Journal of the Learning Sciences
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t775653672

Design and Reflection Help Students Develop Scientific Abilities: Learning in Introductory Physics Laboratories

Eugenia Etkina [a]; Anna Karelina [a]; Maria Ruibal-Villasenor [a]; David Rosengrant [b]; Rebecca Jordan [c]; Cindy E. Hmelo-Silver [d]
[a] Department of Learning and Teaching, Rutgers University, [b] Department of Biology and Physics, Kennesaw State University, [c] Department of Ecology, Evolution, and Natural Resources, Rutgers University, [d] Graduate School of Education, Rutgers University,

Online publication date: 25 January 2010

To cite this Article Etkina, Eugenia, Karelina, Anna, Ruibal-Villasenor, Maria, Rosengrant, David, Jordan, Rebecca and Hmelo-Silver, Cindy E.(2010) 'Design and Reflection Help Students Develop Scientific Abilities: Learning in Introductory Physics Laboratories', Journal of the Learning Sciences, 19: 1, 54 — 98

To link to this Article: DOI: 10.1080/10508400903452876

URL: http://dx.doi.org/10.1080/10508400903452876

THE JOURNAL OF THE LEARNING SCIENCES, 19: 54–98, 2010

ISSN: 1050-8406 print / 1532-7809 online
DOI: 10.1080/10508400903452876

Design and Reflection Help Students Develop Scientific Abilities: Learning in Introductory Physics Laboratories

Eugenia Etkina, Anna Karelina, and Maria Ruibal-Villasenor
Department of Learning and Teaching
Rutgers University

David Rosengrant
Department of Biology and Physics
Kennesaw State University

Rebecca Jordan
Department of Ecology, Evolution, and Natural Resources
Rutgers University

Cindy E. Hmelo-Silver
Graduate School of Education
Rutgers University

Design activities, when embedded in an inquiry cycle and appropriately scaffolded and supplemented with reflection, can promote the development of the habits of mind (scientific abilities) that are an important part of scientific practice. Through the Investigative Science Learning Environment (*ISLE*), students construct physics knowledge by engaging in inquiry cycles that replicate the approach used by physicists to construct knowledge. A significant portion of student learning occurs in *ISLE* instructional labs where students design their own experiments. The labs provide an environment for cognitive apprenticeship enhanced by formative assessment. As a result, students develop interpretive knowing that helps them approach new problems as scientists. This article describes a classroom study in which the students in the *ISLE* design lab performed equally well on traditional exams as *ISLE* students who did not engage in de-

Correspondence should be addressed to Eugenia Etkina, Graduate School of Education, Rutgers University, 10 Seminary Place, New Brunswick, NJ 08901. E-mail: eugenia.etkina@gse.rutgers.edu

sign activities. However, the design group significantly outperformed the non-design group while working on novel experimental tasks (in physics and biology), demonstrating the application of scientific abilities to an inquiry task in a novel content domain. This research shows that a learning environment that integrates cognitive apprenticeship and formative assessment in a series of conceptual design tasks provides a rich context for helping students build scientific habits of mind.

Designing environments to guide learners in the construction of knowledge is a challenge for instructors of all disciplines (Bransford, Brown, & Cooking, 1999; Brown, 1992; Derry, Seymour, Steinkuehler, Lee, & Siegel, 2004). It is particularly difficult when educators try to help learners develop both flexible knowledge and habits of mind that can be applied to a range of situations. Such environments need to help provide an apprenticeship in thinking (Collins, Brown, & Newman, 1989) and formative assessment (Black & Wiliam, 1998; Cowie & Bell, 1999). The goal of these environments is to prepare students to go beyond direct application transfer and, as Bransford and Schwartz (1999) argued, to prepare students for future learning.

In this article we investigate whether engaging students in experimental design in such an environment affects student development of scientific abilities. As in Bybee (2000), we use the term *scientific abilities* to describe some of the most important procedures, processes, and methods that scientists use when constructing knowledge and solving experimental problems. For scientists, these abilities are internalized and become habits of mind used to approach new problems; they are scientists' cognitive tools. For the students who have not internalized these processes and procedures, scientific abilities are processes that they need to use reflectively and critically (Salomon & Perkins, 1989). After students internalize them, these abilities become their habits of mind as well.

Scientific abilities include but are not limited to collecting and analyzing data from experiments; devising hypotheses and explanations and building theories; assessing, testing, and validating hypotheses and theories; and using specialized ways of representing phenomena and of communicating ideas (Duschl, Schweingruber, & Shouse, 2007; Ford & Forman, 2006).

Such abilities are generative in nature and, if the students take ownership of them, should develop interpretive knowing (Schwartz & Martin, 2004). *Interpretive knowing* is a way of framing and perceiving a problem, noticing and paying attention to certain features and ignoring others; in a way, "knowing with," not "knowing what" (Bransford & Schwartz, 1999; Broudy, 1977). Interpretive knowing is different from a traditional understanding of knowing as replicative (recalling information) and applicative (using knowledge to solve problems). Interpretive knowing is necessary for transfer, if transfer is viewed as a process in which individuals have an understanding of (an ability to explain and use flexibly) the acquired resources (both concepts and process) that they use to generate new knowledge, that is, to continue learning (Campione, Shapiro, & Brown, 1995). As it cannot be directly observed, interpretive knowing can be only inferred from students' approaches to new tasks (i.e., transfer tasks).

We explore experimental design because of its opportunity to involve students in genuine scientific tasks that replicate real-world challenges and require innovative solutions to solve problems. This study involves a learning system in which students have the opportunity to learn science by actively engaging in scientific practices (Etkina & Van Heuvelen, 2007). The Investigative Science Learning Environment (*ISLE*) is a complex, multifaceted intervention that synthesizes the principles of cognitive apprenticeship with the methods of formative assessment and seeks to develop interpretive knowing with respect to solving scientific problems. Interpretive knowing affects students' framing of a problem, the specific features they focus on, and the constraints of the problem they perceive. For example, when solving an experimental problem, a scientist needs to decide which features of the problem are relevant and which can be ignored; how to represent the problem in different ways, including through the use of mathematical expressions; how to use available equipment to collect necessary data; how to evaluate the quality of the measurements; and how to make sense of the results.

Experimental design is an essential component of *ISLE* that occurs mainly in instructional labs. In these labs, students generate scientific evidence and explanations while designing and conducting their own experimental investigations. We hypothesize that these processes should help students develop scientific abilities and subsequently interpretive knowing with respect to solving scientific problems. To test this hypothesis, we examined the effect of removing the design component of the *ISLE* system as we addressed the following research questions:

1. How does the need to design their own experiments affect the types of activities in which learners engage?
2. How does designing their own experiments affect students' approaches to experimental inquiry?
3. How does designing their own experiments affect students' development of experimental procedures, processes, and methods (scientific abilities)?
4. Does devising, designing, and conducting their own experiments affect students' acquisition of science concepts?
5. How does engaging in design affect students' development of scientific abilities as evidenced in their performance on transfer tasks?

In the article we first describe *ISLE* and the theoretical foundations of design labs, then provide details of the *ISLE* design labs, and finally describe the intervention and the results of the study.

WHY DESIGN?

One way to help students develop scientific habits of mind is to engage them in designing scientific investigations. Multiple successful demonstrations of the usefulness of design for student learning are based on student design of real experiments (Bell & Linn, 2000; Gallagher, Stepien, Sher, & Workman, 1995; Kolodner, 2002) or virtual experiments (Hmelo, Nagarajan, & Day, 2002; Wilensky & Reisman, 2006). Design helps create rich contexts for learning (Harel & Papert, 1991). Through engaging in design, learners become more accountable for their learning through planning, evaluating, revising activities, and reflecting on the process (Hmelo, Holton, & Kolodner, 2000). By means of this process, learners construct meaning and internalize the knowledge they created (Kafai & Resnick, 1996). Even when students engage in design activities with an engineering focus (i.e., with the aim of obtaining improved outcomes) instead of with the goal of construction or refinement of knowledge, they may progress in the direction of the development of an "empirical attitude" in knowledge construction situations (Apedoe & Ford, 2009). Design requires students to engage in metacognitive thinking, which, in turn, might lead to a productive scientific discourse (Davidowitz & Rollnick, 2003; Driver, Asoko, Leach, Scott, & Mortimer, 1994; Germann, 1989; Gourgey, 1998; Hofstein, Shore, & Kipnis, 2004). Students engage in metacognitive thinking (Campione et al., 1995) such as planning, monitoring, and evaluating in order to manage investigative tasks. To communicate with one another and with the instructors, they need to become fluent in scientific discourse. Such communication includes the use of scientific language and representations as well as scientific ways of using that language, thinking, evaluating, acting, and interacting. These are the behaviors that identify individuals as members of the socially meaningful group of scientists (Gee, 1999).

