Research Article

Bridging the Design-Science Gap With Tools: Science Learning and Design Behaviors in a Simulated Environment for Engineering Design

Jie Chao,[1] Charles Xie,[1] Saeid Nourian,[1] Guanhua Chen,[1] Siobhan Bailey,[1] Molly H. Goldstein,[2] Senay Purzer,[2] Robin S. Adams,[2] and M. Shane Tutwiler[3]

[1]*The Concord Consortium, 25 Love Lane, Concord, Massachusetts 01742*
[2]*School of Engineering Education, Purdue University, West Lafayette, Indiana*
[3]*Harvard Graduate of School of Education, Harvard University, Cambridge, Massachusetts*

Received 15 July 2016; Accepted 12 April 2017

Abstract: Many pedagogical innovations aim to integrate engineering design and science learning. However, students frequently show little attempt or have difficulties in connecting their design projects with the underlying science. Drawing upon the Cultural-Historical Activity Theory, we argue that the design tools available in a learning environment implicitly shape knowledge development as they mediate students' design actions. To explore the roles of tools in design-science integrated learning environments, this study investigated how secondary students' tool-mediated design actions were linked with their science learning in a tool-rich design environment with minimal explicit guidance. Eighty-three ninth-grade students completed an energy-efficient home design challenge in a simulated environment for engineering design supported by rich design tools. Results showed that students substantially improved their knowledge as a result of designing with the tools. Further, their learning gains were positively associated with three types of design actions—representation, analysis, and reflection—measured by the cumulative counts of relevant computer logs. In addition, these design actions were linked with learning gains in ways that were consistent with their theoretical impacts on knowledge development. These findings suggest that, instead of being passive components in a learning environment, tools considerably shape design processes, and learning paths. As such, tools offer possibilities to help bridge the design-science gap. J Res Sci Teach 9999:XX–XX, 2017
Keywords: engineering design; science learning; precollege engineering; design tools; computer-aided design; computer simulation; computer logs; cultural-historical activity theory

In engineering design, designers practice science as they create models, test ideas, analyze data, and construct new knowledge, which is then applied to conceive, compare, and optimize design solutions (Crismond & Adams, 2012; Lewis, 2006). The nexus between design and science is the basis for pedagogical innovations that integrate engineering design and science learning. Evidence favoring this design-science integration approach has been accumulating (e.g., Crismond, 2001; Fortus, Dershimer, Krajcik, Marx, & Mamlok-Naaman, 2004;

Jie Chao, Learning Scientist, The Concord Consortium, Concord MA.
Conflict of interest: The authors of this article have no conflict of interest to declare.
Contract grant sponsor: National Science Foundation; Contract grant number: DRL-1348547.
Correspondence to: J. Chao; E-mail: jchao@concord.org
DOI 10.1002/tea.21398
Published online in Wiley Online Library (wileyonlinelibrary.com).

Hmelo, Holton, & Kolodner, 2000; Kolodner et al., 2003). However, it is also widely documented that students frequently show little attempt or have difficulties in making connections between the design challenges they are tackling and the underlying scientific concepts and processes (Carroll et al., 2010; Crismond, 2001; Mentzer, 2014; Roth, Tobin, & Ritchie, 2001). Many cross-sectional studies have shown that students spent much less effort on gathering background information and analyzing their designs than experienced designers (Lammi & Gero, 2011; Lammi et al., 2014; Mentzer, Becker, & Sutton, 2015). Even when they did, the information they gathered was often not related to science (Mentzer, 2014) and the knowledge they learned was highly contextualized and lacked connections to science concepts (Crismond, 2001). These novice behaviors are resistant to change even after substantial exposure to pre-college engineering learning opportunities (Kannengiesser, Gero, Wells, & Lammi, 2015; Lammi et al., 2014). Even in classrooms focused on science learning, students often got too absorbed in building physical models and failed to reflect upon the underlying science concepts (e.g., Carroll et al., 2010; Roth et al., 2001). In many cases, students copied peers' design ideas or searched the Internet to find existing designs to follow instead of resorting to scientific principles and practices (e.g., Roth, 1995). When asked to explain their designs, students tended to focus on the functions of their designs instead of the underlying mechanisms (Carlsen, 1998).

This "design-science gap" (Vattam & Kolodner, 2008) undermines the very promise of the design-science integration approach that is critical in precollege engineering. It is imperative for science educators to gain a deeper understanding of this issue and explore strategies to bridge the gap. Among many factors that mediate the connection between engineering design and science learning, design tools are often overlooked due to their seemingly passive status. Yet, research on design cognition has shown that design tools have significant influence on designers' thoughts (e.g., Cross, 1999; Purcell & Gero, 1998), especially for novice designers (e.g., Lemons, Carberry, Swan, Jarvin, & Rogers, 2010; Welch, 1999). The purpose of this study was to explore how design tools shape students' design behaviors as well as their science knowledge development. Specifically, we examined how students' science learning gains were associated with their design behaviors mediated by three main types of design tools—representation tools, analysis tools, and reflection tools—in a computer-based, tool-rich design environment. The findings of study offered insight on the impacts of design tools on science learning and implications for implementing design-science integration approach.

Theoretical Framework

To understand the root cause of the design-science gap and the potential roles of design tools, one needs to look beyond individual students' behaviors and reinterpret the phenomenon in a larger framework of human activities. The Cultural-Historical Activity Theory (CHAT, Engeström, Miettinen, & Punamäki, 1999) provides a powerful lens to examine this issue. CHAT views human activities as goal-oriented systems that comprise dynamic interactions among subject, object, mediating artifacts, community, rules, and division of labor. Consider a design project that challenges students to design insulation materials to minimize heat loss in a house within a given budget. The students are provided with premade mini model houses, a variety of materials such as fiberglass and cellulose, temperature sensors, heating lamps, and design notebooks. Working in small groups, the students start with some vague ideas of what good insulation materials should look like. They construct models using the given model houses and materials and test the models with the heating lamps and temperature sensors. In this design activity system (Figure 1), the individual student is considered as the *subject* of the design activity system. The initial *object* would be her or his initial vague ideas about insulation. The student transforms her or his design ideas through multiple cycles until they

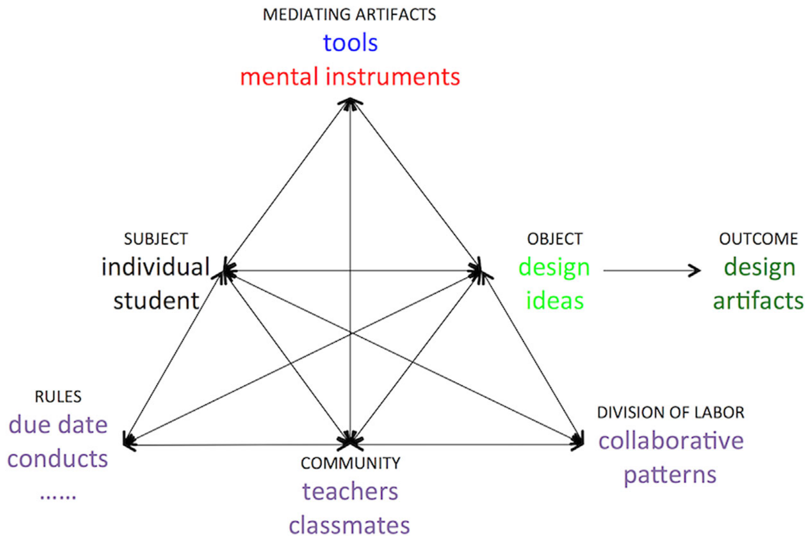


Figure 1. A design activity system of a student designer (adapted from Engeström, 1987, p.78). [Color figure can be viewed at wileyonlinelibrary.com].

stabilize as a desired *outcome* in the form of design artifacts. This *transformation* is made possible by *mediating artifacts* including *tools* such as the premade model houses and the temperature sensors and *mental instruments* such as the knowledge, skills, and strategies shared within her or his team. In the *community* of her or his design team, the students continuously negotiate their *division of labor* as well as explicit and implicit *rules* such as the pace of work, use of resources, and appropriate behaviors.

Where is science learning in this design activity system? From the CHAT perspective, learning is a by-product of the activity system orientated toward the design goal. In trying to create designs that serve a need, students must acquire, use, and adapt mediating artifacts, which include both tools and mental instruments. The development of mental instruments is equivalent to the conventional notion of learning. Situating science learning in a design activity system helps explain the design-science gap. Because science learning may not be the goal of the design activity, as long as students can achieve the desired design outcomes by other means, they would have no strong reason to learn or apply science. Students may naturally "divide labors" to search for solutions from the Internet or their classmates when these practices are allowed by the rules. They may inch toward the desired outcome only using weak methods that do not take advantage of domain knowledge, particularly the science principles involved. From the students' point of view, imitation and tinkering are legitimate, effective, and efficient ways to get the job done.

Research has identified several ways to help students connect their design projects and the underlying science. The most common strategy is to have the teacher monitor students' progresses and intervene with scientific principles specific to their design problems. Yet, this strategy is much less effective with larger class size or less experienced teachers (e.g., Bamberger, Cahill, Hagerty, Short, & Krajcik, 2010). Another common strategy is to have students collaborate and review each other's designs so that students would develop deeper understanding as they try to explain their ideas and justify their design choices (Roth, 2001). The effectiveness of this approach is contingent upon students' attitudes toward peers, communication skills, and domain knowledge. A relatively scalable and effective strategy is to divert students to inquiry activities oriented toward science learning in the middle of the design process (Kolodner et al., 2003; Schnittka & Bell, 2011). However, the drawback of this approach is that students may face different difficulties and work at different paces, so the imposed switch of focus is likely to be disruptive and misaligned with students' own goals and trajectories.

In contrast to the social mediation discussed above, tools are often perceived as passive components in learning environments. However, as indispensable actors in the activity system, tools can significantly influence students' mental development. Tools developed over time to enable people to solve certain problems, served as the "functional organs" to augment human natural capacities (Kaptelinin, 1996), and became the "carriers" of the knowledge and culture of communities (Wertsch, 1994). For example, computer-aided design (CAD) software was developed to assist designers to perform design tasks, thus its functionalities reflect the common knowledge of the design community. Tools mediate actions by implicitly specifying the *modes of operation* for the subjects (Norman, 1999; Vygotsky, 1981). The mediated actions are gradually *internalized* into mental processes and structures that can be used independently from the presence of artifacts. For example, the shape of a scissor suggests the action of cutting materials with its two blades. With repeated use of scissors, people become able to mentally simulate the process of cutting-with-a-scissor and develop the embodied schema of "cutting" (Johnson, 1987). Therefore, tools are not only instrumental for designers but also influential on their mental development through mediating their design actions.

In the simplest design activity system, there are at least three interconnected design actions mediated by tools: *representation*, *analysis*, and *reflection* (Figure 2). In this simple design action cycle, students first transform their design ideas into design artifacts using *representation tools*. Then they transform the design artifacts into functions and behaviors of the designs using *analysis tools*, which support model testing and in-depth inquiry. Their design experience comprises the design representation and analysis actions as well as the associated feedback on the structures, functions, and behaviors of the designs.

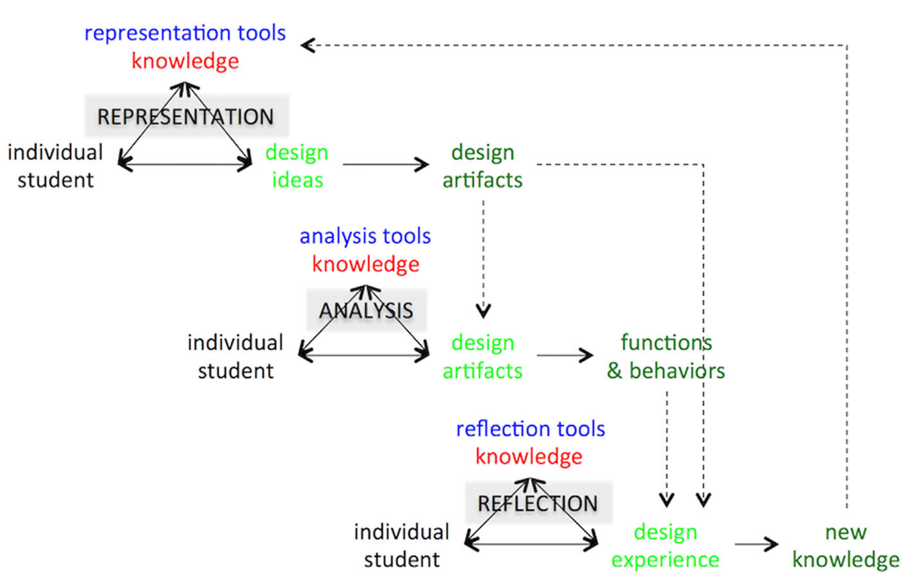


Figure 2. A cycle of design representation, analysis, and reflection actions. [Color figure can be viewed at wileyonlinelibrary.com].

Students further transform the design experience into new knowledge using the *reflection tools*. The new domain knowledge will then mediate a new cycle of design actions by spurring new design ideas and analysis techniques, enabling new interpretations of design feedback, and even promoting new ways to reflect upon the design experience.

In this design activity system, the designer's behavior and knowledge state coevolve through reciprocal influence. Tools shape this coevolution in several ways. First, they can elicit or suppress certain aspects of students' current domain knowledge, constraining or expanding the design space, and highlighting the underlying science concepts. The literature of design cognition reveals that physical or virtual representations help designers to notice emergent features (Cross, 1999; Purcell & Gero, 1998) and augment their imagination (Verstijnen, van Leeuwen, Goldschmidt, Hamel, & Hennessey, 1998). Student designers were found to be even more reliant on representations to understand design challenges, generate ideas, and iterate solutions (Cardella, Atman, & Adams, 2006; Lemons et al., 2010; Welch, 1999). Research in science learning repeatedly showed that the presence of artifacts could funnel students' reasoning onto productive paths (Clement, Brown, & Zietsman, 1989; Roth, 2001; Schoultz, Säljö, & Wyndhamn, 2001).

Tools also provide valuable feedback on the structure, functions, and behaviors of designs. Multiple researchers have pointed out that students' oversight on scientific principles may be attributed to the lack of feedback on the functionalities of the designs (Barnett, 2005; Hmelo et al., 2000). This pattern can be changed when students have access to test and analysis tools that can generate rich feedback on how certain design features function. For instance, Apedoe and Schunn (2013) observed that students developed deeper understanding of quake-resistant structures through iteratively testing their models.

In addition, tools can mimic social mediation that nudges students to shift focus to conceptual understanding. This function manifests most strongly in tools designed to promote deep reflection. Experienced designers constantly engage in a reflective conversation with the design problem, frequently transitioning between problem scoping activities (e.g., gathering information and determining constraints) and problem solving activities (e.g., modeling and evaluation, Adams, Turns, & Atman, 2003). For student designers, reflection is critical to learning. In many noteworthy design-science integration approaches, students are asked to keep design journals with structured prompts that mimic the practices of informed designers (e.g., Chiu et al., 2013; Puntambekar & Kolodner, 2005). In a case study with elementary students, Wendell and Lee (2010) demonstrated a close connection between reflective note-taking behaviors and conceptual development in materials science.

