. *J Res Sci Teach* 41(10):1081–1110. doi: [10.1002/tea.20040](https://doi.org/10.1002/tea.20040)
- Fortus D, Krajcik JS, Dershimer RC, Marx R, Mamlok-Naaman R (2005) Design-based science and real-world problem-solving. *Int J Sci Educ* 27(7):855–879. doi: [10.1080/09500690500038165](https://doi.org/10.1080/09500690500038165)
- Germann PJ, Haskins S, Auls S (1996) Analysis of nine high school biology laboratory manuals: promoting scientific inquiry. *J Res Sci Teach* 33(5):475–499. doi: [10.1002/\(SICI\)1098-2736\(199605\)33:5<475::AID-TEA2>3.0.CO;2-O](https://doi.org/10.1002/(SICI)1098-2736(199605)33:5<475::AID-TEA2>3.0.CO;2-O)
- Gleitman L, Papafragou A (2005) Language and thought. In: Holyoak KJ, Morrison RG (eds) *The Cambridge handbook of thinking and reasoning*. Cambridge University Press, New York, pp 633–661
- Graham S, Taylor AZ, Hudley C (1998) Exploring achievement values among ethnic minority early adolescents. *J Educ Psychol* 90(4):606–620. doi: [10.1037/0022-0663.90.4.606](https://doi.org/10.1037/0022-0663.90.4.606)
- Haberman M (1991) Pedagogy of poverty versus good teaching. *Phi Delta Kappan* 73:290–294
- Hestenes D, Wells M, Swackhamer G (1992) Force concept inventory. *Phys Teach* 30:141–158. doi: [10.1119/1.2343497](https://doi.org/10.1119/1.2343497)
- Hmelo CE, Holton DL, Kolodner JL (2000) Designing to learn about complex systems. *J Learn Sci* 9(3):247–298. doi: [10.1207/S15327809JLS0903\\_2](https://doi.org/10.1207/S15327809JLS0903_2)
- Holbrook JK, Gray JT, Fasse BB, Camp PJ, Kolodner JL (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 (7th and 8th grades). Retrieved from <http://timss.bc.edu/timss1995i/TIMSSPDF/BSItems.pdf>
- Johnson MA, Lawson AE (1998) What are the relative effects of reasoning ability and prior knowledge on biology achievement in expository and inquiry classes. *J Res Sci Teach* 35(1):89–103. doi: [10.1002/\(SICI\)1098-2736\(199801\)35:1<89::AID-TEA6>3.0.CO;2-J](https://doi.org/10.1002/(SICI)1098-2736(199801)35:1<89::AID-TEA6>3.0.CO;2-J)
- Jones MG, Howe A, Rua MJ (2000) Gender differences in students' experiences, interests, and attitudes toward science and scientists. *Sci Educ* 84:180–192. doi: [10.1002/\(SICI\)1098-237X\(200003\)84:2<180::AID-SCE3>3.0.CO;2-X](https://doi.org/10.1002/(SICI)1098-237X(200003)84:2<180::AID-SCE3>3.0.CO;2-X)
- Kahle JB, Meece J, Scantlebury K (2000) Urban African-American middle school science students: does standards-based teaching make a difference? *J Res Sci Teach* 37(9):1019–1041. doi: [10.1002/1098-2736\(200011\)37:9<1019::AID-TEA9>3.0.CO;2-J](https://doi.org/10.1002/1098-2736(200011)37:9<1019::AID-TEA9>3.0.CO;2-J)
- Kanari Z, Millar R (2004) Reasoning from data: how students collect and interpret data in science investigations. *J Res Sci Teach* 41(7):748–769. doi: [10.1002/tea.20020](https://doi.org/10.1002/tea.20020)
- Klahr D, Li J (2005) Cognitive research and elementary science instruction: from the laboratory, to the classroom, and back. *J Sci Educ Technol* 14(2):217–238. doi: [10.1007/s10956-005-4423-5](https://doi.org/10.1007/s10956-005-4423-5)
- Kolodner JL, Camp PJ, Crismond D, Fasse B, Gray J, Holbrook J, Puntambekar S, Ryan M (2003a) Problem-based learning meets case-based reasoning in the middle-school science classroom: putting learning by design™ into practice. *J Learn Sci* 12(4):495–547. doi: [10.1207/S15327809JLS1204\\_2](https://doi.org/10.1207/S15327809JLS1204_2)
- Kolodner JL, Gray JT, Fasse BB (2003b) Promoting transfer through case-based reasoning: rituals and practices in learning by design™ classrooms. *Cogn Sci Q* 3(2):183–232