Designing and carrying out investigations also helps engage learners in important cognitive and social activities that promote the development of interpretive knowing. Perkins (1986) defended the view of knowledge itself as design: In this view, knowledge is a structure that has to be built and adapted for specific purposes. In addition, the social component of the design process, which is usually carried out in groups, might promote the dispositions to think critically and creatively (Bereiter, 1995).

Design activities also help learners activate prior knowledge and notice relevant features of phenomena or processes that enable them to take advantage of learning opportunities and prepare them for future learning (Hmelo et al., 2002; Schwartz & Martin, 2004).

ISLE

ISLE (Etkina & Van Heuvelen, 2007) is a learning system that was developed for large- or small-enrollment introductory, non-major, college physics courses that follow a traditional structure of lectures (we call them *large-room meetings*), problem-solving sessions, and labs. The goal of *ISLE* is not only to help students learn fundamental physics concepts but most important to help them learn to approach problems as scientists by engaging them in processes similar to those that scientists use while constructing and applying new knowledge. Thus, in this spirit *ISLE* is similar to many other approaches that engage students in design and authentic problem solving (Barron et al., 1998; Hmelo et al., 2000; Hmelo-Silver, 2004; Kolodner et al., 2003; Merrill, 2002). The main difference is that *ISLE* does this within the traditional structure of a physics course; therefore, students work on smaller problems that can be solved during an 80-min problem-solving session or a 3-hr instructional lab. However, problem types that students encounter are repeated throughout the semester. This allows students to develop experimental approaches relevant to scientific and engineering design (Kolodner, 2002).

ISLE engages students in collaborative knowledge construction to acculturate them in the practice of physics. The goal of the environment is not only to help students learn the concepts and laws of physics but most important to help them learn *how* this knowledge is constructed. To learn a new concept, students go through a cycle that repeats multiple times during the semester and takes place in all settings: large-room meetings, problem-solving recitations, and instructional labs. They first observe a series of carefully selected experiments; then they use available tools (such as motion diagrams, force diagrams, energy bar charts, ray diagrams) to analyze the data to find patterns; then, when possible, they devise explanations or mechanisms for the patterns. Later they test the explanations by using them to predict the outcomes of new experiments with the goal of ruling out the explanation instead of proving it, and finally they apply new knowledge to solve practical problems (see Figure 1). For further details, see Etkina and Van Heuvelen (2007). The *ISLE* cycle is a blend of Karplus's (1977) learning cycle and Lawson's (2002) science cycles.

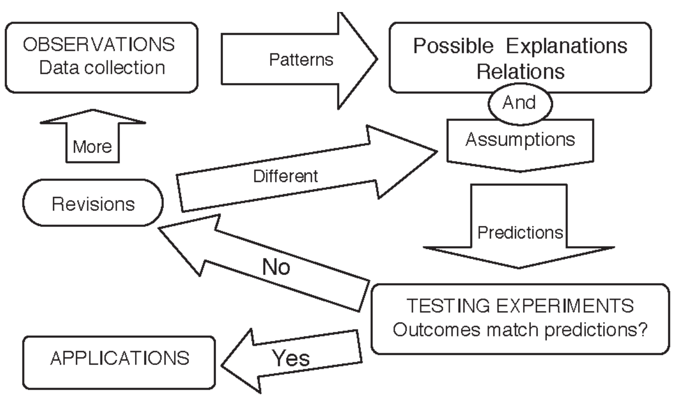


FIGURE 1 Investigative Science Learning Environment cycle: students' activities that emulate the processes of scientific knowledge construction.

One of the most important features of *ISLE* is that students learn collaboratively in all settings. Even in large-room meetings students do not listen passively to a lecture but work in groups to come to a consensus of what they observed, how to analyze and represent the observations, how to explain them, or how to test the explanations. They work on activities in the Physics Active Learning Guide (Van Heuvelen & Etkina, 2006) and use electronic clickers to express their opinions and often call out to present the ideas of their group. Some *ISLE* instructors use a short paper method to obtain formative feedback from the students during the large-room meeting. Group work continues in problem-solving sessions and instructional labs. In this article we focus mostly on the *ISLE* laboratory environment (Etkina, Murthy, & Zou, 2006; see an example of a handout for an *ISLE* design lab in Appendix A).

THEORETICAL FOUNDATIONS

The development of interpretive knowing with respect to solving experimental problems is the goal of the labs. In response to this goal, *ISLE* labs closely integrate cognitive apprenticeship and formative assessment. To help students design their own experiments and be reflective about their work (which is a component of successful design), we provide them with expert models of such activities and scaffolding using the model of cognitive apprenticeship. Formative assessment helps us make this expertise visible and provide feedback to the students. Here we describe the theoretical foundations of interpretive knowing, cognitive apprenticeship, and formative assessment (see Figure 2).

Interpretive Knowing

Helping students develop interpretive knowing (Schwartz, Bransford, & Sears, 2005) is one of the goals of *ISLE* labs. Interpretive knowing empowers students so that they can approach new phenomena and problems similar to how a scientist would approach them. For example, in one lab students need to determine the specific heat of a particular object. When confronted with such a challenge, students normally search for physics equations that are relevant to the question and then try to conduct a single experiment to measure relevant quantities. In contrast, a scientist, after thinking about relevant physics ideas and laws, immediately realizes that one experiment will not be sufficient (i.e., different experiments that will produce similar results are needed). Furthermore, a scientist knows that the physics laws that he or she will use to analyze collected data are mathematical models of some ideal situation, and he or she will need to make sure that the experimental setup satisfies the mathematical model criteria. Finally, a scientist keeps in mind that any result that he or she obtains has experimental uncertainties because of the instruments and procedures used (Alberts, 2000). In summary, a scientist would notice

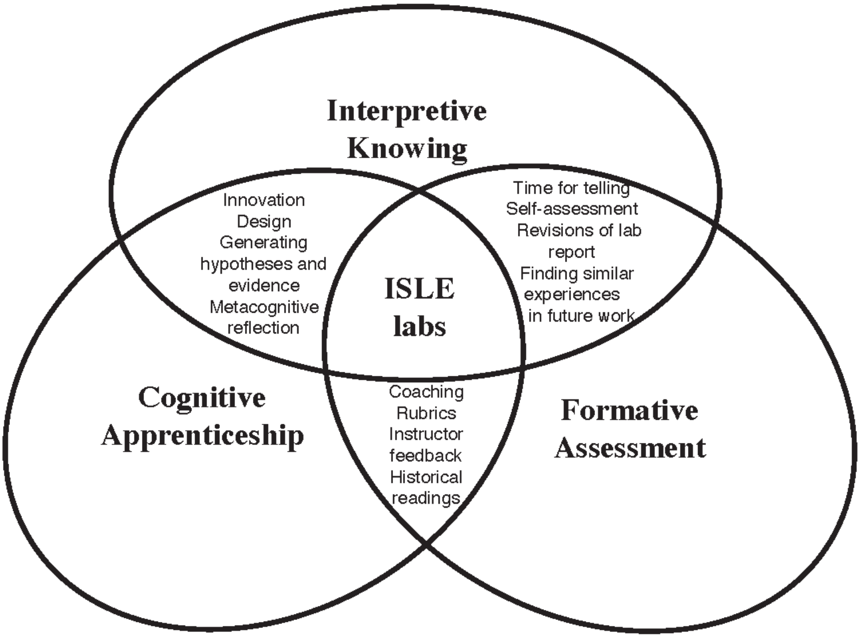


FIGURE 2 Theoretical foundations of Investigative Science Learning Environment (*ISLE*) labs.

that this task has particular features about which he or she needs to think while solving the problem.

The goal of *ISLE* labs is to help students learn to interpret experimental and conceptual problems in ways similar to how scientists would. We would like students to think of relevant physics principles, assumptions in the mathematical procedure, uncertainties in the experimental results, the need to confirm the results with an independent method, and so on, when faced with an experimental problem. Marton (2006) argued that these idiosyncratic, highly specialized, and finely attuned ways of interpreting certain situations are the invariable signal of expertise and that knowing anything requires the development of particular ways of perceiving the world for certain purposes. The work of Goodwin (1994) has shown that scientists have their own distinctive ways of looking at and making sense of the world. This "scientific glance" or "scientific framing" is what enables them to construct scientific knowledge; in other words, it is what prepares them for their science learning. We may conclude that the development of specific interpretive knowing not only helps students pay attention to certain relevant features of a particular problem but also assists them in the development of a whole new "scientific" way of looking at the world.