Literature Review

Despite the importance of tools in design, their roles in design-science integrated learning environments are not well understood due to two typical sources of confounding. First, in some cases, the learning environments used in the research lacked mission-critical tools. Many earlier studies reported that students failed to attend to the mechanisms underlying their designs due to the lack of construction and testing tools (e.g., Barnett, 2005; Hmelo et al., 2000). For instance, in their reflection on an artificial lungs design challenge for sixth-graders, Hmelo et al. (2000) pointed out:

> "The lung challenge afforded generating questions and coming up with ideas, but it did not afford coming up with solutions that could be tested......As a result, students had a hard time getting to a point where it made sense to grapple with the details of implementation. Yet it is in implementation that design challenges have their most powerful affordances for

> learning... ...The best design challenges for promoting learning are those that afford construction, testing, timely and authentic feedback, and revision." (Hmelo et al., 2000, p.287)

In more recent studies, construction and testing tools have been recognized as important elements in design-based learning environments and commonly provided in the reported studies (Fortus, Krajcik, Dershimer, Marx, & Mamlok-Naaman, 2005; Kolodner et al., 2003; Wendell & Lee, 2010). Students were provided with adequate materials and construction tools to create models as well as a variety of devices and sensors to test their models. However, these materials and tools cannot reveal the science concepts and processes that explain design success or failure. As a result, student learning often remains implicit and ambiguous, not immediately connected with science.

An ideal design environment to support design-science integration would provide a collection of tools that can support the entire process of engineering design. Such a tool-rich design environment is necessary to study tool-mediated design and learning. These tools include but not limited to: (i) adequate and carefully curated set of physical and/or virtual materials to allow students to fully explore the design space and flexibly represent different design ideas; (ii) science visualizations and simulations, ideally overlaid on students' design artifacts, to help them understand the underlying science concepts and consequences of their design choices; (iii) space and guidance for students to reflect upon their intermediate design solutions and design processes. These design tools, by implicitly shaping students' design behaviors, could infuse the cultural practices of expert designers into classroom learning.

Second, as aforementioned, the rules, community, and division of labor in a design activity system mediate students' design behaviors and knowledge development. In many studies aimed to evaluate entire curriculum models, students followed prescribed learning cycles, which were often directed by teachers (e.g., Fortus et al., 2005; Kolodner et al., 2003). As a result, even when the research was focused on specific tools (Puntambekar & Kolodner, 2005; Wendell & Lee, 2010), the effects of the tools were likely overshadowed or confounded by the strong social mediation present in the learning environments.

To highlight the roles of tools, a tool-rich design environment with *minimal explicit guidance* would be strongly preferred. There have been a few studies that reflected such a context, yet they either lacked assessment of science learning or did not formally examine the relationship between tool-mediated design behaviors and science learning. For example, Sun, Wang, and Chang (2013) investigated the relationships among students' thinking style, design strategy, and tool usage during a bridge design task using the West Point Bridge Designer construction simulation software. As a tool-rich design environment, the software allowed students to represent and test their designs, analyze and diagnose issues, and document and review design history. While the study reported some interesting correlations among certain thinking styles, design strategies, and tool usage, students' science learning was neither measured nor inferred from their design behaviors.

In a study of student thinking and learning in a robotics design environment, Sullivan (2008) found that students' systems understanding significantly improved after the design activity. Sullivan attributed the learning outcomes partially to the tool-rich design environment, which provided immediate feedback to motivate students to iteratively improve their design and knowledge. However, no further investigation was conducted to substantiate this attribution.

In a recent study with fifth-grade students challenged to solve physics problems (e.g., accurately measure the velocity of moving objects), Kim, Suh, and Song (2015) reported student science learning evident in their design behaviors, classroom conversations, and design artifacts. The learning outcomes were attributed to the abundant design tools including physical materials for prototyping and testing, and particularly mobile phones that allowed the students to use a variety of apps to gather information from the Internet, perform measurement, analyze data, and document design progress. However, similar to Sullivan's (2008) study, this study also lacked in-depth investigation into the relationships between tool-mediated design behaviors and science learning. The present study aimed to close this knowledge gap by answering three research questions as follow:

1. To what extent do secondary students learn science concepts through completing an engineering design project in a tool-rich design environment with minimal explicit guidance?
2. How are students' learning gains associated with three types of tool-mediated design actions—representation, analysis, and reflection?
3. What are the unique contributions of these three types of design actions in their collective association with learning gains?

Method

Research Design

This study aimed to investigate how secondary students' design actions relate to their science knowledge development in the design topic in a tool-rich design environment with minimal explicit guidance. This research focus required a combination of features of the controlled observation method and naturalistic observation method. That is, observations were made in a natural setting so that participants would behave in normal modes, but some aspects of the setting were manipulated to reduce noise and highlight the phenomenon of interest. To maintain ecological validity (Johnson & Christensen, 2008), we conducted this study in an authentic classroom setting. To reduce noise from other sources of mediation, students were provided with only an overview of the software and an introduction of the design challenge and then they were asked to work individually as much as they could. Note that these instructional methods (i.e., minimal explicit guidance and individual work) were chosen not because they represent the best classroom practices but because they permit clearer observations of the impacts of design tools on students' knowledge development. Alternative instructional methods such as explicit learning scaffolds and collaborative learning may be more desirable or even necessary in real classroom settings.

Participants and Context

This study was conducted in a suburban high school in northeastern America. Table 1 shows the school's demographics. Participants were 111 9th-grade students enrolled in five physical sciences honors class sections taught by a male teacher. The teacher had over 17 years experience teaching physical science and 5 years experience implementing engineering design projects. The following students were eliminated from the sample for various reasons: one student withdrew from the study in the middle of the project; one student did not complete the posttest; two students' user logs were lost due to hardware issues; 24 students had absent days during the project (their absence considerably compromised the fidelity of implementation). After eliminating these students, the resulting sample was 83 (40 girls and 43 boys).

Simulated Environments for Engineering Design Supported by Rich Tools

To meet the tool-rich criterion previously described, we conducted this study using a *Simulated Environment for Engineering Design* (SEED). As its name implies, SEEDs attempt to simulate the

Table 1
Demographics of the participating school

Gender	Plans of Graduates
Male 47%	4-year college 87%
Female 53%	2-year college 8%
Race	Work 3%
African American 4.2%	Other 2%
Asian 10.7%	Mobility
Hispanic 4.6%	First language not English 13.5%
Native American 0.2%	English language learner 1.1%
White 76.7%	Students with disabilities 10.5%
Native Hawaiian/Pacific Islander 0.2%	High Needs 18.7%
Multi-Race/Non-Hispanic 3.4%	Economically Disadvantaged 8.2%

entire engineering design process by integrating tools for representation, analysis, and reflection in “one stop shop.” Three characteristics of SEEDs are necessary for helping students to learn and apply science concepts in design activities. First, the tools in SEEDs must reflect authentic practices in the professional science and engineering communities. Second, the tools must be easy for science and engineering novices to pick up and use properly. Third, the tools can guide students to approach design goals by providing feedback on major design processes. With these characteristics, the tools may infuse the knowledge and practices of professional communities to classrooms by mediating students’ design actions, and through repeated use of these tools students may eventually internalize the knowledge and practices as their own mental instruments.

Energy3D

The specific SEED used in this study was Energy3D (Figure 3), a simulated environment for designing, analyzing, and constructing green buildings that utilize renewable energy (Xie & Nourian, n.d.). In Energy3D, users can quickly sketch up a realistic-looking building and evaluate its energy performance for any given day and location. Based on computational physics, Energy3D can rapidly generate time graphs, heat flux, and heat maps for in-depth analyses. An embedded notepad is available for keeping design journals. Prompts can be preloaded in the notepad to guide students through reflection. As such, Energy3D embodies a SEED by integrating representation tools, analysis tools, and reflection tools. Several recent studies have shown that students spontaneously learned and applied science during design in Energy3D. Goldstein, Purzer, Zielinski, and Adams (2015) found increasing number of students citing science concepts and analysis results in their reflective notes as they moved through design iterations. In a related case study, Purzer, Goldstein, Adams, Xie, and Nourian (2015) discovered that deep learning of science concepts was motivated by students’ desire to balance design benefits and trade-offs, which were design issues made visible and assessable by Energy3D’s capability of simulating consequences of any design choices.

Energy-Plus Home Design Challenge

The participants in this study completed an Energy-Plus Home design project (Figure 4). The project challenged students to build homes that, over the course of a year, produce more renewable energy than the energy consumed. The design challenge specified three house styles, budget, and various structural criteria. Students were also advised to complete their designs by following an engineering design cycle: starting with understanding design specifications and planning,

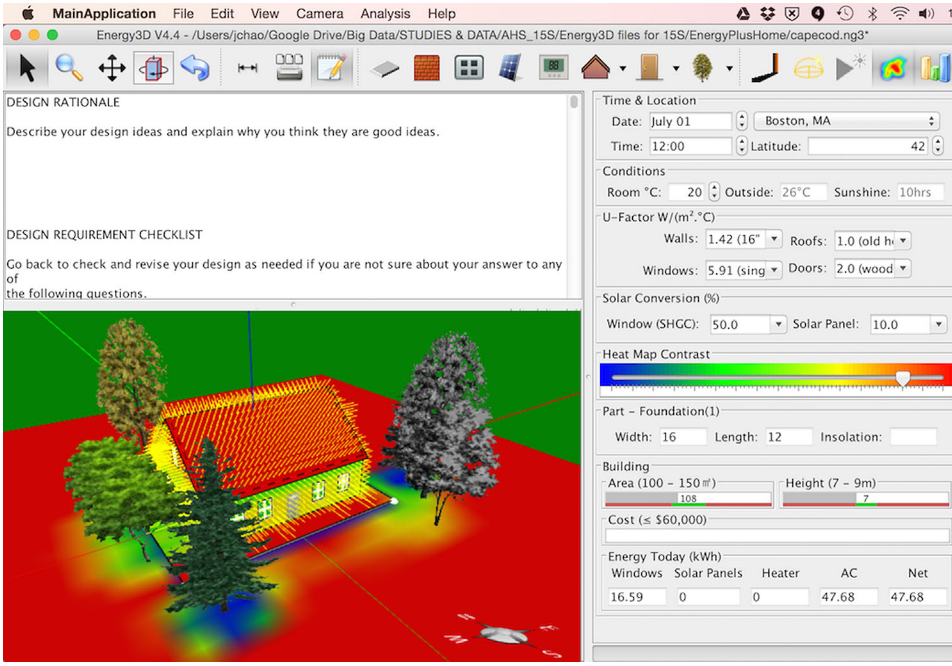


Figure 3. Energy3D, a simulated environment for designing, analyzing, and constructing green buildings that utilize renewable energy. [Color figure can be viewed at wileyonlinelibrary.com].

proceeding iteratively through construction, testing, evaluation, and revision, and finally reporting design choices and justifications. Learning scaffolds were implicitly provided through the design tools available in the Energy3D. First, a set of simulation tools was available to allow students to analyze the physical environment (i.e., the Sun path and shadow patterns), thermal processes (i.e., the heat flux through building envelop), sources of cost (i.e., cost breakdown by building components), and energy performance of their designs (i.e., annual energy consumption and production). Second, as part of the design project, students were required to complete design journals with prompts for (i) documenting design ideas and rationales; (ii) checking through all design specifications; (iii) recording and reflecting upon each iteration; and (iv) summarizing lessons learned from the project.

The design challenge targeted the domain of green building science, a multi-disciplinary field that integrates physics, earth sciences, geometry, and systems thinking. This domain was new to the students as it was rarely formally taught. However, instructions and assessments in physics and earth sciences in the school district sometimes use energy-efficient buildings as contexts to explore science concepts such as heat, temperature, and solar energy. Also, as home energy efficiency has been greatly promoted in the local society, students may have absorbed information from the media and other sources outside school. Thus, the domain was relatively accessible to high school students.

Green Building Science Test

The green building science test consisted of 19 two-tier items asking students to make choices among design alternatives for given situations and explain their choices (see appendix 1). The research team developed the test to measure student learning as a result of completing the design

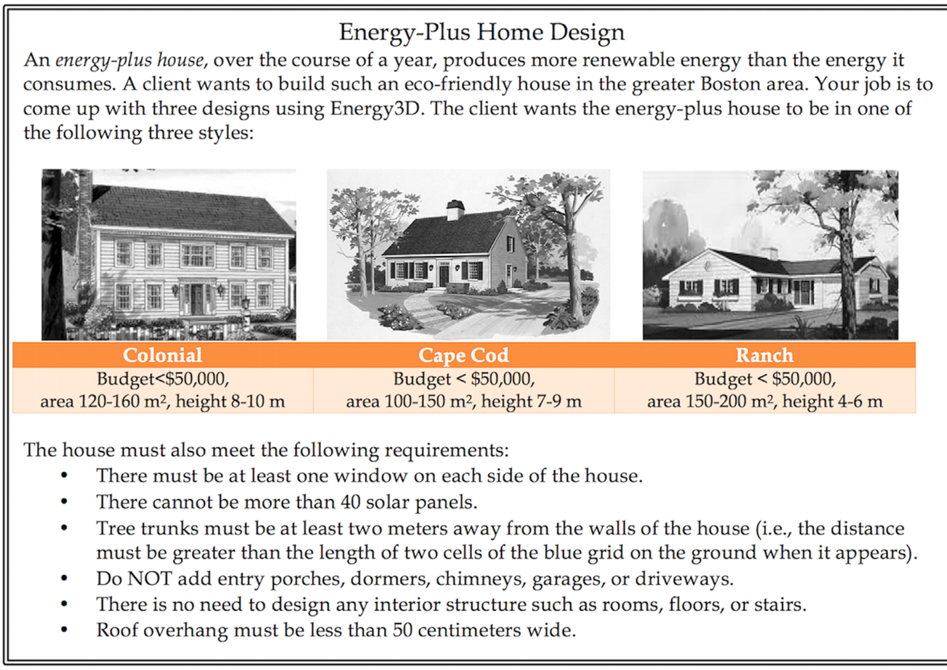


Figure 4. Energy-Plus Home design challenge. [Color figure can be viewed at wileyonlinelibrary.com].

challenge. Evaluating design alternatives is an ecologically valid assessment task as it is a common practice among engineering design professionals. It is a cognitively complex task as it requires coordinated use of knowledge of multiple types and from multiple domains. To ensure content validity, the assessment items were drawn from green building science textbooks (e.g., Hens, 2011; Montoya, 2010). To ensure instructional sensitivity, the items were further selected based on the learning opportunities offered in the Energy-Plus Home design challenge in the Energy3D design environment. The test items were reviewed by a panel of experts including green building science experts, engineering design professors, learning scientists, and high school science teachers to ensure their content validity and developmental appropriateness.