- Kuhn D (2007) Reasoning about multiple variables: control of variables is not the only challenge. *Sci Educ* 91(5):710–726. doi: [10.1002/sce.20214](https://doi.org/10.1002/sce.20214)
- Kuhn D, Dean D Jr (2008) Scaffolded development of inquiry skills in academically disadvantaged middle-school students. *J Psychol Sci Technol* 1(2):36–50

- Lawrenz F, Huffman D, Welch W (2001) The science achievement of various subgroups on alternative assessment formats. *Sci Educ* 85(3):279–290. doi:[10.1002/sce.1010](https://doi.org/10.1002/sce.1010)
- Lawson AE (1978) The development and validation of a classroom test of formal reasoning. *J Res Sci Teach* 15(1):11–24. doi:[10.1002/tea.3660150103](https://doi.org/10.1002/tea.3660150103)
- Lawson AE (1987) Classroom test of scientific reasoning: revised paper and pencil version. Arizona State University, Tempe
- Lee O, Deaktor RA, Hart JE, Cueva P, Enders C (2005) An instructional intervention's impact on the science and literacy achievement of culturally and linguistically diverse elementary students. *J Res Sci Teach* 42(8):857–887. doi:[10.1002/tea.20071](https://doi.org/10.1002/tea.20071)
- Leonard MJ (2005) Examining tensions in a “Design for Science” activity system: science versus engineering goals and knowledge. *J Res Teach Educ* 3:132–146 *Tidskrift for Lararutbildning och Forskning*
- Lewis T (2006) Design and inquiry: bases for an accommodation between science and technology education in the curriculum? *J Res Sci Teach* 43(3):255–281. doi:[10.1002/tea.20111](https://doi.org/10.1002/tea.20111)
- Li J, Klahr D, Siler S (2006) What lies beneath the science achievement gap: the challenges of aligning science instruction with standards and tests. *Sci Educ* 15(1):1–12. doi:[10.1007/s11191-004-4812-9](https://doi.org/10.1007/s11191-004-4812-9)
- Lundy GF (2003) School resistance in American high schools: the role of race and gender in oppositional culture theory. *Eval Res Educ* 17(1):6–30
- Lynch S, Kuipers J, Pyke C, Szesze M (2005) Examining the effects of a highly rated science curriculum unit on diverse students: results from a planning grant. *J Res Sci Teach* 42(8):912–946. doi:[10.1002/tea.20080](https://doi.org/10.1002/tea.20080)
- McCarthy CB (2005) Effects of thematic-based, hands-on science teaching versus a textbook approach for students with disabilities. *J Res Sci Teach* 42(3):245–263. doi:[10.1002/tea.20057](https://doi.org/10.1002/tea.20057)
- Mehalik MM, Doppelt Y, Schunn CD (2008) Middle-school science through design-based learning versus scripted inquiry: better overall science concept learning and equity gap reduction. *J Eng Educ* 97(1):71–85
- National Research Council (1996) National science education standards. National Academy Press, Washington, DC
- Norman O, Ault CR Jr, Bentz B, Meskimen L (2001) The black-white “achievement gap” as a perennial challenge of urban science education: a sociocultural and historical overview with implications for research and practice. *J Res Sci Teach* 38(10):1101–1114. doi:[10.1002/tea.10004](https://doi.org/10.1002/tea.10004)
- O'Neill T, Calabrese Barton A (2005) Uncovering student ownership in science learning: the making of a student created mini-documentary. *Sch Sci Math* 105(6):292–301
- Palincsar AS, Herrenkohl LR (2002) Designing collaborative learning contexts. *Theory Pract* 41(1):26–32. doi:[10.1207/s15430421tip4101\\_5](https://doi.org/10.1207/s15430421tip4101_5)