To be specific, in our study the development and assessment of interpretive knowing is done through the development and assessment of scientific abilities mentioned previously (Etkina, Van Heuvelen, et al., 2006). The scientific abilities that we help our students develop include the ability to represent physical processes in multiple ways; the ability to devise and test a qualitative explanation or quantitative relationship; the ability to design an experimental investigation; the ability to collect and analyze data; the ability to evaluate claims, solutions, and models; and the ability to communicate (see http://paer.rutgers.edu/scientificabilities for a complete list). The list of abilities was constructed based on the analysis of the history of the practice of physics (Holton & Brush, 2001; Lawson, 2000, 2003), research on the goals of science curricula and taxonomies of cognitive skills (Bybee & DeBoer, 1994; Marzano, 2001), recommendations of cognitive scientists (Schunn & Anderson, 1999), and an analysis of science process test items (National Assessment Governing Board, 2005). Some of the abilities are similar to science skills and practices described by Kolodner (2002). Each of the scientific abilities involves many sub-abilities. For example, the ability to collect and analyze data includes the following sub-abilities: (a) identify sources of experimental uncertainty, (b) evaluate how experimental uncertainties might affect the data, (c) minimize experimental uncertainty, (d) record and represent data in a meaningful way, and (e) analyze data appropriately.

Cognitive Apprenticeship

The *ISLE* system regards learning as a cognitive apprenticeship of physics practice. In this approach, learning is supported by means of modeling, coaching, and scaffolding (J. S. Brown, Collins, & Duguid, 1989; Collins et al., 1989). *ISLE* labs provide models of scientific inquiry as students read case studies of scientific developments (e.g., students read and reflect on how scientists learned about pulsars, medical prophylaxis, the nature of AIDS). In addition, other forms of modeling are provided as instructors demonstrate hypothetico-deductive reasoning, construct representations, devise mathematical procedures, and demonstrate other scientific practices in the labs. Additional modeling is provided by formative assessment rubrics (described later). Coaching is achieved by (a) the careful selection and organization of the tasks students have to accomplish, (b) the structuring of the tasks by means of prompts and questions given to the students on the lab handouts, (c) instructor feedback, and (d) the breaking of the assigned tasks into subtasks through lab handout hints and questions and by the scientific abilities rubrics.

As students work in teams on different laboratory tasks, the instructors and course materials provide scaffolding. *ISLE* lab handouts do not have any explicit instructions in terms of experimental procedures, but they have a set of questions that focus students' attention on the important aspects of the design process and simultaneously help make visible their thinking about the salient elements of design (see Appendix A). Because we are introducing students to the practices of real science, we need to provide support to enable them to accomplish unfamiliar, complex tasks and, at the same time, learn from the experience. This support is provided through lab handout questions and self-assessment rubrics. Moreover, the system demands reflection from the learners, as they need to be highly metacognitive to complete their design tasks. They have to articulate and refine their ideas when discussing with their partners, answering the reflection questions given in the laboratory handouts, and writing their individual lab reports (samples of student writing are discussed later). *ISLE* labs have embedded reflection and connection-building activities. These activities are critical for transfer, as both direct application and preparation for future learning perspectives suggest (Bransford & Schwartz, 1999).

In addition, the whole *ISLE* system provides scaffolding for students' inquiry activities by introducing the processes of physics in a structured and simplified manner. The *ISLE* cycle (see Figure 1) helps communicate the process of scientific inquiry in physics and gives structure to the series of tasks that students have to accomplish. Students learn to differentiate between observational experiments, testing experiments, and application experiments in physics (Etkina, Van Heuvelen, Brookes, & Mills, 2002), and they learn to conceptualize lab experiments as different variations of these three. This procedural facilitation is essential, as one of the main difficulties students face when they are engaged in the building of new knowledge is that they do not know how the tasks can play a part in their overall insight of a phenomenon. The investigative cycle helps students know where they are in the inquiry process and helps them perceive their learning as an integrated process and better understand how and why each step or task is important (Schwartz, Brophy, Lin, & Bransford, 1999).

Formative Assessment

During their lab work, students use specially designed assessment rubrics that help them organize and revise their work as they progress (Etkina, Van Heuvelen, et al., 2006; see an example of a rubric in Figure 3 and a full set of rubrics at http://paer.rutgers.edu/ScientificAbilities/Rubrics/default.aspx). These rubrics serve two purposes simultaneously. They not only provide the modeling and coaching aspect of the cognitive apprenticeship, but they also simultaneously engage students in self-assessment, which is the most powerful form of formative assessment (Black & Wiliam, 1998; Cowie & Bell, 1999).

The goal of the rubrics is to help students develop scientific abilities through self-formative assessment. Rubrics serve as tools that help students develop these abilities by making expert practices explicit and thus promoting interpretive knowing. They are tools for procedural facilitation and feedback as they help novices to complete inquiry tasks that require complex and unfamiliar scientific abilities. At the be-

Scientific Ability	Missing (0)	Inadequate (1)	Needs some improvement (2)	Adequate (3)
Is able to evaluate the results by means of an independent method	No attempt is made to evaluate the consistency of the result using an independent method.	A second independent method is used to evaluate the results. However there is little or no discussion about the differences in the results due to the two methods.	A second independent method is used to evaluate the results. The results of the two methods are compared using experimental uncertainties. But there is little or no discussion of the possible reasons for the differences when the results are different.	A second independent method is used to evaluate the results and the evaluation is done with the experimental uncertainties. The discrepancy between the results of the two methods, and possible reasons are discussed.

FIGURE 3 An example of a rubric for one sub-ability.

ginning of the semester, students are told which rubrics are appropriate for each task. Toward the end of the semester, however, this scaffolding fades, and students need to decide which rubrics are appropriate. As students work in the lab designing and conducting experiments and writing a report, they use the rubrics for guidance. The instructor assesses their lab reports based on the same rubrics and holds a discussion in a subsequent lab about students' successes and weaknesses. Using the rubrics helps prepare students to learn from the instructor's explanations and lab discussions (Bransford & Schwartz, 1999; Schwartz & Martin, 2004). At the same time, the rubrics provide an opportunity for instructors to make their thinking visible for the students (Collins, Brown, & Holum, 1991; Scardamalia & Bereiter, 1985). After students receive the formative assessment feedback, they can reflect on what they learned and revise their work (Black & Wiliam, 1998; Cowie & Bell, 1999).

In conclusion, *ISLE* laboratories are generative in nature. Students produce a variety of cognitive outcomes encompassing ideas, procedures, and artifacts, as shown in Figure 4. The creation of these sharable products greatly facilitates the process of learners' knowledge construction (Harel & Papert, 1991).

AN INVESTIGATION OF THE ROLE OF DESIGN

To investigate the causal importance of design and reflection, we created an alternative set of labs in which students did not design their own experiments (and con-

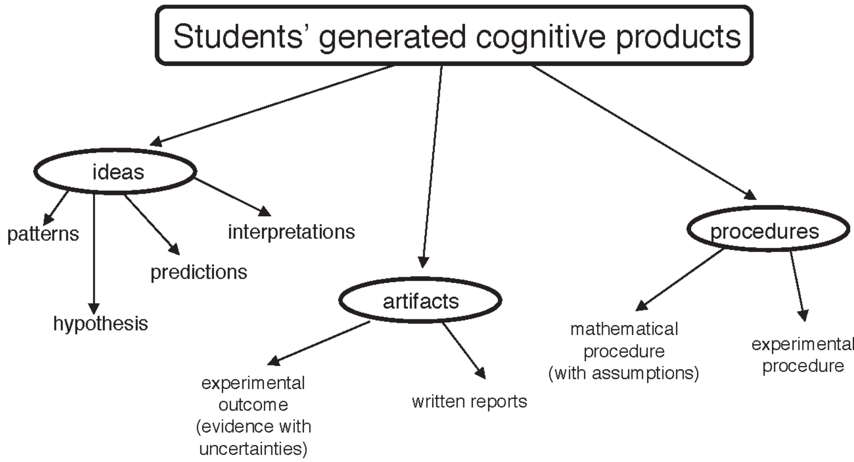


FIGURE 4 Why Investigative Science Learning Environment labs are generative: Different types of products.

sequently did not reflect on their design procedure) to address whether design and reflection affect students' development of scientific abilities. In this non-design alternative, students generated a smaller set of cognitive products as they did not devise their own experimental procedures. They did not use rubrics to self-assess and did not have to engage in constant metacognitive reflection. Instead, they followed a set of clear directions to perform a desired experiment. However, they did use multiple representations and the theoretical introduction in the handout to devise their own mathematical procedure to determine the quantity sought. They repeated the experiment several times to determine the uncertainty in the value of the quantity sought (see Appendix B). Basically, the non-design, non-reflective labs were traditional introductory science college labs with enhanced conceptual and quantitative reasoning. The purpose of having some of the students perform non-design labs was to have a baseline of student approach to experimental tasks when they do not engage in design and reflection on a regular basis.