As shown in Figure 5, each item covered a subset of concepts from the four target domains (sun path & insolation, heat transfer, spatial, and geometric thinking, and thermodynamic representations). For example, one item asked students to choose among four types of trees with varying heights (i.e., tall or short) and biological characteristics (i.e., evergreen or deciduous) to improve the energy efficiency of a house. The best choice was tall deciduous trees and the proper explanation involved allowing exposure to the sun in the winter, creating shade in the summer, and increasing covered area to maximize effect.

Students' responses to the green building science test were scored using a 5-level scoring rubric developed to differentiate different levels of scientific sophistication in students' design justifications. Responses were first analyzed to identify and categorize discrete ideas as (i) normative, which is both scientifically correct and contributes to the ideal response; (ii) alternative, which is incorrect as judged by scientific norms or does not contribute to the ideal response; and (iii) irrelevant ideas including vague statements, violation of conditions, misunderstanding of the questions, etc. Then, scores (0–4) were assigned to the responses based

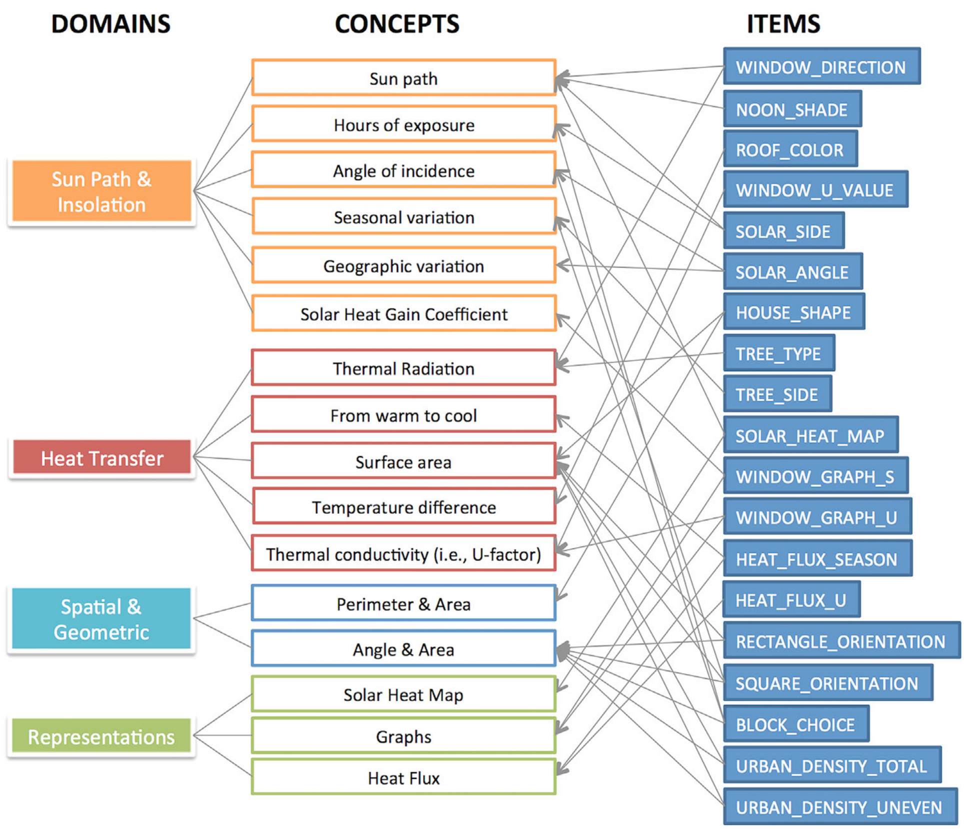


Figure 5. Concepts underlying the green building science test items. [Color figure can be viewed at wileyonlinelibrary.com].

on the number of normative, alternative, and irrelevant ideas and connections among them (Table 2).

The first author and two research assistants first independently scored 20% of students' responses to establish inter-rater reliability. Intraclass correlation coefficients ranged from 0.94 to 1.00 for different test items, indicating substantial inter-rater reliability (Koch, 1982). Then, the first author completed the scoring of the remaining responses. The test also demonstrated good internal consistency, as Cronbach's alpha was 0.82 for the pretest data and 0.83 for the posttest data.

Energy3D User Logs and Design Operation Measures

A powerful feature of Energy3D is its capability of logging all actions, notes, and snapshots unobtrusively in the background. These learner data can be used to reconstruct the entire design process with all the important details restored for analysis. A few studies have demonstrated the usefulness of Energy3D user logs to measure students' level of engagement, discern gender differences in design behaviors, identify design iterations, detect impacts of instructional interventions (Xie, Zhang, Nourian, Pallant, & Bailey, 2014), and reflect high-level design behaviors such as idea fluency (Goldstein, Purzer, Mejia, et al. 2015).

Table 2
Green building science test scoring rubrics

Level	Response Description	Example Response
Level 4	Three or more connected normative ideas without any alternative idea	Two dogwood trees would be the best trees to plant in this situation because the tree sheds leaves, so the trees will block sun going into the windows in the summer (1). And in the winter the leaves will shed so more light will be able to come through the windows (2) and heat the house. Also the trees are tall enough to cover many windows (3), but not tall enough to block the solar panels from getting sun energy, either in the summer or in the winter (4).
Level 3	Two connected normative ideas with no more than one alternative idea	They will provide shade during the summer (1), and not block the sun in the winter (2).
Level 2	One normative idea with no more than two alternative ideas	The deciduous trees would shade the windows in the summer (1) lowering AC cost.
Level 1	Using relevant but alternative ideas	Because the tree this height could collect energy more easily (x) and if they shed their leaves, they could survive a New England winter (x).
Level 0	No answer; irrelevant ideas	Seems most logical.

Using user logs to measure high-level cognitive processes require some theoretical and empirical considerations. Leont'ev (1978), one of the CHAT pioneers, proposed a three-level analytic scheme for analyzing activity systems. In this widely used scheme, activity is the high-level behavior driven by an overall motive. Activity is constituted by multiple actions, each driven by a specific goal. Action is further constituted by multiple low-level operations directly mediated by tools. In a design activity, as illustrated in Figure 6, the overall motive is to create designs that meet all the design specifications. This overall objective can be divided into multiple goals, which orient specific design actions such as *representing ideas*, *conducting experiments*, and *reflecting on process* described in the Informed Design Matrix (Crismond & Adams, 2012). Each action requires a series of operations using design tools. For instance, to conduct an experiment students need to manipulate some aspects of their designs and perform analysis using a certain simulation tool. Note that each operation with a tool has two dimensions, one is the *conventional use* developed through the history of the community and shared by community members, and the other is the *individual use* unique to individual subject and reflective of idiosyncrasy and creativity (Wertsch, 1994). Thus, when inferring mental processes based on user logs, we considered both the conventional use and variation in individual use among students. For example, Energy3D allows students to analyze the energy consumption and production of their designs over 12 months. The analysis results are presented as a graph with rich information about how the building subsystems (e.g., air conditioning, heating, rooftop solar system) perform in different seasons. Students may use this information to learn about the building system, to test different design choices, or to optimize certain parameter values, which may have greater or less connection with science learning. Below we discuss three types of operations logged by Energy3D and how they can be used to indicate higher-level design actions.

The first type of operation is *Model Manipulation,* which includes adding, removing, editing, moving, resizing, and rotating essential building structures (i.e., wall, window, door, roof, tree, and solar panel, and whole building)[1] and changing material properties (i.e., U-factors and colors

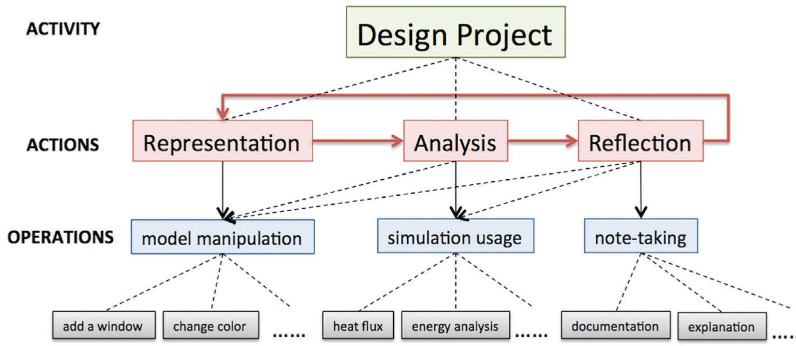


Figure 6. Three levels of a design activity system in Energy3D. [Color figure can be viewed at wileyonlinelibrary.com].

of building materials, solar heat gain coefficients (SHGC) of windows, and efficiency of solar panels). Although multiple design actions may involve model manipulation operations, *represent ideas* (Crismond & Adams, 2012) is the main design action that requires model manipulation operations. This linkage was evident in our previous finding that the amount of model manipulation logs was associated with the level of idea fluency measured by researchers' ratings of students' design replays (Goldstein, Purzer, Mejia, et al., 2015). Model manipulation operations may directly or indirectly contribute to science learning. For instance, a student may add a big window to test whether it can save or lose energy—an action that directly leads to science learning. Yet, the student may do so to express her or his aesthetic preference—an action not intended for science learning. But this action may still indirectly lead to learning when the student conducts energy analysis at some point and discovers the unexpected energy cost of the big window. However, if not concerned with energy efficiency at all, the student would not learn the underlying science. Therefore, as long as the student's high-level motive is consistent with the design specifications, her model manipulation operations will either directly or indirectly contribute to science learning.

The second type of operation is *Simulation Usage,* which includes analysis of annual or daily energy performance (Energy Annual Analysis, Compute Energy), as well as visualizations of the sun path (Show Heliodon and Animate Sun), shadow (Show Shadow), heat map (Compute Energy), and heat flux (Compute Energy). *Analysis actions* intended to build knowledge and conduct experiments (Crismond & Adams, 2012) require simulation usage. The high-level motive behind simulation usage is relatively clear—they are intended to improve the energy efficiency of the building under design. However, the same operation may have different underlying goals that reflect varying levels of expertise. For instance, students may use the annual energy analysis tool to conduct a confounded experiment or a valid experiment. When processing the analysis results, they may only read the aggregated energy performance or they may carefully examine the seasonal patterns of energy consumption and production. These nuanced operations, some of which are not captured by user logs, would account for some variance in learning. Despite the variation in individual use of the simulation tools, it is reasonable to expect that more simulation usage indicate more interactions with the energy functions of various building parts and the building's surroundings.

The third type of operation is *Note-Taking,* specifically the insertion and deletion of characters in the built-in notepad. There are four sets of prompts preloaded in the notepad: (i) describe initial design ideas and rationales; (ii) a checklist of design specifications;

(iii) a series of templates for documenting the design choices made, energy analysis results, and lessons learned in each iteration; (iv) questions about design features that promote or undermine the energy efficiency of buildings. Thus, students' notes generally fall into these four categories. Energy3D generates a record for each second while students are entering notes, so the cumulative count of note logs is a measure of the time spent on taking notes. It is reasonable to assume that more reflective students would spend more time on writing notes and generate more words than less reflective ones. Our previous studies (Goldstein, Purzer, Adams, & Xie, 2015; Goldstein, Purzer, Zielinski, et al., 2015) showed that the level of reflectivity scored by human raters was generally linked to the amount of notes students wrote. However, there are several sources of variation that need to be considered. Students may type at different speeds, resulting in different amount of notes within the same amount of time. Students may take notes with different writing styles—some write complete sentences and some use short phrases, so the same amount of notes may convey different amount of ideas. Also, the four categories of notes require varying levels of reflection, so how students allocate effort in different categories will moderate the link between the amount of notes and the amount of reflection. For example, if most of the notes are in the specification checklist category, which requires much less reflection than the other three, then the amount of notes can barely indicate how much the student has reflected.

Data Collection

Students took the green building science test the day before the design activity and the same test the day following. The test was delivered online through SurveyMonkey survey service. The Energy-Plus Home Design curriculum was implemented during a 9-day period (50-min regular periods for 8 days and an 80-min long period for 1 day, 8 hr in total). On the first day, two researchers demonstrated the Energy3D interface and students practiced how to construct buildings, use embedded simulations, and perform energy analysis. From the second to the 7th day, students created three designs for the design challenge. They were asked to complete their designs individually but allowed to communicate with each other and ask the teacher for help. On the 8th day, students completed a design modification task, in which they had to modify a house with certain structural constraints to improve its energy performance. On the last day, each student presented his or her best design to the class. Each student was given a USB flash drive preloaded with the Energy3D software, three design templates for the three house styles, and a comprehensive user's guide on the software interface and the green building science. Energy3D captured students' operation details in between an opening and a closing of the software in JSON format. All JSON files were saved in a log folder in students' flash drives.

Data Analysis

Students' responses to the green building science test were scored using the 5-level rubric previously described. Then, normalized gain scores (Hake, 1998) were calculated using the pretest and posttest scores. The log files were read into a spreadsheet and the cumulative counts of model manipulation (Manipulation), simulation usage (Simulation), and note-taking (Note) operations were calculated for each student.

As gender and social context (defined as class section in the present study) have known effects on learning behaviors (Jones et al., 2000; Jordan & McDaniel, 2014; Teng, Cai, & Yu, 2014), a two-way ANOVA was performed to detect their possible impacts on the normalized gain scores and three design operation measures. Results indicated no statistically significant effects by gender and class section. The only exception was that girls spent more time taking notes than boys.

Given the great similarity across gender subgroups and class sections in the normalized gain scores and design operation measures, a regression analysis was performed for the whole sample to preserve statistical power.

Although the three types of operations can be separately linked to certain design actions, they are intrinsically connected to each other and share the same overall motive. Thus, we used the multiple linear regression model to test whether they collectively predict science learning gain. An *a priori* power analysis indicated that 76 participants were needed to have 80% power for detecting a medium sized effect when employing the conventional .05 criterion of statistical significance. The sample size of this study was 83, which had sufficient statistical power for the planned multiple linear regression analysis.

Results

Descriptive Analysis of Science Learning Gain and Design Operation Measures

A paired-samples t-test indicated that students' performance on the green building science test significantly improved from pretest ($M = 24.82$, $SD = 8.61$) to posttest ($M = 37.81$, $SD = 8.66$), $t = 15.08$, 1-tailed p-level $<.001$, with a large effect size, Cohen's $d = 1.66$. Raw gain scores ($M = 12.99$, $SD = 7.85$) were calculated by subtracting the pretest scores from the posttest scores. However, Pearson's correlation analysis indicated a strong negative correlation between the raw gain scores and the pretest scores, $r = -.449$, p-value $<.001$, suggesting that students with high pretest scores were limited with a smaller possible range of learning gain than those with low pretest scores. To mitigate this limitation, normalized gain scores (nGain; Hake, 1998) were calculated using the raw gain scores divided by possible gain scores. The raw gain scores were obtained by subtracting the pretest scores from posttest scores. The possible gain scores were obtained by subtracting the pretest scores from the maximum possible scores. The test used in this study had 19 items. The scores for each item ranged from 0 to 4. Thus, the maximum possible score was 76.

$$\text{normalized gain score} = \frac{\text{raw gain}}{\text{possible gain}} = \frac{\text{posttest score} - \text{pretest score}}{\text{maximum possible score} - \text{pretest score}}$$

Table 3 shows descriptive statistics of the normalized gain scores and the three design operation measures (i.e., Manipulation, Simulation, and Note). The distributions of the normalized gain scores and Manipulation were approximately symmetric ($-.50 <$ Skewness $<.50$). The distribution of Simulation had a high positive skew (Skewness $=> 1.00$), and the distribution of Note had a moderate positive skew ($.50 <$ Skewnes <1.00). Shapiro-Wilk tests of normality confirmed that normality could be assumed for normalized gain scores and Manipulation but not for Simulation and Note. Normalized gain scores had two mild outliers on the lower end. Manipulation had one mild outlier and Simulation had three mild outliers on the higher end.