- Penner DE, Lehrer R, Schauble L (1998) From physical models to biomechanics: a design-based modeling approach. *J Learn Sci* 7(3/4):429–449. doi:[10.1207/s15327809jls0703&4\\_6](https://doi.org/10.1207/s15327809jls0703&4_6)
- Pine J, Aschbacher P, Roth E, Jones M, McPhee C, Martin C, Phelps S, Kyle T, Foley B (2006) Fifth graders' science inquiry abilities: a comparative study of students in hands-on and textbook curricula. *J Res Sci Teach* 43(5):467–484. doi:[10.1002/tea.20140](https://doi.org/10.1002/tea.20140)
- Puntambekar S, Kolodner JL (2005) Toward implementing distributed scaffolding: helping students learn science from design. *J Res Sci Teach* 42(2):185–217. doi:[10.1002/tea.20048](https://doi.org/10.1002/tea.20048)
- Raghavan K, Sartoris ML, 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
- Renner JW, Stafford DG, Coffia WJ, Kellogg DH, Weber MC (1973) An evaluation of the science curriculum improvement study. *Sch Sci Math* 73(4):291–318
- Roth W-M (2001) Learning science through technological design. *J Res Sci Teach* 38(7):768–790. doi:[10.1002/tea.1031](https://doi.org/10.1002/tea.1031)
- Roth W-M, Tobin K, Ritchie S (2001) Re/Constructing elementary science. Peter Lang Publishing, New York
- Sadler PM, Coyle HP, Schwartz M (2000) Engineering competitions in the middle school classroom: key elements in developing effective design challenges. *J Learn Sci* 9(3):299–327. doi:[10.1207/S15327809JLS0903\\_3](https://doi.org/10.1207/S15327809JLS0903_3)
- Schauble L, Klopfer LE, Raghavan K (1991) Students' transition from an engineering model to a science model of experimentation. *J Res Sci Teach* 18(9):859–882. doi:[10.1002/tea.3660280910](https://doi.org/10.1002/tea.3660280910)
- Schauble L, Glaser R, Duschl RA, Schulze S, John J (1995) Students' understanding of the objectives and procedures of experimentation in the science classroom. *J Learn Sci* 4(2):131–166. doi:[10.1207/s15327809jls0402\\_1](https://doi.org/10.1207/s15327809jls0402_1)
- Seiler G (2001) Reversing the “standard” direction: science emerging from the lives of African American students. *J Res Sci Teach* 38(9):1000–1014. doi:[10.1002/tea.1044](https://doi.org/10.1002/tea.1044)
- Seiler G, Tobin K, Sokolic J (2001) Design, technology, and science: sites for learning, resistance, and social reproduction in urban schools. *J Res Sci Teach* 38(7):746–767. doi:[10.1002/tea.1030](https://doi.org/10.1002/tea.1030)
- Sirin SR (2005) Socioeconomic status and academic achievement: a meta-analytic review of research. *Rev Educ Res* 75(3):417–453. doi:[10.3102/00346543075003417](https://doi.org/10.3102/00346543075003417)
- Toth EE, Klahr D, Chen Z (2000) Bridging research and practice: a cognitively based classroom intervention for teaching experimentation skills to elementary school children. *Cogn Instr* 18(4):423–459. doi:[10.1207/S1532690XCI1804\\_1](https://doi.org/10.1207/S1532690XCI1804_1)
- Tschirgi JE (1980) Sensible reasoning: a hypothesis about hypotheses. *Child Dev* 51(1):1–10. doi:[10.2307/1129583](https://doi.org/10.2307/1129583)
- Ulrich KT, Eppinger SD (2004) Product design and development, 3rd edn. McGraw-Hill, New York
- Waxman HC, Huang S-YL, Padron YN (1995) Investigating the pedagogy of poverty in inner-city middle level schools. *Res Middle Lev Educ* 18(2):1–22
- Wilson MR, Bertenthal MW (eds) (2006) Systems for state science assessment. The National Academies Press, Washington, DC
- Zimmerman C, Raghavan K, Sartoris ML (2003) The impact of the MARS curriculum on students' ability to coordinate theory and evidence. *Int J Sci Educ* 25(10):1247–1271. doi:[10.1080/0950069022000038303](https://doi.org/10.1080/0950069022000038303)