Participants

A total of 186 students enrolled in an algebra-based physics course followed the *ISLE* curriculum. (The number of students attending different activities varied throughout the semester.) Of these students, 72 male students and 114 female students signed up for different sections of the course. As the course is a requirement for science majors (e.g., biology, ecology, pre-veterinarian, and exercise science majors), students range in academic year from sophomores to seniors. The former are those who fulfill the requirement as soon as they can, and the latter are those who postpone taking physics until their last year in college. During the semester, students were not informed about the study. At the end of the semester, we disclosed the procedure, and students signed a consent form allowing us to use their work for research.

Instructional Context

This study was conducted in the first semester of the course. There were two 55-min lectures or large-room meetings, one 80-min problem-solving session, and a 3-hr lab per week. There were two midterm exams, one final exam, and two lab exams. All students participated through the same *ISLE* curriculum in large-room meetings and in smaller problem-solving sections but had different lab experiences (see below). There were eight problem-solving sections and eight lab sections. Students signed up for lab sections and problem-solving sections separately; thus, students who were in the same lab section were not necessarily in the same problem-solving section. There were three instructors in the lab sections: one instructor taught four sections and the other two taught two sections each. All three instructors were highly skilled in interactive teaching, and each had from 4 to 7 years of teaching experience. These instructors were also active members of the Rutgers physics education research group and, prior to this study, had helped develop and assess different aspects of *ISLE* labs. Each had had coursework in the science of learning and interactive physics teaching methods, and was observed multiple times while teaching.

In four lab sections the students performed design labs supplemented with reflection questions and rubrics, and in the other four sections students performed non-design labs. Hereafter we refer to these as the design and non-design sections, respectively, keeping in mind that the difference between the sections was more than just the presence or absence of design but also the presence or absence of reflection and rubrics. Two of the instructors taught one design and one non-design section each, and the third instructor taught two of each. Our decision as to which sections were design or non-design was based on the pragmatic consideration that we did not want the students from design sections to learn how to perform the experiment from the non-design section students. Thus, the first four sections of the week were assigned to the design category and the other four to non-design. As there is no particular system according to which students sign up for lab sections, we assumed that they would be distributed among the sections randomly.

To ensure that the students' design and non-design lab sections were equivalent in scientific reasoning ability at the beginning of the course, we administered Lawson's test of scientific reasoning (Lawson, 1978) in the first lab session. Students in the two groups did not differ: design (M = 11.04, SD = 3.83), non-design (M = 10.95, SD = 2.96), $F(1, 185) = 0.06$, $p = .81$.

Instructional Procedures

The only difference in the treatment of the groups was the presence or absence of the experimental design and reflection in the laboratories. The material covered in the labs, the lab equipment, the weekly topics, and the instructors were the same.

Design labs. Students in the experimental group had to design their own experiments. Two thirds of the experiments were based on the content that students had already discussed in large-room meetings and problem-solving sessions. In one third of the labs, students designed an experiment to investigate a new phenomenon and find patterns. The scaffolding was provided through lab handout questions that focused students' attention on the elements of the scientific process (or on specific scientific abilities): representing the situation, deciding on the experiment, analyzing experimental uncertainties, and so on.

Although we provided scaffolding, students struggled first to generate possible designs and subsequently to actually implement their designs and to evaluate the results. They had to use different representations, such as force diagrams and energy bar charts, to help devise the mathematical procedure. After they implemented the procedure, they had to figure out whether the result made sense. For example, if the goal of the experiment was to determine the coefficient of friction between their shoe and the floor, and they obtained a certain value through the experiment and calculation, they could not ask the lab instructor whether the value was acceptable. Instead, they had to design an independent experiment to determine the same value and then make a judgment about the results of both experiments.

Students had to compose individual written lab reports to describe what they did and to answer the questions in the lab handout. They could consult relevant self-assessment rubrics to improve their reports as they were writing or at the end. Appendix A provides an example of a lab handout.

The lab instructors did not tell students how to design the experiments, and when students had difficulties, the lab instructors asked questions and provided hints but did not answer the students' questions directly. Usually a lab had two experiments: one that the students had to design from scratch by choosing the equipment, putting it together, and devising the mathematical procedure; and the other one for which the experimental setup was provided but students had to invent their own procedure and decide what measurements to record. Students wrote their reports during the 3-hr lab period and handed them in before they left.

After each experiment, students were asked to reflect on its purpose and on its place in an overall scientific process. Reflection questions focused students' attention on contrasting cases, such as the difference between an observational experiment and testing experiment (Etkina, Murthy et al., 2006). Students also had to reflect on the difference between experimental uncertainties and theoretical assumptions and explain why they had to design two independent experiments to determine the value of a particular quantity.

At the end of each lab, students were asked to address questions that encouraged drawing connections between in-lab practices and out-of-classroom experiences. Lab homework, which students did after each lab, contained reading passages with reflection questions. Students analyzed stories about the historical development of several scientific theories and applications, such as the nature of AIDS, prophylaxis, or pulsars. They had to identify the elements of scientific inquiry that were present when scientists answered new questions or applied knowledge. The purpose of the passages was to help students reflect on the common elements of scientific investigations.

Non-design labs. Students in the non-design labs used the same equipment as those in design labs and performed the same number of experiments. The handouts guided them through the experimental procedure but not through the mathematics. Students had to draw force diagrams, energy bar charts, and other representations to solve experimental problems. The lab handout provided a list of assumptions that students should use when applying a theoretical model to a particular experimental situation. Students were instructed to repeat the measurement several times to determine the uncertainty and to incorporate the uncertainty into the final result.

These non-design labs had homework as well. About 30% of the time the students had to read about experimental uncertainties or assumptions and do some simple exercises. These assignments were similar to those of the design students. About 70% of the time, however, the homework was different than that of the design students. In these assignments, non-design students had to solve traditional physics problems that prepared them to do the next lab versus reflecting on science processes and strategies. Thus, we can say that the non-design labs were more or less traditional, introductory science labs akin to those offered to undergraduates nationwide.

The instructors taught non-design labs differently. They provided an overview of the material at the beginning of the lab and then later, if students had questions, they answered those questions. In Appendix B we provide an example of a lab handout for the non-design group. This handout was used during the same week that the design students performed the experiment in Appendix A. Notice that although the handout provides students with instructions on how to do the experiment, it does not tell them how to solve the physics problem, as in this example (excerpt from Appendix B):

> h) Construct a work-energy bar chart for the process starting with the car resting on the stretched launcher and ending when the car is at its maximum elevation.
> i) Apply the generalized work-energy equation for the process.

> …
> k) Calculate the fractional uncertainty of the elastic potential energy for each launching position—equal to the fractional uncertainty of the vertical distance traveled times the elastic energy for that launching position.

Table 1 summarizes the similarities and differences between the design and non-design labs in terms of student activities.

Procedure for Interpretive Knowing Test Labs

In addition to the 10 standard laboratory sessions described previously, we developed two lab sessions in which both groups designed an experiment and wrote a lab report. We call these lab sessions *interpretive knowing test labs*. In contrast to the other labs, these particular labs were identical for the design and the non-design groups. We devised these labs to assess how students apply scientific abilities to unfamiliar physics content in the same functional context (according to the classification by Barnett & Ceci, 2002).

The content of the first interpretive knowing test lab involved drag force in fluid dynamics; this concept we assumed would be unfamiliar to the students, as it was not covered in the course. To minimize diffusion of information among the students because of different sections performing the lab on three consecutive days of the week, we developed four similar versions of the lab (see Appendix C). Students were provided some necessary and some redundant information in the lab handout and had access to textbooks and the Internet. There were no scaffolding questions in the handout and no reference to the rubrics.

The students performed this lab in Week 13 of the semester. Prior to this, they had performed 10 labs. The drag force lab was attended by 89 students in design sections and by the same number of students in the non-design sections.

The second interpretive knowing test lab involved an experimental problem in biology. Both the design and the non-design groups had to design an experiment to find the transpiration rate of a certain species of plant and subsequently write a report detailing their experimental procedures, calculations, and conclusions. This particular biology problem was selected because (a) measuring transpiration is a task simple enough for students with very little plant physiology background; and (b) students could use multiple measures to determine the transpiration rate, which gave them some room for inventiveness, evaluation, and decision making.