As shown in Figure 7, the majority of model manipulation operations were on solar panel arrangement and window structure. Wall, tree, and roof also took up substantial percentages. Small fractions of operations were on door, whole building, and four material properties (i.e., color, U-factor, solar panel efficiency, and window SHGC).

As for simulation usage (Figure 8), Energy Annual Analysis took up over half of the operations and the other four analysis tools—Animate Sun, Show Shadow, Show Heliodon, and Compute Energy—had similar shares.

Table 3
Descriptive statistics and Shapiro-Wilk normality tests of the normalized gain score and design operation measures

Statistics	nGain	Manipulation	Simulation	Note
Mean	.25	1,163	124	834
S.D.	.14	296	57	437
C.V.	.55	.25	.46	.52
Minimum	−.15	520	39	15
Maximum	.58	2,012	294	1,974
Range	.73	1,492	255	1,959
Median	.25	1,153	115	723
Q1	.16	982	86	501
Q3	.34	1,320	150	1,140
IQR	.18	339	64	639
Skewness	−.27	.41	1.17	.66
Kurtosis	3.52	3.20	4.30	2.73
Shapiro-Wilk W	.986	.982	.908	.952
Shapiro-Wilk p-level	.511	.295	<.001	.004

S.D. stands for standard deviation; C.V. stands for coefficient of variation.

Relationships Between Science Learning Gain and Design Operation Measures

Linear regression analyses were performed to examine the relationships between science learning gain and the three design operation measures. Assumptions for linear regression analysis were examined (see appendix 2 for regression diagnoses). All assumptions were met except that the relationship between Simulation and normalized gain scores appeared to be nonlinear, which

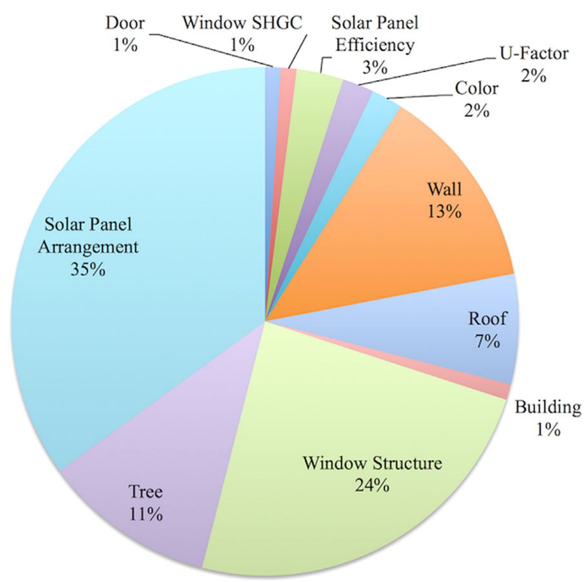


Figure 7. Average frequency distribution of different types of model manipulation. [Color figure can be viewed at wileyonlinelibrary.com].

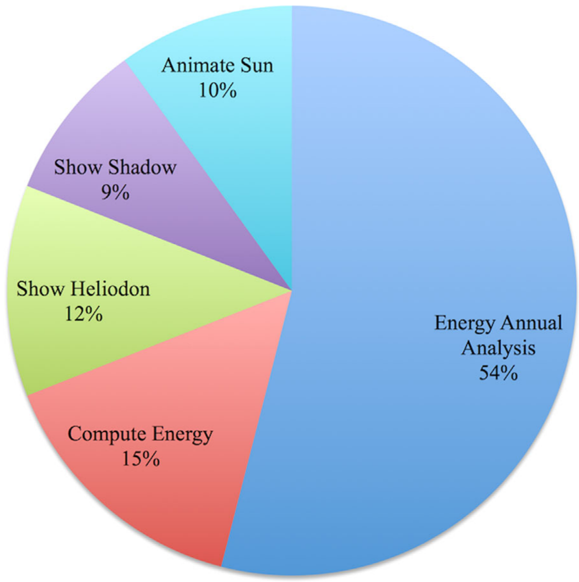


Figure 8. Average frequency distribution of different types of simulation usage. [Color figure can be viewed at wileyonlinelibrary.com].

might be caused by the highly skewed distribution of Simulation. Thus, Simulation was transformed using base-2 logarithm. The base-2 log-transformed Simulation (log2(Simulation)) appeared to be linearly associated with the normalized gain scores.

Pearson's correlation analysis (Table 4) indicated that the normalized gain scores had moderate positive correlations with Manipulation and Simulation but no significant correlation

Table 4
Pearson's correlations among normalized gain score and design operation measures

	nGain	Manipulation	log2(Simulation)	Note
nGain	—			
Manipulation	.25 (.024)	—		
log2(Simulation)	.27 (.012)	.27 (.013)	—	
Note	.18 (.100)	−.08 (.478)	−.08 (.478)	—

Table 5
Comparison of three regression models

	Results of Regression Models					Comparison of Models			
Models	R^2	*df*	*F*	*p*-value	f^2	R^2-change	*df*	*F*	*p*-value
M1: M	.061	1.81	5.307	.024	.065				
M2: M+S	.109	2.80	4.835	.010	.122	.048 (M1→M2)	1.80	4.155	.045
M3: M+S+N	.154	3.79	4.805	.004	.182	.045 (M2→M3)	1.79	4.340	.040

M, manipulation; S, log2(Simulation); N, Note.

Table 6
Association between the normalized gain score and the three design operation measures (results from multiple linear regression analysis)

Variable	*B*	SE	β	*t*-value	*p*-level
Manipulation	9.404×10^{-5}	5.039×10^{-5}	.201	1.866	.066
Log2(Simulation)	5.125×10^{-2}	2.328×10^{-2}	.237	2.202	.031
Note	6.852×10^{-5}	3.289×10^{-5}	.217	2.083	.041

with Note. Manipulation and Simulation were positively correlated, however, neither of them was correlated with Note.

Three regression models were specified and compared to examine the unique contributions of Manipulation, Simulation, and Note to the prediction of nGain. As shown in Table 5, Manipulation alone explained 6.1% of the variance in nGain. Manipulation and Simulation together explained 10.9% of the variance, which was significantly higher than Manipulation alone. The three measures together explained 15.4% of the variance, which was significantly higher than the model without Note. These results indicated that all three design operation measures had unique contributions in the prediction of the normalized gain scores.

Based on the third model, the three design operation measures together significantly predicted the normalized gain scores with a medium effect size, $R^2 = .154$, $F(3, 79) = 4.805$, $p = .004$, $f^2 = .182$. As shown in Table 6, student's normalized gain score would increase 9.404×10^{-5} for each model manipulation, 5.125×10^{-2} for each doubling in simulation usage, and 6.852×10^{-5} for each second spent on taking notes. Manipulation was only marginally significant ($p = .066$) in the presence of the other two measures, even though it alone was significantly predictive of nGain. This might be attributed to the moderate correlation between Manipulation and Simulation (although multicollinearity was not a concern based on the variance inflation factors of the three predictors) or the loss of degree of freedom in multiple linear regression. Both the base-2 log-

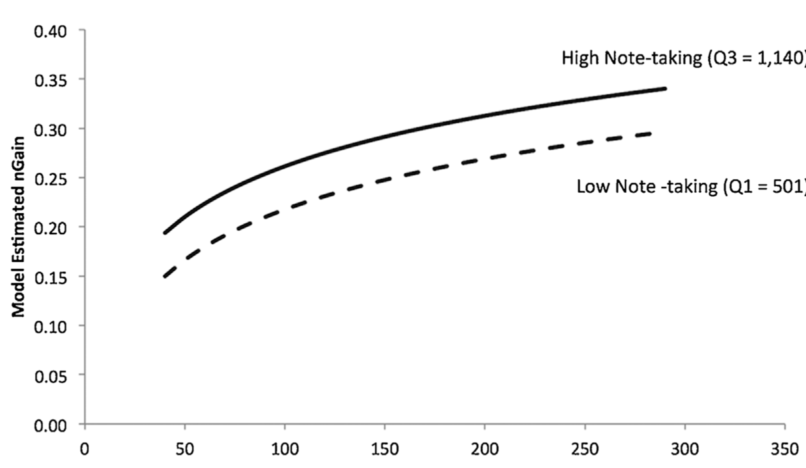


Figure 9. A summary of the third regression model: normalized gain scores as outcome and Manipulation, logtransformed Simulation, and Note as predictors.

transformed Simulation and Note were statistically significant predictors of the normalized gain scores, suggesting that students who used more simulations and/or spent more time taking notes learned more science knowledge. However, because the log-transformed Simulation was used in the model, the effect of each additional simulation usage on learning gain would gradually diminish with the increase in the total count.

The joint relationship between the three design operation measures and normalized gain scores is summarized in Figure 9. We note that a doubling of simulation usage (e.g., from 100 to 200) is related to approximately a 0.05 unit increase in student gain scores, on average in the population and controlling for model manipulation and note-taking behaviors ($d = 0.37$). As previously mentioned, this trend is monotonic but non-linear (logarithmic). We also note that, on average in the population, students who spent relatively more time taking notes (set at the third quartile value of 1,140) had gain scores that were approximately 0.04 units higher than their peers who spent less time taking notes (set at the first quartile value of 501), controlling for their model manipulation and simulation usage behaviors ($d = 0.31$). In other words, a doubling in the amount of simulation usage and the difference between high and low note-taking behaviors (as defined by the inter-quartile range), are each related to about a third of a standard deviation unit difference in normalized gain scores, controlling for all other factors in the model.

To check whether data points from a few students significantly influenced the regression analysis results, Cook's distances were calculated for all cases and the conventional cutoff value ($4/n = 4/83 = .048$) was used to identify influential cases. Results are presented in appendix 2. We note that, while the controlled values of some coefficients did fluctuate when various influential cases were removed, the global F-statistic remained statistically significant. This indicates that, taken as a whole, our three predictors jointly predicted variation in the outcome, even when influential cases were removed. We therefore choose to retain all cases in evaluating our final model.

Discussion

Connecting the design space and science space is critical to all types of design-science integration approaches. We argue that design tools play important roles in shaping design behaviors as well as science learning processes and outcomes. This study aimed to investigate secondary students' science knowledge development and its relationship with their tool-mediated design actions in a tool-rich design environment with minimal explicit guidance. The students in this study substantially improved their understanding of green building science as a result of completing an energy-efficient home design project in a simulated environment for engineering design (SEED). Three types of tool-mediated design actions—representation, analysis, and reflection—were identified and measured by the cumulative counts of model manipulation, simulation usage, and note-taking operations logged by the computer during the design challenge. The three design operation measures were jointly associated with the learning gains with a medium effect size. However, the three design operation measures were linked with learning gains in different ways. Below we discuss each of these findings in detail.

Model Manipulation

On average in the population, students who performed more model manipulation showed higher learning gains than their peers (p-value $< .05$). When taking simulation usage into account, there was still a positive relationship between model manipulation and learning gains, though it is less clear that this relationship is also present in the population (p-value $= .066$). This unstable relationship may be attributed to the multiple design actions that possibly require the model manipulation operations. Substantial amount of model manipulation happened as students created initial designs or perfected their final designs. To create an initial design, they needed to erect a building and add a roof, doors, windows, and solar panels. Many students also took great effort to edit various aspects of their designs. These structure-focused construction operations should not directly lead to conceptual understanding.

However, students need to manipulate certain features of their designs to analyze their functions and behaviors. For example, students had to move the solar panels to different sides of the roof to test which side produced most energy output. So more simulation usage was accompanied by more model manipulation, as evident in the positive correlation between these two measures. Thus, simulation usage may have somewhat confounded the association between model manipulation and learning gain. This explains why when simulation usage was controlled, the predictive strength of model manipulation became weaker.

Yet, the association between model manipulation and learning gains was still notable with marginal statistical significance, after controlling for simulation usage and note taking. It is possible that the models students created helped them recognize and understand some design issues that they were unaware of (Cardella et al., 2006). For example, one student initially put several tall trees on the south side of her house, thinking that the trees would reduce solar heat coming through the windows to keep the house cool in the summer. While she was right about solar heat reduction, the tall trees largely comprised the energy production of the solar panels installed on the south-facing roof. Several iterations later, she realized the problem with the trees and removed them. She would not have learned the lesson if her initial design did not include trees. This type of unintentional model manipulation operations might have helped students develop understanding of the underlying science.

Simulation Usage

Controlling for the model manipulation and note-taking behaviors of a prototypical student, a doubling in simulation usage (e.g., 100–200) was associated with a .05 unit difference in normalized gain. This represents an effect size of approximate 0.37 (Cohen's *d*, Cohen, 1992). This may be attributed the changing mental processes underlying simulation usage. During the early stage of the design project, students might use simulations to build knowledge of the building system. For instance, many students in the present study extensively explored the simulations of the Sun's position, shadow patterns, and insolation patterns during the first and second design sessions to learn the impacts of the Sun. As their understanding reached certain level, their usage of the simulation tools were more focused on testing and optimizing their designs.

The other possible explanation is that students who used simulation tools less frequently actually processed the analysis outputs more deeply than those who used more frequently. Take the Energy Annual Analysis for an example. It provides rich and complex information about a building' net energy, energy generated by solar panels, and energy consumed by heating and cooling over 12 months in one time graph. Ideally, students would spend some time to read the time graphs to understand the seasonal patterns of energy consumption and generation. However, some students may only quickly scan the net energy and go back to change their designs. This superficial use still would allow students to acquire some basic knowledge about the design space, but it was not helpful for developing more sophisticated understanding. This finding is consistent with Apedoe's and Schunn's (2013) study showing that highly successfully design teams created and tested more designs than lower performing teams. But note that in this previous study, participants' simulation usage behaviors were compared based on their design performance or design expertise. The present study showed instead the linkage between simulation usage behaviors and knowledge development.