We provided students with handouts that had definitions of transpiration and humidity. The handout (the same for both groups) also included a table with saturated water vapor density as a function of temperature, as the course did not cover humidity at all. In addition, the students could consult the Internet.

TABLE 1
Features of Student Work in the Design and Non-Design Lab Sections

Feature	*Design Lab*	*Non-Design Lab*
Students work in groups and submit individual lab reports	Yes	Yes
Students conduct physics experiments related to the same physics content of the week	Students design their own experiments.	Students perform prescribed experiments following guidelines.
Students devise their own mathematical procedure and are encouraged to use different representations	Yes, the specific representations (force diagrams, energy bar charts, etc.) are sometimes recommended and are sometimes not.	Yes, the representations are recommended.
Students are aware of the assumptions in the mathematical procedure	Students think of assumptions and evaluate them themselves.	The assumptions are provided.
Students evaluate the uncertainties in the result	Students decide how to evaluate uncertainties and how to represent the result themselves.	The evaluation of uncertainties is a part of the prescribed procedure.
Students reflect on the experience and self-assess their work using rubrics	Yes	No
Students do lab homework	For 50% of the homework students read passages about scientific discoveries and answer reflection questions; the other 50% of the homework involves reading a text describing uncertainties or assumptions with relevant exercises.	All homework is related to the labs. It involves either solving physics problems related to the lab next week or reading a text describing uncertainties or assumptions with relevant exercises.
Lab instructors	At the beginning of the lab, instructors provide a summary of issues related to scientific abilities in the lab reports of the previous week; they do not assist students in experimental design.	At the beginning of the lab, instructors provide a theoretical summary of the lab content; they help students when asked.

Data Collection Procedures for Physics Knowledge and Scientific Abilities

In Table 2 we list the data sources that we used to answer the research questions. Here we describe how we collected the data and how we established the reliability of our instruments.

Observations of student behavior during labs. To observe student behavior during the labs, we tracked the time spent by a group of students on different activities (Karelina & Etkina, 2007). A trained observer sat with a single group of students from each condition for the whole lab period during 10 weeks (twenty 3-hr observations). This observer timed and recorded everything that students did. Students' behaviors were then coded using a coding scheme modified from the work of Lippmann and colleagues (Lippmann & the Physics Education Research Group, 2002; Lippmann Kung, Danielson, & Linder, 2005).

Lippmann's scheme had three codes: sense-making, logistic, and off-task. According to Lippmann, during sense-making episodes, students are talking to one another, working on figuring out the answer, and holding a coherent conversation. During the logistic mode (here, *procedure*) students gather equipment, operate equipment, collect data, read, and write. Off-task mode involves the time intervals when students are not directly engaged in the lab task.

In addition to Lippmann's three-item coding scheme we used a code for writing and for instructor help (see Table 3). Observations of more than 30 lab groups con-

TABLE 2
Data Sources Related to the Research Questions

Research Question	*Data Source*
How does the need to design their own experiments affect the types of activities in which learners engage?	Observations of student behavior during Labs 1–10 of the semester (20 observations, 10 randomly chosen student groups in the design sections and 10 in non-design sections)
How does designing their own experiments affect students' approaches to experimental inquiry?	Observations of student behavior during the physics and biology interpretive knowing test labs (8 design and 8 non-design groups)
How does designing their own experiments affect students' development of experimental procedures, processes, and methods?	Students' rubric scores (for the students in the design group) on relevant abilities during the semester based on the rubrics
Does devising, designing, and conducting their own experiments affect students' acquisition of science concepts?	All students' scores on regular exams that included multiple choice and open-ended questions (2 midterms and 1 final exam)
How does engaging in design affect students' development of interpretive knowing (i.e., student application of scientific abilities in new situations)?	Student rubric scores for physics and biology interpretive knowing test labs

TABLE 3
Codes for Observations of Lab Behaviors

Code	*Description*
Sense-making	Engaging in discussions about physics concepts, experimental design, the mathematics procedure, assumptions, uncertainties in the data, revisions of the experiment, and questions in the handout
Writing	Describing the experiment, recording data, calculating values, and explaining results
Procedure	Gathering equipment, mounting the setup, and taking data
Instructor help	Listening to the instructor, who is explaining and answering student questions (for non-design groups) or providing feedback on the previous week's lab reports and the design/procedure (for design groups)
Off-task	Any activity not related to the laboratory task

ducted over the 2 years prior to this study showed that all behaviors fit into one of the five coding categories. Also, within the sense-making code we noted instances when students discussed the issues of design, the physics concept involved, the mathematical procedure, assumptions inherent in the mathematical procedure, experimental uncertainties, and revisions of the experiment based on the outcome. To establish the interrater reliability for the coding, two observers independently coded four of the first lab observations (20% of the observations). They achieved 84% agreement on the codes before discussion and 100% agreement after discussion. After reliability was established, a single observer monitored student lab groups. The observer timed and recorded one design group and one non-design group each week, observing twenty 3-hr labs.

Observations of student behavior during the physics drag force and bio transpiration interpretive knowing test labs. During the two labs in which all students had to design experiments without any guidance, we observed one group per lab section using the same coding scheme as described in the previous section. In total, we collected observations of four student groups from the design lab sections and four student groups from the non-design sections.

Design students' rubric scores. After the semester was over, three trained raters scored the reports of three design sections (one section per instructor) using the same scientific abilities rubrics that the students used during the labs for self-assessment. All three scorers used the chosen rubrics to independently score the lab reports of two or three students for each lab. Then they discussed any discrepancies in scores to make sure that the details of the particular labs were taken into account. They then scored an additional 7 to 10 randomly chosen lab reports (a total of 22% of lab reports) until they achieved an agreement of more than 85% of the given scores. For many labs, the scorers achieved almost 100% agreement after the second scoring. Following this, each rater scored 15 to 17 additional reports.

Physics knowledge. Regular exams consisted of multiple choice and open-ended questions. Midterms were 55 min and had 11 questions each. Of these 11 questions, 8 were multiple choice questions (equally split between conceptual and quantitative) and 3 were open-ended problems. A 3-hr final exam had 18 questions, 12 of which were multiple choice questions (9 quantitative and 3 conceptual) and 6 of which were open-ended problems. The multiple choice part of the final exam was more equation oriented than the multiple choice parts of the midterms. The grades for the open-ended exam questions were based on a rubric devised by the professor of the course.

Rubric scores for the interpretive knowing test labs. During the drag force lab and the transpiration lab, students in each lab section worked in the same group of three or four as they had done during the semester and submitted individual reports for grading. The four design sections had both labs earlier in the week than the non-design sections. After the semester was over, the researchers used the scientific abilities rubrics to code student work using the same procedure for ensuring the reliability of the scores as described previously. The rubrics chosen for scoring were for the following sub-abilities: the ability to evaluate the effects of assumptions in the mathematical procedure, the ability to evaluate the effects of experimental uncertainties on the result, the ability to evaluate the results by means of an independent method, and the ability to communicate (see an example of a rubric in Figure 3). These rubrics were chosen based on two criteria: They were the most relevant for those labs, and they assessed some abilities that we hoped students would develop in a lab course.

RESULTS

How Does the Need to Design Their Own Experiments Affect the Types of Activities in Which Learners Engage?

The amount of time that students spent on different activities is shown in Table 4. To test for differences among groups for the total time students spent in the labs and for each of the coding categories listed in Table 4, we conducted a two-way analysis of variance (ANOVA) between the design and non-design groups across the coding scheme. Differences between the groups in terms of the total time spent in the lab were noted, $F(1, 180) = 45.16, p < .001$. Differences in the duration of the various activities in which students engaged during the labs were also noted, $F(5, 180) = 40.12, p < .001$; too few observations in the reading category (see Table 4) caused us to eliminate this category from the analysis. Inspection of the data showed that these differences were most evident between sense-making and overall time spent in the lab (see Tables 4 and 5). In general, the design students spent a

TABLE 4
Total Time in Minutes Spent by Students on Different Activities

	Design							Non-Design						
Lab	*SM*	*Wr*	*Pro*	*Rd*	*TA*	*Off*	*Tot*	*SM*	*Wr*	*Pro*	*Rd*	*TA*	*Off*	*Tot*
1	39	53	7		9	12	120	22	48	11		5	0	86
2	26	50	34		58	7	175	30	60	33		5	1	129
3	52	71	22		17	2	164	19	39	37		39	2	136
4	47	71	12		1	0	131	14	57	20		28	1	120
5	33	74	21		13	31	172	14	24	6		11	2	57
6	44	64	20		16	2	146	17	60	17		31	5	130
7	44	93	24	7	11	3	182	12	33	14	6	24	2	91
8	20	60	10	6	10	8	114	4	33	11	7	15	2	72
9	27	63	49	5	31	4	179	6	36	21	0	9	0	72
10	41	65	40	3	12	15	176	3	17	33	3	2	2	60

Note. SM = sense-making; Wr = writing; Pro = procedure; Rd = reading; TA = instructor help; Off = off-task; Tot = total.

greater time sense-making, and this difference became more evident toward the end of the semester, with the non-design students decreasing the amount of time they spent sense-making with each lab (see Figure 5).