Note-Taking

The time spent on taking notes alone was not significantly predictive of learning gain, but it became a significant positive predictor when model manipulation and simulation usage were held constant. It is likely that the variance explained by model manipulation and simulation usage masked the effect of note taking. The removal of this variance revealed the significant positive effect of note taking over and above the other two operations. It is reasonable that with similar amount of model manipulation and simulation usage, students who spent more time taking notes were more likely to develop deeper understanding of science concepts. For instance, Emma and Tim (pseudo names), two students in the present study, started with similar pretest score ($PRE_{Emma} = 30$ *vs* $PRE_{Tim} = 28$) but ended up with a substantial difference in the posttest ($POST_{Emma} = 49$ *vs* $POST_{Tim} = 32$). Their user logs revealed that while they had similar amount of model manipulation ($Manimuplation_{Emma} = 1522$, $Manimpulation_{Tim} = 1605$) and simulation usage ($Simulation_{Emma} = 129$, $Simulation_{Tim} = 122$), Emma spent much more time taking notes than Tim ($Note_{Emma} = 1846$, $Note_{Tim} = 320$). As shown in Table 7, their notes also showed different levels of completeness and reflectivity. Emma recorded 13 iterations in total and wrote complete and thorough notes for each, while Time only recorded two iterations and his notes lacked clarity and reflection.

Table 7
Comparison of notes between two students, Emma and Tim

Emma (13 Iterations in Total)	Tim (2 Iterations in Total)
Iteration 1 Net energy usage: "48,252" Note: "I make the basic design for the shape of the house and put in windows. The house is white. There is no insulation."	*Iteration 1* Net energy usage: "- 4120" Note: none
Iteration 2 Net energy usage: "45,733" Note: "I paint the house colors that are not white, and put trees in front of the windows on the south side of the house. I still haven't insulated or put in any solar panels."	*Iteration 2* Net energy usage: "- 3059" Note: "just made a couple arrangements with the windows and roof to look nicer"
Iteration 3 Net energy usage: "5,101" Note: "I add 20 solar panels at level 15, and insulate everything. Insulation makes a difference, guys."	
Iteration 4 Net energy usage: "778" Note: "I add more solar panels and place trees strategically in front of the windows. I am almost out of money. Time to power through."	
Iteration 5 Net energy usage: "- 515" Note: "I moved my solar panels closer together so that I could fit more on the southern wall. Then, I added more solar panels so that I had exactly 40 solar panels. It also turned out that one of my walls was not perpendicular to the other sides of the house, but was slightly tinted. Once I fixed that, my net energy usage went down significantly. I'm in the negatives."......	

This finding is consistent with previous studies showing the positive association between reflective activity and science learning outcome (Goldstein, Purzer, Zielinski, & Adams et al., 2015; Wendell & Lee, 2010). However, our analysis showed more nuanced role of note taking. Specifically, the positive effect of note taking is contingent on sufficient model manipulation and simulation usage.

To sum up, there are three pieces of evidence that point to the impacts of tools on design behaviors and science learning. First, the students in this study substantially improved their understanding of green building science as a result of completing an energy-efficient home design project in a SEED. We intentionally minimized social mediation and direct instruction so that the SEED would be the main source of learning. Thus, it is reasonable to attribute students' learning gains to their interactions with the tools provided in the SEED. Second, students' design actions, particularly those theoretically inducive of learning, were positively linked with their learning gains. This linkage further corroborates the tools as the sources of learning, as tools mediated the design actions. Third, the three design operations (i.e., model manipulation, simulation usage, and note-taking) were linked with learning gains in different ways and also connected among themselves in a certain pattern. Specifically, it appears that model manipulation serves as entry to the concrete design space; use of simulations functions as a bridge onto the science space; and note-taking as an elevator to lift the mind off ground and overlook the paths taken—learning happens here and the new knowledge instigates a new cycle of design. This pattern likens the theoretical design activity system we described in the introduction. The movement of this "three-gear cycle" is fueled by the feedback generated or induced by the design tools.

So what caused students to perform different amount of design operations and learned different amount of science knowledge? There are two personal choices that may lead students to different learning paths. First, students' initial and intermediate designs present different learning opportunities. For example, if students prefer big windows for their houses, they may struggle with both the thermal insulation issue and the excessive solar heat issue, which present great opportunities to learn the concepts of thermal conduction and thermal radiation. However, if they start out with small windows, these issues would be much less prominent, resulting in little opportunity to learn. Second, individual students may respond to the same analysis results differently. For instance, some students may give up after several attempts to keep their big windows, but some may explore more possibilities to make it work. This persistence may lead to discovery of new design elements such as the windows' solar heat gain coefficient or design principles such as the optimal orientation of big windows. These personal choices are functions of students' interest levels, aesthetic preferences, prior domain knowledge, and metacognitive skills. The potential impacts of these factors on learning paths warrant further investigations.

While the present study was focused on design tools, it is important to not overlook the complexity of classroom environments. According to the regression analysis results, there was still a large amount of variance (84.6%) in students' learning gains remained unexplained. Several factors may be at play. First, as the test responses were scored based on scientific sophistication demonstrated in students' justifications for their design choices, different levels of argumentation skills among the students might have affected their scores. Second, although students worked individually on their designs, they did sometimes communicate with each other about the project. The different amount of peer communication might explain some variance in the learning gains. Furthermore, even though the curriculum was implemented under the condition of minimal explicit guidance, the teacher did help the students who struggled to ensure that they fully benefitted from the learning opportunity. This need-based intervention possibly somewhat weakened the link between students' design behaviors and their learning gains.

Implications

With the rapid advances in technology and increased interest in integrating science and engineering learning, we expect that more SEEDs will be developed or adapted from industrial-grade CAD tools. Based on the findings of this study, we offer several suggestions for developers to conceptualize SEEDs for particular settings or educators to evaluate and select appropriate SEEDs for their classrooms.

First, this study clearly demonstrates the efficacy of SEEDs in supporting design-science integrated learning. Attention to the provision of design tools can substantially complements the limited or lack of scaffolding from social environment (e.g., teacher's feedback and peer review). This great potential of SEEDs warrants investment into their development and evaluation. However, it should be noted that there have been concerns about using CAD tools to teach design. For instance, Lawson (1999) argued that CAD tools could impede creativity because they tended to limit design to the structures available in the software. Kimbell and Stables (2007) found that CAD tools hindered reflective practices among students due to the additional effort required to master the software. In light of these healthy skepticisms, we must emphasize that SEED is a special type of CAD environments specifically designed for engineering and science learning. The design tools offered in SEEDs are intended to scaffold the process of engineering design and scientific inquiry. While generic CAD tools may offer abundant modeling capabilities to maximize flexibility for designers, SEEDs would offer a smaller but curated collection to direct students' attention to the targeted science and engineering concepts. SEEDs would also reduce the difficulty and number of operations for students to transform their abstract ideas into concrete artifacts. For instance, providing premade structures would allow students to achieve their goals without struggling with spatial composition. Similarly, the analysis and reflection tools provided by SEEDs may not be available or tailored for student learning in generic CAD environments. These qualitative differences should be duly noted and used as criteria for SEED development or selection.

However, questions remain about how SEEDs may function in learning environments with rich social mediation. As noted by many engineering educators (e.g., Bucciarelli, 1994; Dym, Agogino, Eris, Frey, & Leifer, 2005), design is a highly social activity in contemporary engineering fields and designing in a team environment is a critical skill that students should develop through engineering education. The present study was limited in this respect. We intentionally chose to control social mediation by minimizing explicit guidance and asking students to work individually. This methodological decision inevitably prevented us to observe the social interactions around the design project and design tools. The importance and ubiquitousness of social mediation in classroom learning warrants future investigations that focus on how students' design behaviors and science learning are shaped by both design tools and interactions with others.

Secondly, this study demonstrates that cultural-historical activity theory (CHAT) is a useful theoretical framework to interpret how learning occurs through design. Under this framework, knowledge acquisition is not considered as the end goal from the students' perspective. Instead, knowledge is the mental instrument that students acquire, use, and adapt to pursue their goals. As students use tools, they gradually internalize the knowledge carried by the tools. Thus, SEED designers must consider how to represent knowledge in the potential ways students may interact with tools. Such considerations may lead to unconventional design choices for the user interface. For instance, generic CAD tools may set the default values of material properties at the optimal levels for convenience, but in SEEDs these default values would be set at the suboptimal levels to offer learning opportunities. To guide the design of SEEDs, future research should be focused on how individual features and collections of features shape learning paths through qualitative analysis. CHAT will be particularly useful to frame such analysis as it has significantly influenced the research on human-computer interaction (Nardi, 1996)

For teachers, SEEDs offer many benefits, but they also pose new challenges. On top of the pedagogical content knowledge of science (Magnusson, Krajcik, & Borko, 1999) and design (Crismond & Adams, 2012) that teachers must have to teach science and engineering design, they also need to master the technologies and understand how students learn with the technologies. This technological pedagogical content knowledge (TPACK) (Koehler & Mishra, 2009) is critical to the successful use of SEEDs. Yet, teachers must invest much time and effort to become fully prepared. To alleviate the burden on teachers, future research should be focused on building robust TPACK frameworks to guide SEED developers and teacher educators to develop effective TPACK-based teacher training programs.

Limitations

The findings of this study can only be generalized to student population similar to the sample used in this study. Based on the school's demographics, the majority of the participants were college-bound. Our classroom observation indicated that they had high-level computer literacy and sufficient ability to learn new technologies. Yet, the participants had very little formal engineering design experience. Furthermore, the curriculum was intentionally designed to limit the amount of explicit guidance. In doing so, we were not suggesting certain instructional approaches (e.g., constructivism vs. cognitivism; Hmelo-Silver, Duncan, & Chinn, 2007; Kirschner, Sweller, & Clark, 2006) were more advantageous than others. Instead, this condition was chosen to reduce noise from social mediation so that tool-mediated design actions could be better observed. These student characteristics and research design features should be considered when using the findings of this study to inform curriculum design or formulate new research hypotheses.

Additionally, as the findings of this study emerged from a SEED, their generalizability to generic learning environments may be limited. This is a trade-off for SEED's user logging capability that permits the fine-grained observations of students' design actions. A related limitation is that the user interface of the SEED undoubtedly shapes students' design behaviors and science learning outcomes. Changes to the user interface, even within the same SEED would inevitably alter users' behaviors, resulting in somewhat different learning pathways. Thus, interpretation of the findings should always take into account the particular user interface deployed in the study.

Lastly, the computer logs were aggregated under three broad categories of design operations (i.e., model manipulation, simulation usage, and note taking). This practice inevitably incurred loss of information in the temporal order of these design operations. The sequence of design operations may indicate qualitatively different ways of exploring the design space and science space. These distinct pathways may explain different learning outcomes on top of the sheer quantities of design operations investigated in the present study. Future research should investigate the temporality of design behaviors using techniques such as latent sequential analysis (Gardner, 1995; Gottman & Roy, 1990).

Conclusion

Design tools have the potential to help students connect design projects and the underlying science, bridging the design-science gap frequently observed in design-science integrated learning environments. This study showed that secondary students were able to develop considerable amount of science knowledge through completing a design project in a simulated environment for engineering design that provided many useful design tools. Further, their learning gains were positively associated with three types of design actions—representation, analysis, and reflection—measured by the cumulative counts of relevant computer logs. In addition, these design actions were linked with learning gains in ways that were consistent with their hypothetical impacts on knowledge development. These findings suggest that, instead of being passive components in a learning environment, tools considerably shape design processes and learning paths. As such, tools offer possibilities to help bridge the design-science gap. Attention to the provision of design tools may substantially complements the limited or lack of scaffolding from social environment (e.g., teacher's feedback and peer review). Technologies such as the simulated environment for engineering design used in this study are well worth further development and research effort.

Note

[1] Actions on floor, foundation, sensor, and human were excluded because they have no impact on the energy performance of the building.

References

Adams, R. S., Turns, J., & Atman, C. J. (2003). Educating effective engineering designers: The role of reflective practice. Design Studies, 24(3), 275–294. https://doi.org/10.1016/S0142-694X(02)00056-X

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

Bamberger, Y., Cahill, C., Hagerty, J., Short, H., & Krajcik, J. S. (2010). Learning science by doing design: How can it work at the middle school level? Journal of Education, Informatics and Cybernetics, 2(2), 41–46.

Barnett, M. (2005). Engaging inner city students in learning through designing remote operated vehicles. Journal of Science Education and Technology, 14(1), 87–100. https://doi.org/10.1007/s10956-005-2736-z

Bucciarelli, L. L. (1994). Designing engineers. Cambridge, MA: MIT Press.

Cardella, M. E., Atman, C. J., & Adams, R. S. (2006). Mapping between design activities and external representations for engineering student designers. Design Studies, 27(1), 5–24. https://doi.org/10.1016/j.destud.2005.05.001

Carlsen, W. S. (1998). Engineering design in the classroom: Is it good science or is it revolting. Research in Science Education, 28(1), 51–63.

Carroll, M., Goldman, S., Britos, L., Koh, J., Royalty, A., & Hornstein, M. (2010). Destination, imagination and the fires within: Design thinking in a middle school classroom. International Journal of Art & Design Education, 29(1), 37–53.

Chiu, J. L., Malcolm, P. T., Hecht, D., DeJaegher, C. J., Pan, E. A., Bradley, M., & Burghardt, M. D. (2013). WISEngineering: Supporting precollege engineering design and mathematical understanding. Computers and Education, 67, 142–155. https://doi.org/10.1016/j.compedu.2013.03.009

Clement, J., Brown, D. E., & Zietsman, A. (1989). Not all preconceptions are misconceptions: Finding "anchoring conceptions" for grounding instruction on students' intuitions. International Journal of Science Education, 11(5), 554–565.

Cohen, J. (1992). A power primer. Psychological Bulletin, 112(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Crismond, D. P. (2001). Learning and using science ideas when doing investigate-and-redesign tasks: A study of naive, novice, and expert designers doing constrained and scaffolded design work. Journal of Research in Science Teaching, 38(7), 791–820. https://doi.org/10.1002/tea.1032

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

Cross, N. (1999). Natural intelligence in design. Design Studies, 20(1), 25–39. https://doi.org/10.1016/S0142-694X(98)00026-X

Dym, C. L., Agogino, A. M., Eris, O., Frey, D. D., & Leifer, L. J. (2005). Engineering design thinking, teaching, and learning. Journal of Engineering Education, 94(1), 103–120.

Engeström, Y. (1987). Learning by expanding: an activity-theoretical approach to developmental research. Helsinki: Orienta-Konsultit https://doi.org/10.1016/j.intcom.2007.07.003

Engeström, Y., Miettinen, R., & Punamäki, R. L. (1999). Perspectives on activity theory. Cambridge, MA: Cambridge University Press.

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

Fortus, D., Krajcik, J., Dershimer, R. C., Marx, R. W., & Mamlok-Naaman, R. (2005). Design-based science and real-world problem-solving. International Journal of Science Education, 27(7), 855–879. https://doi.org/10.1080/09500690500038165

Gardner, W. (1995). On the reliability of sequential data: Measurement, meaning, and correction. In J. M. Gottman (Ed.), The analysis of change (pp. 339–360). Mahwah, N.J.: L. Erlbaum Associates.

Goldstein, M. H., Purzer, Ş., Adams, R. S., & Xie, C. (2015). Exploring the relationship between student reflectivity and their understanding of informed design. *Paper Presented at Research in Engineering Education Symposium (REES)*. Dublin, Ireland.

Goldstein, M. H., Purzer, Ş., Mejia, C. V., Zielinski, M., & Douglas, K. A. (2015). Assessing idea fluency through the student design process. Paper Presented at Frontiers in Education Conference (FIE). El Paso, TX.

Goldstein, M. H., Purzer, Ş., Zielinski, M., & Adams, R. S. (2015). High school students' ability to balance benefits & tradeoffs while engineering green buildings. In *Proceedings of the 122nd American Society for Engineering Education Annual Conference and Exposition*. Seattle, WA.

Gottman, J. M., & Roy, A. K. (1990). Sequential analysis: a guide for behavorial researchers. New York: Cambridge University Press.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. American Journal of Physics, 66(1), 64. https://doi.org/10.1119/1.18809

Hens, H. (2011). Applied Building Physics. Berlin, Germany: Ernst & Sohn.

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: a response to Kirschner, Sweller, and Clark (2006). Educational Psychologist, 42(2), 99–107. https://doi.org/10.1080/00461520701263368

Hmelo, C. E., Holton, D. L., & Kolodner, J. L. (2000). Designing to learn about complex systems. Journal of the Learning Sciences, 9(3), 247–298. https://doi.org/10.1207/S15327809JLS0903_2

Johnson, B., & Christensen, L. B. (2008). Educational research: Quantitative, qualitative, and mixed approaches (3rd ed.). Los Angeles: Sage Publications.

Johnson, M. (1987). The body in the mind: the bodily basis of meaning, imagination, and reason. Chicago, IL: University of Chicago Press.

Jones, M. G., Brader-Araje, L., Carboni, L. W., Carter, G., Rua, M. J., Banilower, E., & Hatch, H. (2000). Tool time: Gender and students' use of tools, control, and authority. Journal of Research in Science Teaching, 37(8), 760–783. https://doi.org/10.1002/1098-2736(200010)37:8<760::AID-TEA2>3.0.CO;2-V

Jordan, M. E., & McDaniel, R. R. (2014). Managing uncertainty during collaborative problem solving in elementary school teams: The role of peer influence in robotics engineering activity. Journal of the Learning Sciences, 23(4), 490–536. https://doi.org/10.1080/10508406.2014.896254

Kannengiesser, U., Gero, J. S., Wells, J., & Lammi, M. D. (2015). Do high school students benefit from pre-engineering design education? Paper Presented at International Conference on Engineering Design – IC ED2015. Milan, Italy.

Kaptelinin, V. (1996). Activity theory: Implications for human-computer interaction. In B. A. Nardi (Ed.), Context and consciousness: Activity theory and human-computer interaction (pp. 53–59). Cambridge, MA: MIT Press. https://doi.org/10.1093/ntr/ntr215

Kim, P., Suh, E., & Song, D. (2015). Development of a design-based learning curriculum through design-based research for a technology-enabled science classroom. Educational Technology Research and Development, 63(4), 575–602. https://doi.org/10.1007/s11423-015-9376-7

Kimbell, R., & Stables, K. (2007). esearching design learning: Issues and findings from two decades of research and development. Dordrecht, Netherlands: Springer Science & Business Media.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. Educational Psychologist, 41(2), 75–86. https://doi.org/10.1207/s15326985ep4102

Koch, G. G. (1982). Intraclass correlation coefficient. In S. Kotz, & N. L. Johnson (Eds.), Encyclopedia of Statistical Sciences (pp. 213–217). New York: John Wiley & Sons.

Koehler, M. J., & Mishra, P. (2009). What is technological pedagogical content knowledge (TPACK)? Contemporary Issues in Technology and Teacher Education, 9(1), 60–70. https://doi.org/10.1016/j.compedu.2010.07.009

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

Lammi, M. D., Denson, C., Wells, J., Paretti, M. C., Williams, C., & Gero, J. S. (2014). Initial findings of high school pre-engineering and non-engineering students' design cognition. In *Frontiers in Education Conference (FIE)*. Madrid, Spain.

Lammi, M. D., & Gero, J. S. (2011). Comparing design cognition of undergraduate engineering students and high school pre-engineering students. In *Frontiers in Education Conference (FIE)*. Rapid City, SD, USA.

Lawson, B. (1999). "Fake" and "Real" creativity using computer aided design. In Proceedings of the third conference on Creativity & cognition−C&C '99 (pp. 174–179). https://doi.org/10.1145/317561.317591

Lemons, G., Carberry, A., Swan, C., Jarvin, L., & Rogers, C. (2010). The benefits of model building in teaching engineering design. Design Studies, 31(3), 288–309. https://doi.org/10.1016/j.destud.2010.02.001

Leont'ev, A. N. (1978). Activity, consciousness, and personality. Englewood Cliffs, N.J.: Prentice-Hall Inc.

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. https://doi.org/10.1002/tea.20111

Magnusson, S., Krajcik, J., & Borko, H. (1999). Nature, sources, and development of pedagogical content knowledge for science teaching. In Examining pedagogical content knowledge (pp. 95–132). Netherlands: Springer. https://doi.org/10.1007/0-306-47217-1_4

Mentzer, N. (2014). High school student information access and engineering design performance. Journal of Pre-College Engineering Education Research, 4(1), 32–42. https://doi.org/10.7771/ 2157-9288.1074

Mentzer, N., Becker, K., & Sutton, M. (2015). Engineering design thinking: High school students' performance and knowledge. Journal of Engineering Education, 104(4), 417–432. https://doi.org/10.1002/jee.20105

Montoya, M. (2010). Upper Saddle River, N.J.: Prentice Hall.

Nardi, B. A. (1996). Context and consciousness: activity theory and human-computer interaction. Cambridge, MA: MIT Press.

Norman, D. A. (1999). Affordance, conventions, and design. Interactions, 6(3), 38–43.

Puntambekar, S., & Kolodner, J. L. (2005). Toward implementing distributed scaffolding: Helping students learn science from design. Journal of Research in Science Teaching, 42(2), 185–217. https://doi.org/10.1002/tea.20048

Purcell, A. T., & Gero, J. S. (1998). Drawings and the design process: A review of protocol studies in design and other disciplines and related research in cognitive psychology. Design Studies, 19(4), 389–430.

Purzer, Ş., Goldstein, M. H., Adams, R. S., Xie, C., & Nourian, S. (2015). An exploratory study of informed engineering design behaviors associated with scientific explanations. International Journal of STEM Education, 2(1), 9. https://doi.org/10.1186/s40594-015-0019-7

Roth, W.-M. (1995). Inventors, copycats, and everyone else: The emergence of shared resources and practices as defining aspects of classroom communities. Science Education, 79(5), 475–502. https://doi.org/10.1002/sce.3730790502

Roth, W.-M. (2001). Learning science through technological design. Journal of Research in Science Teaching, 38(7), 768–790. https://doi.org/10.1002/tea.1031

Roth, W.-M., Tobin, K., & Ritchie, S. (2001). Re/Constructing elementary science. New York, NY, USA: Peter Lang Publishing Inc.

Schnittka, C., & Bell, R. L. (2011). Engineering design and conceptual change in science: Addressing thermal energy and heat transfer in eighth grade. International Journal of Science Education, 33(13), 1861–1887. https://doi.org/10.1080/09500693.2010.529177

Schoultz, J., Säljö, R., & Wyndhamn, J. (2001). Heavenly talk: Discourse, artifacts, and children's understanding of elementary astronomy. Human Development, 44, 103–118.

Sullivan, F. R. (2008). Robotics and science literacy: Thinking skills, science process skills and systems understanding. Journal of Research in Science Teaching, 45(3), 373–394. https://doi.org/10.1002/tea

Sun, C. T., Wang, D. Y., & Chang, Y. Y. (2013). Effects of thinking style on design strategies: Using bridge construction simulation programs. Educational Technology and Society, 16(1), 309–320.

Teng, P. S., Cai, D., & Yu, T. K. (2014). The relationship between individual characteristics and ideation behavior: An empirical study of storyboards. International Journal of Technology and Design Education, 24, 459–471. https://doi.org/10.1007/s10798-014-9264-1

Vattam, S. S., & Kolodner, J. L. (2008). On foundations of technological support for addressing challenges facing design-based science learning. Pragmatics & Cognition, 16(2), 406–437. https://doi.org/10.1075/p&c.16.2.08vat

Verstijnen, I. M., van Leeuwen, C., Goldschmidt, G., Hamel, R., & Hennessey, J. M. (1998). Sketching and creative discovery. Design Studies, 19(4), 519–546. http://doi.org/10.1016/S0142-694X(98)00017-9

Vygotsky, L. S. (1981). The instrumental method in psychology. In J. V. Wertsch (Ed.), The concept of activity in Soviet psychology (pp. 134–143). Armonk, NY: M.E. Sharpe.

Welch, M. (1999). Analyzing the tacit strategies of novice designers. Research in Science & Technological Education, 17(1), 19–34. https://doi.org/10.1080/0263514990170102

Wendell, K. B., & Lee, H.-S. (2010). Elementary students' learning of materials science practices through instruction based on engineering design tasks. Journal of Science Education and Technology, 19(6), 580–601. https://doi.org/10.1007/s10956-010-9225-8

Wertsch, J. V. (1994). The primacy of mediated action in sociocultural studies. Mind, Culture, and Activity, 1(4), 202–208. https://doi.org/10.1080/10749039409524672

Xie, C., & Nourian, S. (n.d.). Energy3D: Learning to build a sustainable future. Retrieved February 1, 2017, from https://energy.concord.org/energy3d/

Xie, C., Zhang, Z., Nourian, S., Pallant, A. M. Y., & Bailey, S. (2014). On the instructional sensitivity of CAD logs. International Journal of Engineering Education, 30(4), 760–778.

Appendix 1 Green Building Science Test

Name: ______________________________ Class Section: _______________________

[WINDOW_DIRECTION]
1. You are designing a house for a client in Boston, Massachusetts. The client would like to have a large window on one side of the house. There are no trees or other buildings around this house. To maximize the energy efficiency of the house in the winter, on which side of the house would you choose to install the large window?

a) East side
b) West side
c) North side
d) South side
e) I am not sure

Please explain your choice:
____
____
____
____
____

Answer key: D. The large south-facing window will be most energy efficient in the winter. In Chicago, which is in the northern hemisphere, the sun rises in the southeast, peaks out at a low angle above the southern horizon, and then sets in the southwest. Because the sun is on the south side of the house all day long, the south-facing window will receive most heat through thermal radiation. The larger the window, the more heat gain. Thus, having the large window on the south wall is the most energy efficient choice.

[NOON_SHADE]
2. A house is built on a flat lot with no trees or other buildings around in Boston, Massachusetts. Which side of the house would be shaded most at NOON regardless of the time of the year?

a) East side
b) West side
c) North side
d) South side
e) All sides are equally shaded
f) I am not sure

Please explain your choice:

Answer key: C. The north side of the house will be shaded most at noon regardless of time of the year. In Boston, which is in the northern hemisphere, the sun rises in the southeast, peaks out at a low angle above the southern horizon, and then sets in the southwest. Thus at noon, the house blocks sunlight from shining on the north wall.

[ROOF_COLOR]
3. Which of the following colors would you choose for the WALLS of your house in order to reduce your cooling cost in the summer?

a) Black
b) Dark gray
c) Light gray
d) Color doesn't matter
e) I am not sure

Please explain your choice:

Answer key: C. Light colored roof would be more energy efficient. Light color reflects more light than dark color, thus it absorbs less heat and stays relatively cooler. The cooler roof in the summer reduces the temperature difference between the roof and the inside, which in turns slow down heat transfer from the inside to the roof. As a result, the energy is required to cool the house.

[WINDOW_U_VALUE]
4. If you were to replace a large single-pane window (U-factor: 5.91 W/(m^2. °C)) in your house with a double-pane one of the same size (U-factor: 3.12 W/(m^2. °C)), what will happen to the energy required to cool your house in the summer?

a) It will stay the same
b) It will increase
c) It will decrease
d) I am not sure

Please explain your choice:

Answer key: C. The energy required to cool the home will decrease. The U-factor is a measure of the rate of heat loss of a particular building material. The smaller the U-factor, the slower heat transfers through the window, requiring less energy to keep the house cool.

[SOLAR_SIDE]
5. For a house located in Boston, Massachusetts, which side of the roof would you choose to install solar panels in order to maximize their ability to convert solar energy to electricity?

a) East side
b) West side
c) North side
d) South side
e) I am not sure

Please explain your choice:

Answer key: D. The sunlight in the northern hemisphere is mostly from the south. Solar panels need to receive direct sunlight to maximize their efficiency. Thus, installing solar panels on the south side of the roof will maximize their efficiency.

[SOLAR_ANGLE]
6. For a house located in Miami, Florida, on which type of roof would solar panels be most efficient?

a) A steep roof
b) A slightly sloped roof
c) A flat roof
d) The angle of the roof does not matter
e) I am not sure

Please explain your choice:

____

____

____

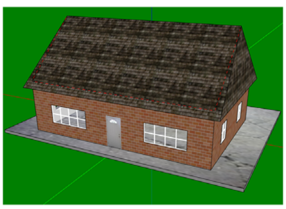
____

____

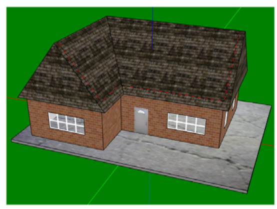
Answer key: B. A slightly sloped roof would maximize the efficiency of solar panels. The sun is higher in locations closer to the equator. A slightly sloped roof would increase the amount of insolation, which in turn improves the efficiency of installed solar panels.

[HOUSE_SHAPE]

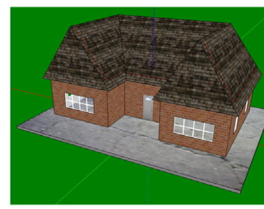
7. The three houses below have identical square footage, windows, and roof height and material. The only difference among them is their shape. In a WINTER NIGHT, which house would lose the most heat?



Rectangle



L-shaped



U-shaped

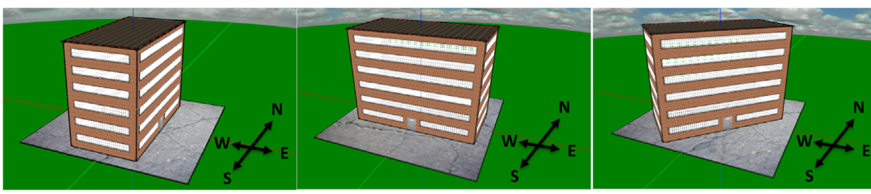
a) The rectangle house
b) The L-shaped house
c) The U-shaped house
d) They would lose the same amount of heat
e) I am not sure

Please explain your choice:

Answer key: C. The U-shaped house loses heat the fastest, the L-shaped house the moderate, and the rectangle house the slowest. According to Fourier's law of thermal conduction, thermal conduction, the large surface area the more heat transfer. With the same square footage, the U-shaped house has the largest surface area, while the rectangle house has the smallest surface area. With all other conditions the same, the U-shaped house loses heat the fastest and the rectangle house loses heat the slowest.

[RECTANGLE_ORIENTATION]

8. A building will be built on a 50m by 50m block with no trees or other buildings around in Boston, Massachusetts. The dimensions of the building are set to be 40m long, 20m wide, and 30m high. Below are three ways to orient the building on the block. Which orientation would allow the building to gain the MOST energy from the sun in the WINTER?