Within the time spent on sense-making, the distribution of discussion issues was different for the two groups. We coded instances when students discussed the issues of design, the physics concept involved, the mathematical procedure, assumptions inherent in the mathematical procedure, experimental uncertainties, and revisions of the experiment based on the outcome. Figure 6 shows the average cumulative time over 10 laboratory sessions that design and non-design students spent on these activities. Students designing their own experiments spent the highest percentage of sense-making time discussing issues associated with the design, whereas the non-design group spent most of its sense-making time discussing the mathematical procedure. In addition, in the non-design labs there was no time spent

TABLE 5
Average Time in Minutes Spent by Students on Different Activities in the Labs

Labs	*Sense-Making*	*Writing*	*Procedure*	*Reading*	*Instructor Help*	*Off-Task*	*Total*
Design group							
1–10	37 (10.0)	66 (12.0)	24 (13.0)	5 (1.7)	18 (16.0)	8 (9.2)	159 (25.9)
Non-design group							
1–10	14 (8.4)	41 (15.1)	20 (10.7)	4 (3.2)	17 (12.8)	2 (1.4)	96 (30.8)

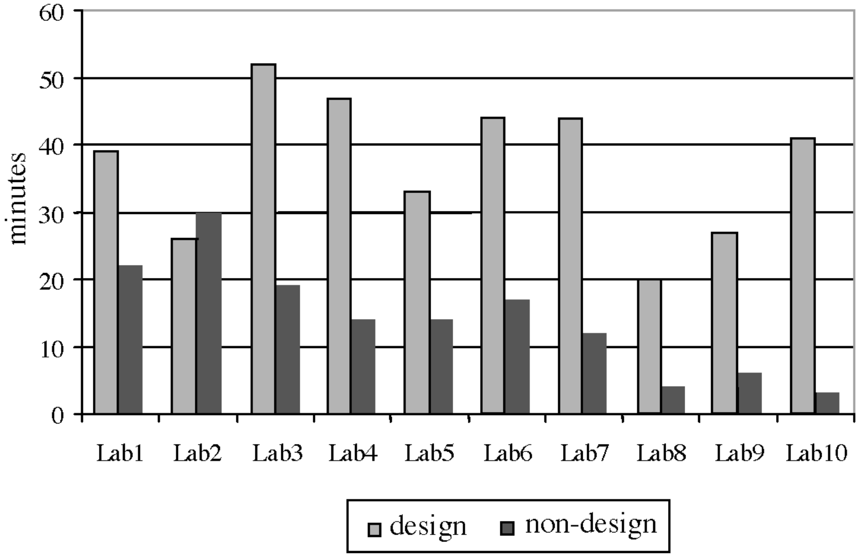


FIGURE 5 Time spent by students on sense-making discussions in design and non-design labs during the semester.

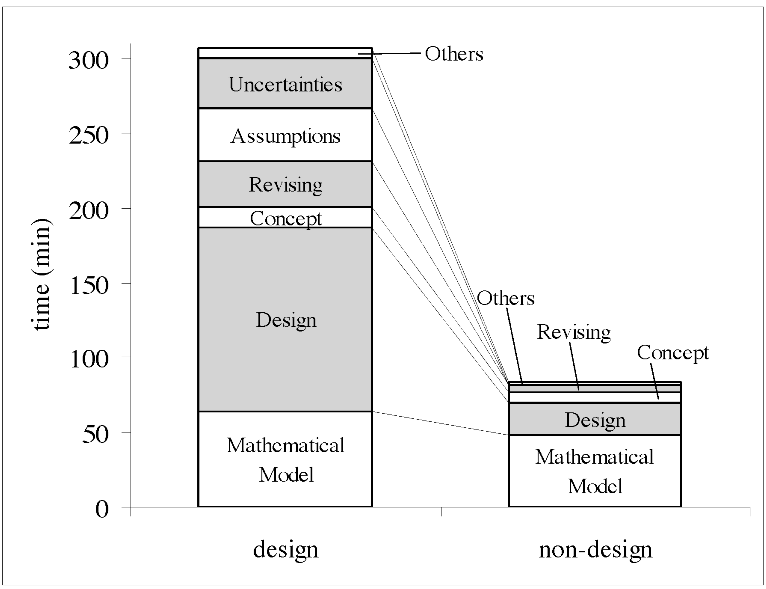


FIGURE 6 Topics of students' discussions during sense-making episodes.

on the discussions of assumption or uncertainties, as these were described in the lab handout. These findings are very similar to our findings for student behaviors in non-design labs in a different course (Karelina & Etkina, 2007). We believe that they accurately describe how students spend time in traditional introductory labs.

With respect to the total time spent in the lab, whereas the design student teams spent close to 3 hr in lab on average throughout the semester, the non-design student teams spent much less time, and this time decreased on average toward the end of the semester. Overall, by the end of the semester the design student teams were spending much more total time on average in the lab and a greater percentage of the total time on sense-making (see Table 5) compared to the non-design students. Although the lab was 3 hr long, non-design students chose to leave early.

How Does Design Affect Students' Approaches to Experimental Inquiry?

The observations showed that there was a difference in the behavior of design and non-design students during the physics and bio interpretive knowing test labs. The patterns described previously persisted, although during those labs both groups had to design experiments with no scaffolding. On closer inspection, we found that the physics drag force lab took more time for design students. The design teams spent more than 40 min more time in the lab room than the non-design students. The duration of the physics drag force lab was 162 ± 17 min (*SD*) for the design students versus 120 ± 25 min for the non-design students. The duration of the bio transpiration lab was 176 ± 26 min for the design students versus 153 ± 26 min for the non-design students.

The differences between the groups and activities in the drag force lab were significant: two-way ANOVA between group (design vs. non-design), $F(1, 42) = 14.33$, $p < .001$; among treatment (among coded categories), $F(6, 42) = 130.39$ $p < .001$. In the biology transpiration lab, differences between groups were not significant, but differences among coded categories were: two-way ANOVA between group, $F(1, 42) = 3.70$, $p = .061$; among coded categories, $F(6, 42) = 141.80$, $p < .001$.

Design students spent considerable time sense-making in both labs; non-design students spent little time sense-making in both labs. Figure 7 shows differences in sense-making discussions. The sense-making lasted a mean of 52 min ($SD = 7$) and 42 min ($SD = 7$) in the design groups in the physics and bio labs, respectively, but 15 min ($SD = 3$) and 19 min ($SD = 3$) in non-design groups. The time that students spent on other activities was more similar between the two groups.

How Does Design Affect Students' Development of Scientific Abilities?

During the semester, students in the design sections had multiple opportunities to develop scientific abilities. Table 6 lists the abilities used for data collection in the study and sample tasks/questions that students had in the labs that addressed

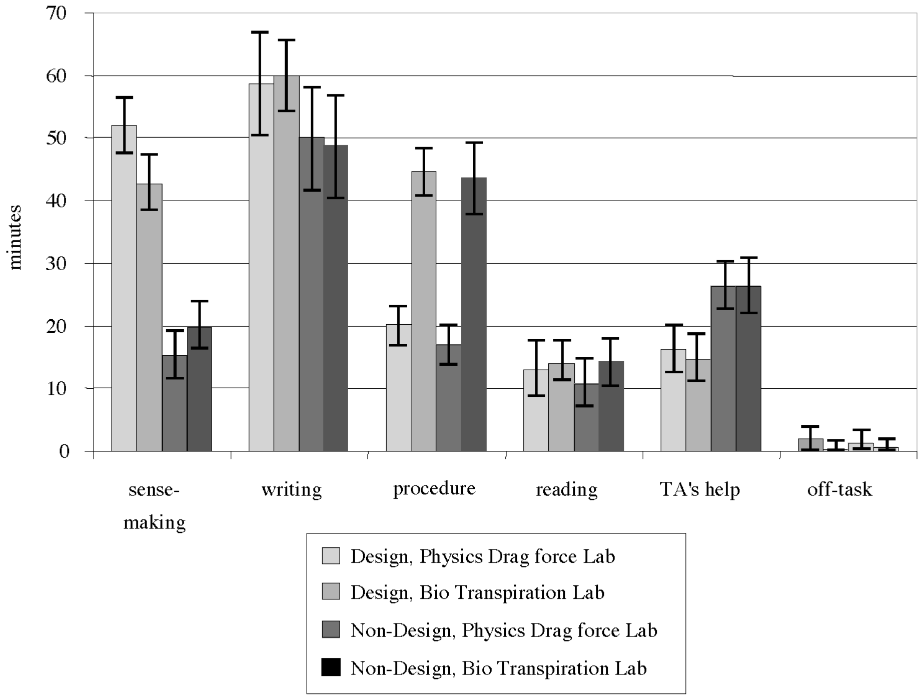


FIGURE 7 Time spent by teams of students on different activities during the physics drag force and bio transpiration labs ($N = 4$ groups). TA's help = instructor help.

the development of those abilities. The table also provides the letter/number identifier of a specific rubric that helped students self-assess the ability. Figure 8 presents a sample student lab report and how it was scored using the rubrics. In Figure 9 we present the scores of the students in the design sections on the relevant abilities at the beginning and the end of the semester. Students in the design sections improved their performance on the abilities chosen for assessment.

Does Design Impede Students' Acquisition of Normative Science Concepts?

With regard to the normative science concepts that were assessed via multiple choice and free-response exam questions and problems, students in the design and non-design groups performed similarly on both midterms and the final exam: Midterm Exam 1, $F(1, 182) = 0.25, p = .62$; Midterm Exam 2, $F(1, 180) = 1.31, p = .25$; final exam, $F(1, 180) = 0.45, p = .502$ (to make three contrasts, we used the sequential Bonferroni correction, critical value of 0.017; see Table 7).

TABLE 6
Scientific Abilities Rubrics Used for Assessment and Self-Assessment in Design Labs

Scientific Ability (Designated Rubric)	*Task/Handout Question/Prompt*
To evaluate the results by means of an independent method (D5)	Design two independent experiments to determine the maximum coefficient of static friction between your shoe and the sample of floor tile provided. Include in your report the following for each independent experiment.
To communicate the details of an experimental procedure clearly and completely (F1)	Draw a sketch of your experimental design. Write a brief outline of the procedure you will use. Think of how you will represent the data so the person who was not present during the experiment can understand.
To identify the assumptions made in using the mathematical procedure (D8) To determine specifically the way in which assumptions might affect the results (D9)	Decide what assumptions about the objects, interactions, and processes you need to make to solve the problem. How might these assumptions affect the result? Be specific. Considering one of the relevant assumptions, evaluate its effect on the results. For example, estimate how the normal force will change if you pull the shoe not horizontally but at an angle of 5° relative to the horizontal direction.
To identify sources of experimental uncertainty (G1) To evaluate specifically how identified experimental uncertainties may affect the data (G2)	What are possible sources of experimental uncertainty? Which measurement gives you the highest uncertainty?
To evaluate specifically how identified experimental uncertainties may affect the data (G2) To describe how to minimize experimental uncertainty and actually do it (G3)	Name two different methods to measure the angle with the available equipment. Estimate uncertainty for each of them. Which method gives less uncertainty and allows you to minimize the angle measurement uncertainty?
To describe how to minimize experimental uncertainty and actually do it (G3)	Make sure you take steps to minimize uncertainties. What is the outcome of the experiment?
To make a judgment about the results of the experiment (D4) To evaluate the results by means of an independent method (D5)	When finished with both experiments, compare the two values you obtained for the maximum coefficient of static friction. Decide, using assumptions and uncertainties, if these values are different or not. If they are different, what are possible reasons?

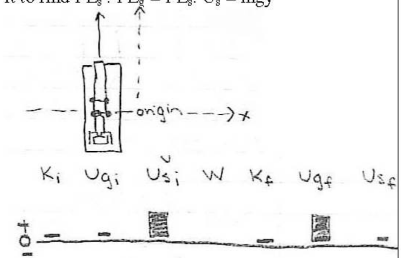
Lab report of a student group	**Commentary**	**Rubric score**
Experiment 1: For each notch, shoot car up into the air, measure the distance that it goes in the air until it starts to fall. Then we can find PE_g and use it to find PE_s : $PE_g = PE_s$. $U_s = mgy$ 	Experiment solves the problem Communication: Explanation and justification of the method with the sketch and energy bar chart	D2: Adequate F1: Adequate
We measured mass of the car, the distance it travels into the air (shoot against the wall and mark the wall where the car reaches max) $U_s = mgy$	Data collection: All of the chosen measurements can be made and all details about how they are done are provided and clear.	D3: Adequate
We assume that the car is point particle, that it shoots straight up and that we can accurately measure the vertical distance. If we cannot measure accurately our values cannot be accurate	Assumptions: Most of relevant assumptions are identified. Effect of assumptions: The effect of an assumption is mentioned but described vaguely and confused with the uncertainty.	D8: Need some improvement D9: Inadequate
y / mgy 1^{st} notch 0.515m 0.141 J 2^{nd} notch 0.720m 0.198 J 3^{rd} notch 0.945m 0.259 J 4^{th} notch 1.175m 0.322 J	Data: All important data are present, organized, and recorded clearly. Analysis: Correct mathematical procedure	G4: Adequate D7: Adequate
We estimate our measurement of the distance is accurate to ±2cm. We estimate that uncertainty from assuming car is a point particle is about ±6cm (length of the car). This uncertainty is the largest so we will ignore others. This is a relative uncertainty of 11%.	Uncertainties: Most of important uncertainties are identified. Random uncertainty is not evaluated. Evaluation of uncertainty: The final result does not incorporate uncertainty No attempt to minimize uncertainty	G1: Adequate G2: Inadequate G3:Missing
Experiment 2: Launch car horizontally let car roll on floor, mark 1m from the rear of the car and measure the time it takes the car to reach the 1m mark. Thus we can calculate the car's kinetic energy.	Experiment solves the problem Explanation and justification of the method with the sketch and energy bar chart	D2: Adequate F1: Adequate

FIGURE 8 Sample student lab report with rubric coding.

(continued)

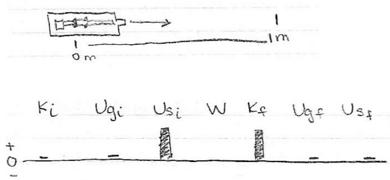
 We measure mass of the car and time it takes. $U_s = K = ½ mV^2$; $V= x/t$		
We assumed that floor is frictionless, and that all the potential energy of rubber band is transferred to the car. If floor is not frictionless our calculated V will be less than V immediately after layncer	Correct assumptions Acceptable evaluation of the effect of an assumption	D8: Adequate D9: Adequate
time / velocity / $K = U_s$ 1st notch: 0.87s, 2.3m/s, 0.074 J 2nd notch: 0.60s, 3.3m/s, 0.152 J 3rd notch: 0.40s, 5.0m/s, 0.350 J 4th notch: 1.35s, 5/7m/s, 0.455 J	All important data are present, organized, and recorded clearly. Correct mathematical procedure	G4: Adequate D7: Adequate
Our reaction time is the largest uncertainty in this experiment (±1s, relative uncertainty of 115%)	Uncertainty evaluated incorrectly (reaction time is about 0.2s, that gives about 20% uncertainty) The final result does not incorporate uncertainty Random uncertainty is not evaluated. No attempt to minimize uncertainty	G2: Inadequate G3: Missing
	No discussion is presented about the results of the experiment No discussion about the differences in the results due to the two methods	D4: Missing D5: Inadequate

FIGURE 8 *(Continued).*

How Does Engaging in Design Affect Students' Development of Scientific Abilities as Evidenced in Their Performance on Transfer Tasks?