Orientation 1 Orientation 2 Orientation 3

a) Orientation 1
b) Orientation 2
c) Orientation 3
d) The building would absorb the same amount of solar energy no matter how it is oriented.
e) I am not sure.

Please explain your choice:

Answer key: B. The long axis of the building runs east-west. In the northern hemisphere, where Washington, DC is located, the east and west walls of a building gain most solar energy among all four walls. Setting the long axis of the building along the east-west direction will reduce the surface areas of the east and west walls, resulting in decreases in solar energy gains on these two walls.

[TREE_TYPE]
9. A two-story south-facing house is about 30 feet high and located in Massachusetts. The house owner wants to plant trees to improve energy efficiency of the house. Which of the following trees would you choose, assuming the total costs of these options are the same?

a) Ten boxwood trees, 5 feet tall and 4 feet wide, Evergreen (doesn't shed leaves annually)
b) Ten Japanese barberry trees, 5 feet tall and 4 feet wide, Deciduous (sheds leaves annually)
c) Two Weeping Alaskan Cedar trees, 25 feet tall and 20 feet wide, Evergreen (doesn't shed leaves annually)
d) Two White Dogwood trees, 25 feet tall and 20 feet wide, Deciduous (sheds leaves annually)
e) I am not sure

Please explain your choice:
____
____
____
____
____

Answer key: D. The taller the trees, the more hours of shade they can provide to reduce solar heat gain in the summer. However, in the winter, solar gain needs to be maximized. Thus, deciduous trees are better than evergreen trees because their bare branches block less sunlight during winter.

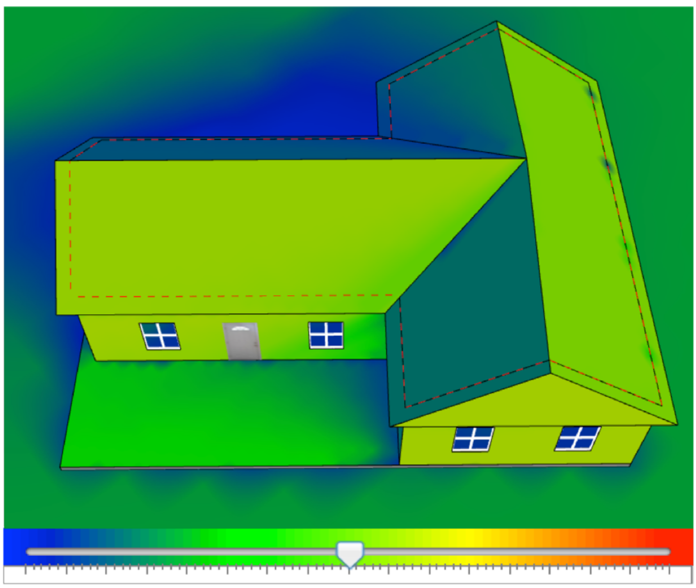
[TREE_SIDE]
10. For the house described in the last question (south-facing, two-story, located in Massachusetts), on which side of the house would you plant the trees?

a) East
b) Southeast
c) South
d) Southwest
e) West
f) I am not sure

Please explain your choice:

Answer key: E. West side. The west side and east side will receive a lot of heat during summer, and the west side feels even hotter than the east because the afternoon sunshine adds to the existing high temperature at the time. Planting trees on the west side of the house would block much sunlight coming from the west.

[SOLAR_HEAT_MAP]
11. The image below shows the temperature distribution of a house located in Boston, Massachusetts at noontime on a sunny day in January. As shown in the temperature gradient bar below the image, warmer colors (i.e., red side) represent higher temperature, and colder colors (i.e., blue side) represent lower temperature. Which direction does the front side face (the front side has the main door)?



a) East
b) Southeast
c) South
d) Southwest
e) West
f) I am not sure

Please explain your choice:

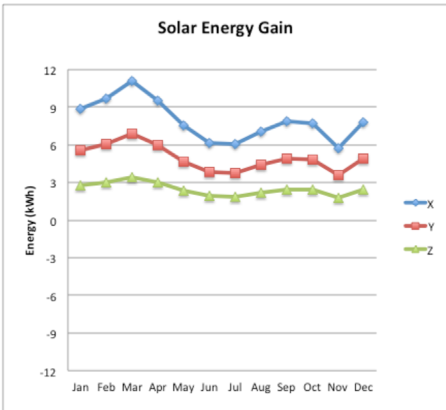
____

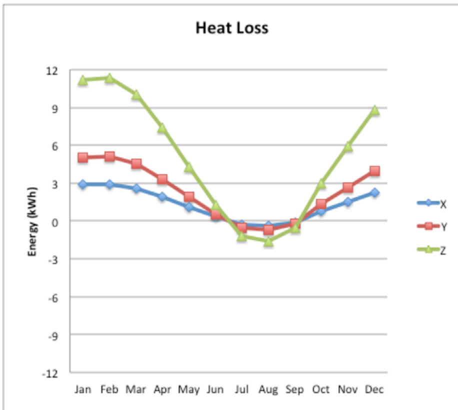
____

____

Answer key: D. The house faces southwest. According to the temperature color coding, green is warmer than blue. First, the roof on the front has warmer than that on the back, so the front must face south in general. Second, the right side of the secondary roof is warmer than the left side, so the right side must face the south. Thus, the house front must face southwest to generate both effects.

[WINDOW_GRAPH_S] [WINDOW_GRAPH_U]
12. Three windows of the same size but different insulation and solar properties are installed on the south side of a house located in Boston, Massachusetts. The graphs below show the solar energy gain and heat loss through these three windows throughout the year. In the table below the graphs, please mark the appropriate Solar Heat Gain Coefficient (SHGC) and U-Factor for each window.





	Solar Heat Gain Coefficient (SHGC)	U-Factor W/(m^2. °C)
Window X (blue)	☐25% ☐50% ☐80%	☐1.53 ☐2.66 ☐5.91
Window Y (red)	☐25% ☐50% ☐80%	☐1.53 ☐2.66 ☐5.91
Window Z (green)	☐25% ☐50% ☐80%	☐1.53 ☐2.66 ☐5.91
	☐I am not sure	☐I am not sure

Please explain your choices

____

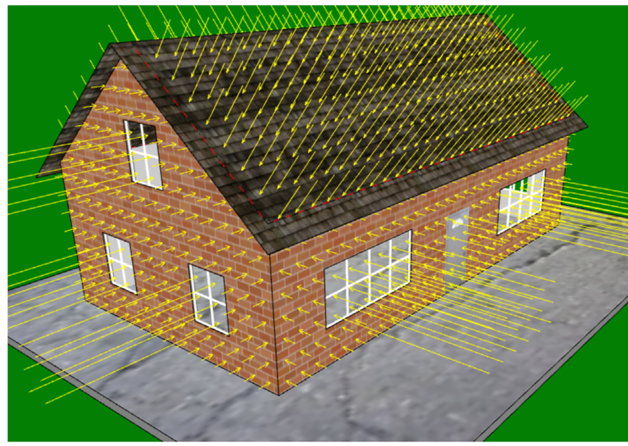
____

____

Answer key:
Window X: SHGC 80%, U-Factor 1.53
Window Y: SHGC 50%, U-Factor 2.66
Window Z: SHGC 25%, U-Factor 5.91
Higher SHGC results in higher solar energy gain, thus Window X has the highest SHGC and Window Z has the lowest SHGC. Higher U-Factor means faster heat loss, or greater negative heat gain, thus Window X has the lowest U-factor and Window Z has the highest U-Factor.

[HEAT_FLUX_SEASON]
13. The image below shows the heat flux analysis of a house. The arrows point to the direction of heat transfer, and the length of the arrows represents the rate of heat transfer (longer arrows indicate faster heat transfer). Which season is most likely selected for this heat flux analysis?

a) Spring
b) Summer
c) Fall
d) Winter
e) I am not sure



Please explain your choice:

Answer key: B. Heat transfers from the outside to the inside of the house, so the outside must be hotter than the inside. Thus, it is most likely to in the summer.

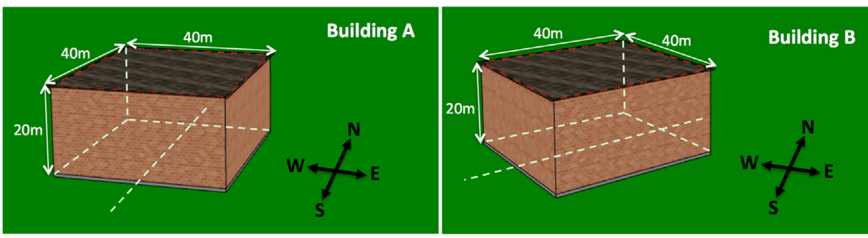
[HEAT_FLUX_U]
14. Based on the heat flux analysis of the house shown in the last question, which building element has the LOWEST U-factor?

a) Walls
b) Windows
c) Doors
d) Roof
e) I am not sure

Please explain your choice:

Answer key: A. As the arrows along the walls are the shortest, heat transfer is slowest there. So the walls must have a low rate of heat loss, which is indicated by the U-factor.

[SQUARE_ORIENTATION]
15. The two square buildings below are 40 meters wide, 40 meters long, 20 meters tall, and located in Boston, Massachusetts. They are identical except for their orientations. The axis of building A runs north-south, while the axis of building B runs southeast-northwest. Assuming there are no trees or other buildings around the two buildings, how do their TOTAL solar energy gains THROUGHOUT THE DAY compare?



Building A Building B

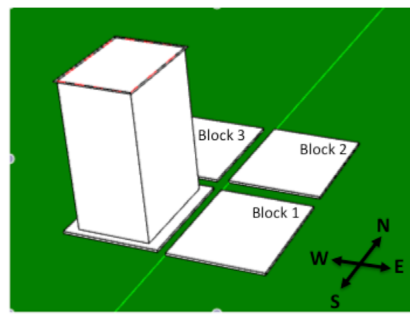
a) Building A gains more solar energy than building B.
b) Building A gains less solar energy than building B.
c) They have the same solar energy gain.
d) I am not sure.

Please explain your choice:

Answer Key 1: B. Larger surface area absorbs more solar energy. While all sides of building B receive sunlight, only three sides of building A do so. Thus, building A gains less solar energy than building B.
Answer Key 2: B. Insolation is determined by the width of the sunlight beam striking on a surface regardless of the angle between the beam and the surface. For both buildings, the width of the sunlight beam fluctuates between the length of the side (minimum) and the length of the diagonal (maximum) over the day. For building A, the width of the sunlight beam reaches the maximum for twice and the minimum for three times. For building B, the width of the sunlight beam reaches the maximum for three times and the minimum for twice. Therefore, the average width of sunlight beam striking on building A is less than that on building B, and it follows that the building A gains less solar energy than building B.

[BLOCK_CHOICE]

16. The four city blocks below are located in Boston, Massachusetts. All of them are 40 meters long and 40 meters wide. There is an existing high-rise building on one of them. You will build another high-rise building similar to the existing one and leave the remaining blocks as open space. You want this new building to have the optimal solar energy gains—as high as possible in the winter and as low as possible in the summer. Which block would you choose for the new building?



a) Block 1
b) Block 2
c) Block 3
d) Any one of them
e) I am not sure

Please explain your choice:

____

____

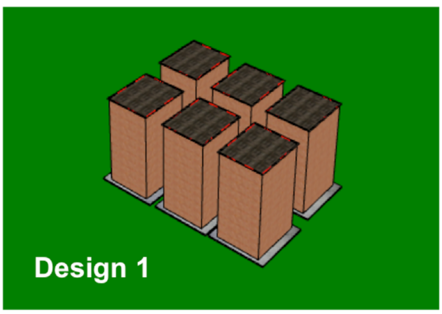
____

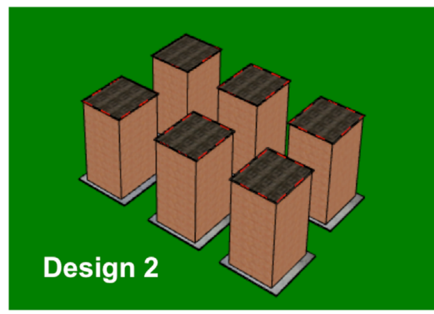
____

____

Answer key: A. building built on block 1 will have the optimal solar energy gains. The building on block 1 receives most sunlight on the south wall, which maximizes its solar energy gain in the winter. Meanwhile, the building on block 1 has the west wall completely shaded by the existing building, which minimizes its solar energy gain in the summer.

[URBAN_DENSITY_TOTAL]
17. Below are two urban designs made for a block in the New York City. The six residential buildings in the two designs are identical except that they are closer to each other in Design 1 than in Design 2. How do the total solar energy gains of the two designs compare?





a) The total solar energy gains of the two designs are the same.
b) The total solar energy gain of Design 1 is greater than that of Design 2.
c) The total solar energy gain of Design 1 is less than that of Design 2.
d) I am not sure.

Please explain your choice:

Answer key: C. The total solar energy gain of Design 1 is less than that of Design 2. With greater distance among buildings, more building surface area can receive sunlight, resulting in an increase in the total solar energy gain.

[URBAN_DENSITY_UNEVEN]
18. For the two urban designs described in the last question, which of the following statements correctly describes how solar energy gains are distributed among the six buildings in the two designs?

a) Solar energy gains are evenly distributed among the buildings within each design.
b) Solar energy gains are unevenly distributed among the buildings within each design, but the degrees of unevenness in the two designs are the same.
c) Solar energy gains are unevenly distributed among the buildings within each design, and the degree of unevenness in Design 1 is greater than that in Design 2.
d) Solar energy gains are unevenly distributed among the buildings within each design, and the degree of unevenness in Design 1 is less than that in Design 2.
e) I am not sure.

Please explain your choice:

____

____

____

____

____

Answer key: C. Solar energy gains are unevenly distributed among the buildings within each design, and the degree of unevenness in Design 1 is greater than that in Design 2. In a cluster of buildings, each building creates shades for others. Depending on their positions within the block, each building has different amount of surface area being shaded. Thus, the solar energy gains are unevenly distributed among them. Greater distance among the buildings reduces the difference in surface area.

Appendix 2 Regression Diagnoses

Linearity. The scatterplots (Figure A1) below suggested that Manipulation and Note were linearly associated with nGain but Simulation appeared not. The partial regression plots (Figure A2) below suggested the similar patterns. The nonlinearity between Simulation and nGain may be due to the highly skewed distribution of Simulation. Therefore, Simulation was transformed using base-2 logarithm. The base-2 log-transformed Simulation appeared more linearly associated with nGain (Figure A1 and Figure A3).

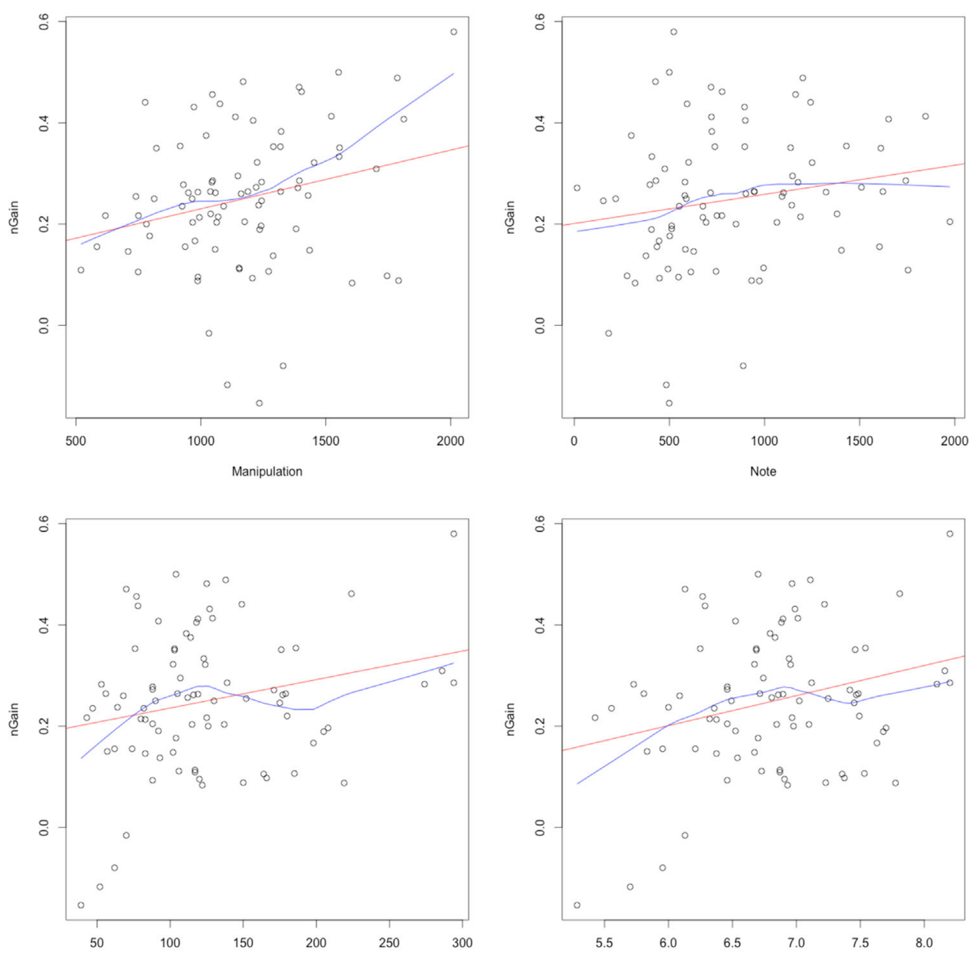


Figure A1. Scatterplots of nGain versus Manipulation, Note, Simulation, and Log2(Simulation). [Color figure can be viewed at wileyonlinelibrary.com].

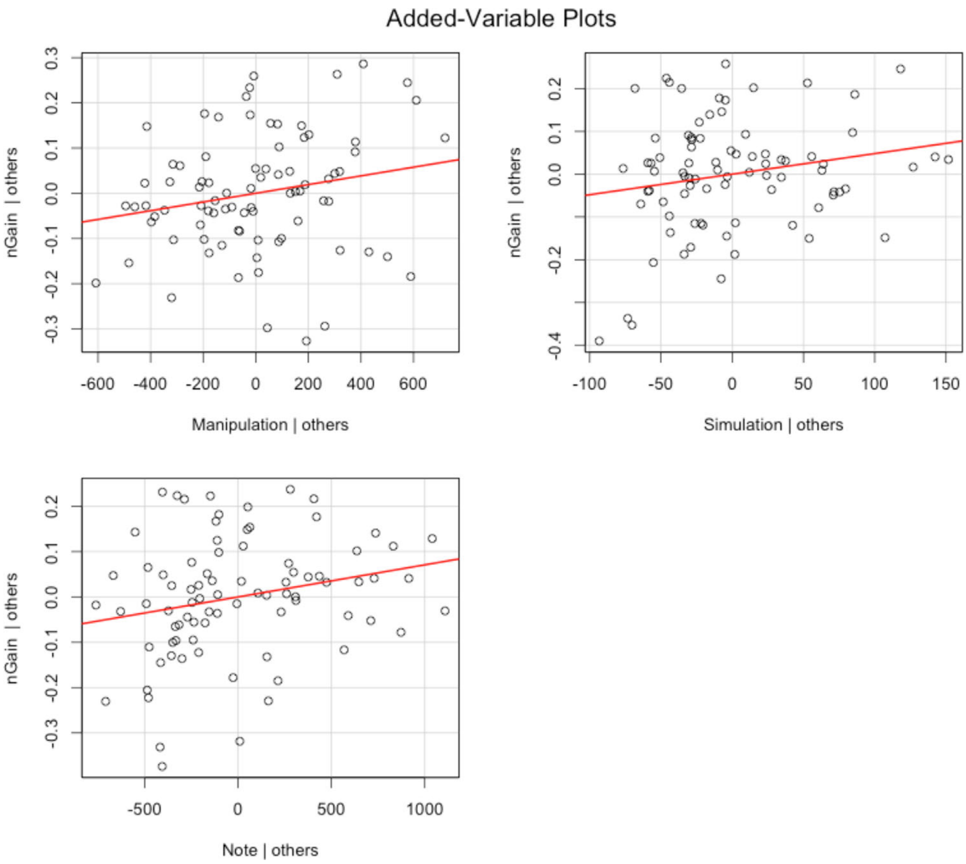


Figure A2. Partial regression plots of nGain versus Manipulation, Simulation, and Note. [Color figure can be viewed at wileyonlinelibrary.com].

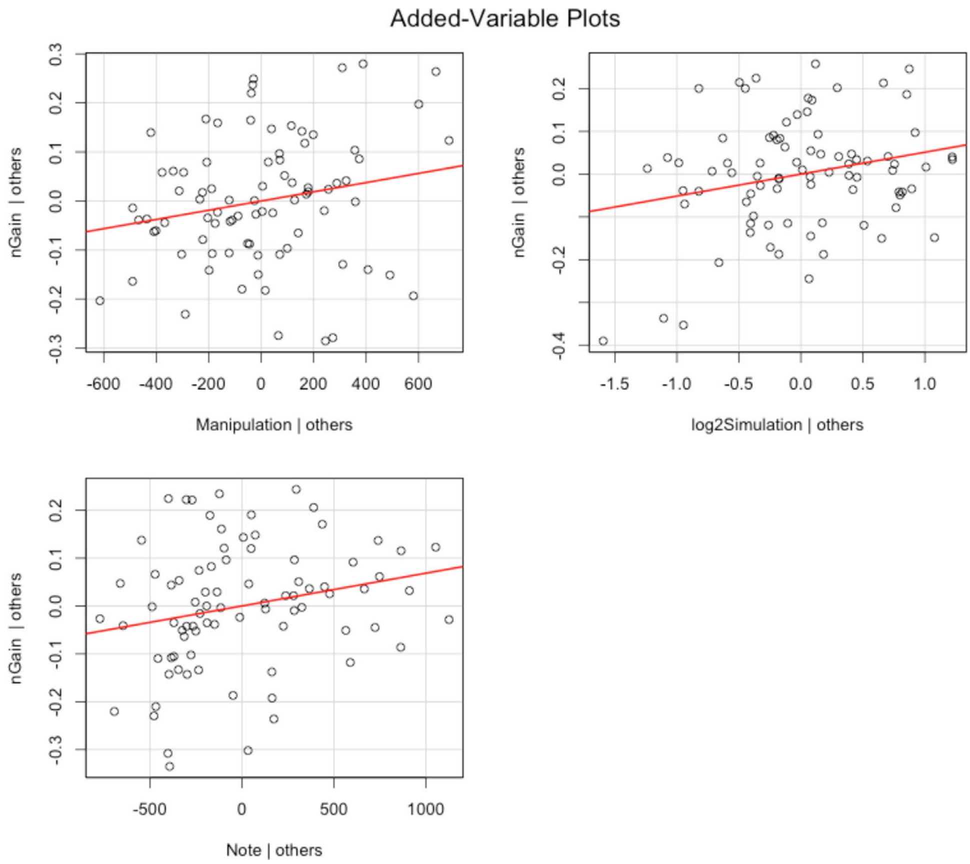
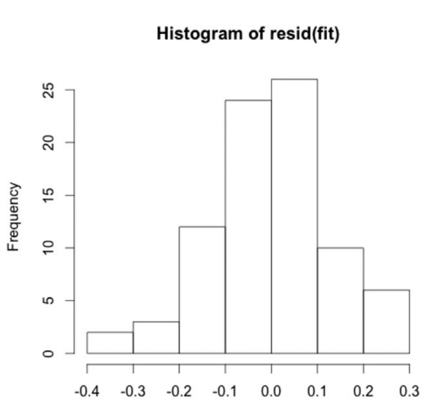
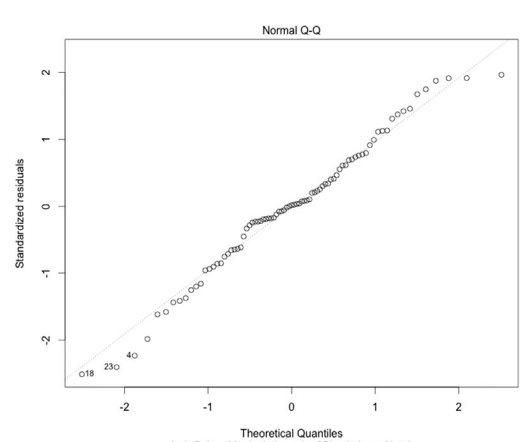


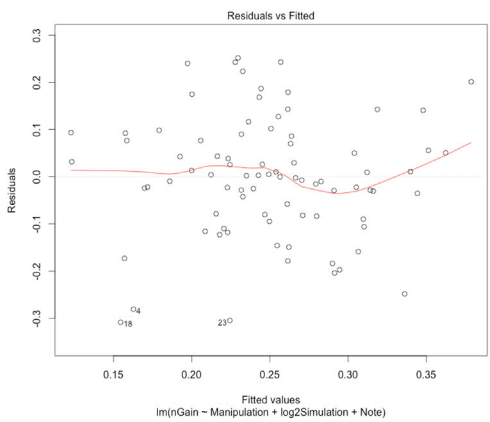
Figure A3. Partial regression plots of nGain versus Manipulation, log2(Simulation), and Note. [Color figure can be viewed at wileyonlinelibrary.com].

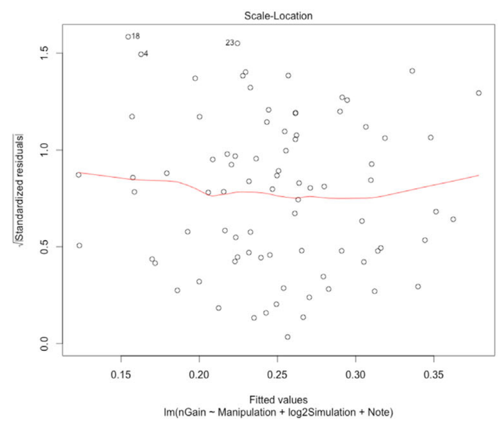
Normality of Residuals. A histogram of the residuals suggested that they were close to being normally distributed. A QQ plot of the standardized residuals showed some slight deviation from normality, but overall there did not appear to be a severe problem with the normality of residuals.



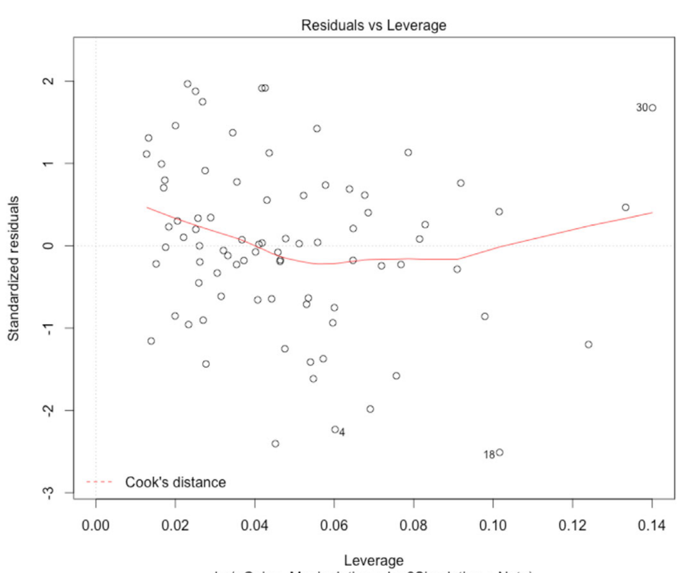


Homoscedasticity. The scatterplots of fitted values versus residuals and square root of standardized residuals showed that the residuals were not distributed in any distinctive pattern with the fitted values. This suggested that the model did not violate the assumption of homoscedasticity.





Multicollinearity was not a concern as the variance inflation factors (VIFs) were low for all three predictors: $VIF_{Construction} = 1.083$, $VIF_{log2(Analysis)} = 1.083$, $VIF_{Note\text{-}Taking} = 1.010$. **Highly influential cases**. Seven influential cases were identified using the Cook's distances with the conventional cutoff value ($4/n = 4/83 = .048$). Their impacts on the overall model fit and the weights of the three predictors were examined in the main text.



Influential Cases. To check whether data points from a few students significantly influenced the regression analysis results, Cook's distances were calculated for all cases and the conventional cutoff value ($4/n = 4/$

83 = .048) was used to identify influential cases. Seven influential cases were identified using the. Table 7 shows the regression analysis results with each case removed from the sample. For all seven reduced samples, the regression model was consistently significant. In other words, none of the cases had any significant impact on the joint effect of the three design operation measures on science learning gain. However, the weights of the three predictors were unstable across the reduced samples. For instance, when Case 18 was removed from the sample, Manipulation became a statistically significant predictor with increased weight while Simulation and Note became statistically not significant with reduced weights. We note that, while the controlled values of some coefficients did fluctuate when various influential cases were removed, the global F-statistic remained statistically significant. This indicates that, taken as a whole, our three predictors jointly predicted variation in the outcome, even when influential cases were removed. We therefore choose to retain all cases in evaluating our final model.

Regression analysis results with influential cases removed from the sample.

					Manipulation		Log2 (Simulation)		Note	
Case no.	Cook's distance	R^2	F	p-value	β	p-value	β	p-value	β	p-value
18	.175	.131	3.911	.012	.241	.032	.158	.157	.199	.065
30	.116	.122	3.623	.017	.148	.177	.208	.059	.233	.032
4	.082	.132	3.966	.011	.217	.052	.193	.083	.200	.063
41	.072	.183	5.815	.001	.246	.023	.241	.026	.228	.030
23	.069	.157	4.853	.004	.237	.032	.197	.074	.227	.033
1	.050	.159	4.919	.004	.164	.134	.251	.023	.241	.024
70	.050	.169	5.279	.002	.231	.034	.244	.026	.195	.063