Design students demonstrated significantly better scientific abilities than non-design students, as shown in Figure 10. A considerable number of design students received scores of 2 (needs some improvement) or 3 (adequate) for identifying assumptions in their lab reports (see Figure 10a). Design students were able to identify relevant and significant assumptions of the theoretical model that they used, whereas only a few non-design students were able to do so: $\chi^2(3, N = 178) = 68, p < .001$, for the physics drag force lab; $\chi^2(3, N = 181) = 120, p < .001$, for the biology transpiration lab. In addition, about half of the design students evaluated the effects of assumptions on the result or validated them in both labs. No students in the non-design sections made an attempt to do this (we do not rep-

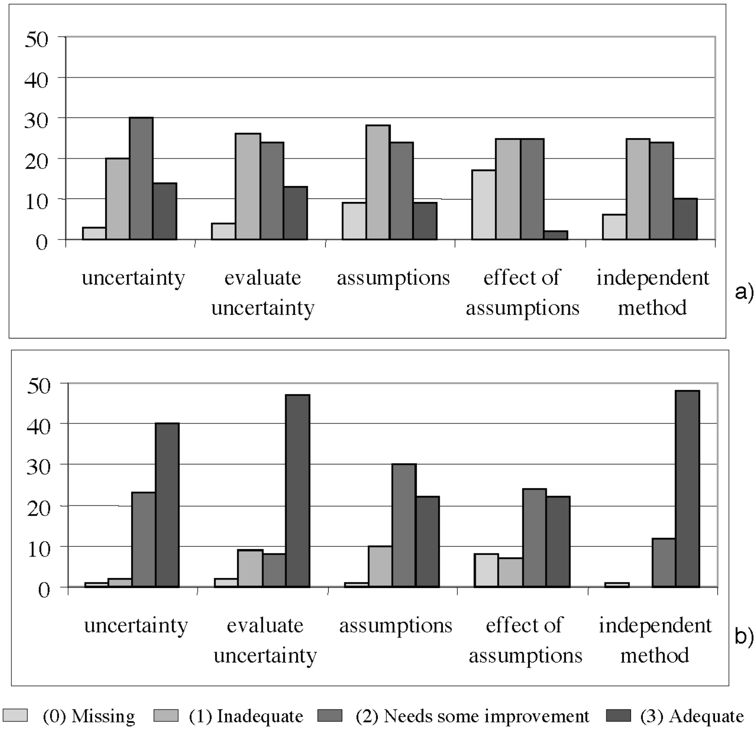


FIGURE 9 Performance of design students on scientific abilities at the (a) beginning and (b) end of the semester. The differently shaded bars show the percentage of students who received scores of 0 = *missing*, 1 = *inadequate*, 2 = *needs some improvement*, and 3 = *adequate*.

resent this result in a figure as the data for a non-design group bars all have zeros).

During the semester, non-design students had to identify sources of uncertainties and how to evaluate their effect on the final answer. Every lab handout had specific instructions on how to do this. Only a few of these students, however, used this ability in the independent experimental investigations (see Figure 10b). More than 50% of design students evaluated the effect of experimental uncertainties in these labs. The difference in the number of students who evaluated uncertainties was statistically significant: $\chi^2(3, N = 178) = 30$, $p < .001$, for the physics drag force lab; $\chi^2(3, N = 181) = 94$, $p < .001$, for the bio transpiration lab.

A high score on the rubric "evaluating the results by means of an independent method" was possible only when a student discussed the discrepancy between the results of two methods and possible reasons for this discrepancy considering assumptions and uncertainties. We found that design students demonstrated a greater

TABLE 7
Exam Scores

	Midterm Exam 1			*Midterm Exam 2*			*Final Exam*		
Group	*MC (80)*	*FR (60)*	*Overall (140)*	*MC (60)*	*FR (80)*	*Overall (140)*	*MC (120)*	*FR (120)*	*Overall (240)*
Design	48.6	47.4	96.0	46.0	58.7	104.7	85.7	89.2	174.9
Non-design	50.0	47.8	97.8	42.1	56.3	98.4	86.2	84.6	170.8

Note. MC = multiple choice questions; FR = free response questions

ability to evaluate the results (see Figure 10c). In all, 72% of design students received a score of 2 or 3 for their reports in the physics drag force lab (i.e., discussed the reasons for the discrepancy). In non-design sections, only 43% of students did this, $\chi^2(3, N = 178) = 16, p < .001$. In the bio transpiration lab, 79% of the design students discussed the difference versus 37% of non-design students, $\chi^2(3, N = 181) = 42.25, p < .001$.

One of the main scientific abilities we wanted students to develop is the ability to communicate their ideas. This includes an ability to draw diagrams and pictures, describe details of the procedure, and explain the methods. The analysis of lab reports showed that more than 60% of design students drew a picture, whereas only 8% of non-design students did. Figure 10d shows the lab report scores for the communication rubric. The difference in scores was statistically significant: $\chi^2(3, N = 178) = 30$, $p < .001$, for the physics lab; $\chi^2(3, N = 181) = 41.65, p < .001$, for the bio lab.

The analysis of the students' reports for the physics drag force lab revealed another feature related to student construction of scientific interpretive knowing. When solving complex problems, scientists spontaneously use different concrete representations such as pictures and diagrams as tools to assist them in problem solving (Kindfield, 1993; Kozma & Russell, 1997). An example of such tools in mechanics are force diagrams (or *free-body diagrams*, as they are called in physics). The quality of force diagrams drawn by students from the two groups was different in spite of the fact that during the semester all students learned to draw force diagrams the same way. In this lab, about 22% of non-design students drew incorrect force diagrams (i.e., with mislabeled or not labeled force vectors, wrong directions, extra incorrect vectors present, or vectors missing), whereas only 2% of design students made a mistake in drawing force diagrams, $\chi^2(3, N = 178) = 18, p < .001$.

In addition, we analyzed whether students constructed consistent representations (e.g., a force diagram vs. mathematics, a picture vs. a free body diagram). We used three codes for the representations: missing, inconsistent, and consistent. We found a difference in the number of students who created inconsistent representations: 22% of design students versus 44% of non-design students, $\chi^2(2, N = 178) = 7.8, p = .02$.

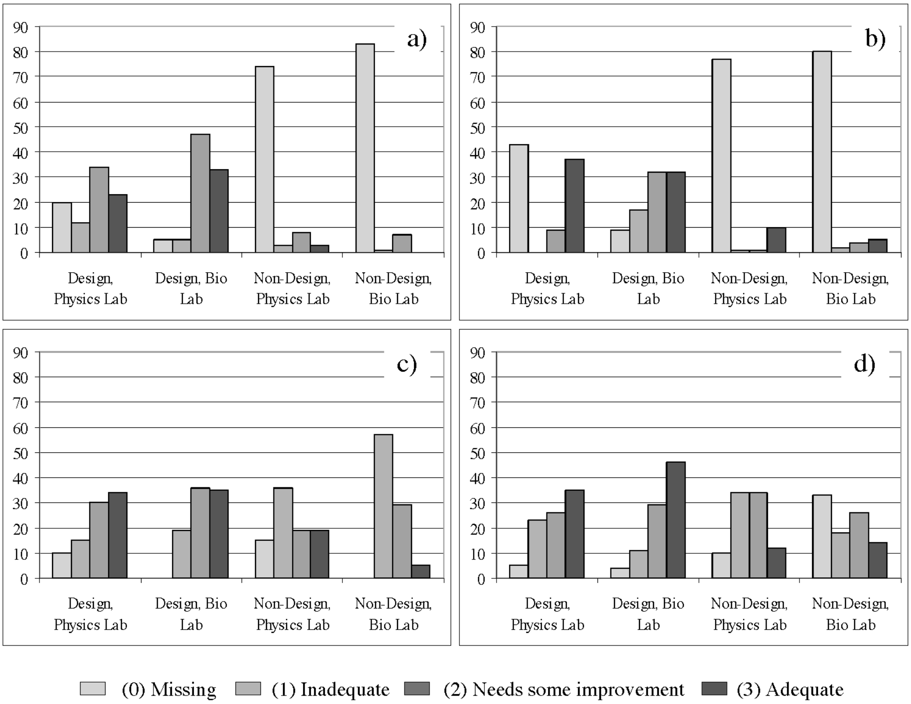


FIGURE 10 The percentage of students whose lab reports received different rubric scores for their ability to (a) consider assumptions in the theoretical model, (b) evaluate uncertainties, (c) evaluate the results by means of an independent method, and (d) communicate ideas during the physics and bio interpretive knowing test labs.

Finally, we analyzed the rubric scores for the lab reports written by design students during regular semester labs and for the physics and bio interpretive knowing test labs to determine how abilities develop during the semester and how students apply them in novel situations. Figure 11 shows the students' lab report scores for the ability to consider assumptions in the theoretical model for five regular labs (scaffolding provided) and two interpretive knowing test labs (no scaffolding). The number of students who identified assumptions in their lab reports increased during the semester and reached about 80% in the labs at the end of the semester.

To illustrate the differences in student lab reports for the design and non-design groups in both the drag force and the transpiration labs, we provide examples of two lab reports, one from each group (the design report is representative of a good report and the non-design report is the best report), annotated using the rubrics (see Figure 12). These two excerpts demonstrate the differences in the students' approach. The

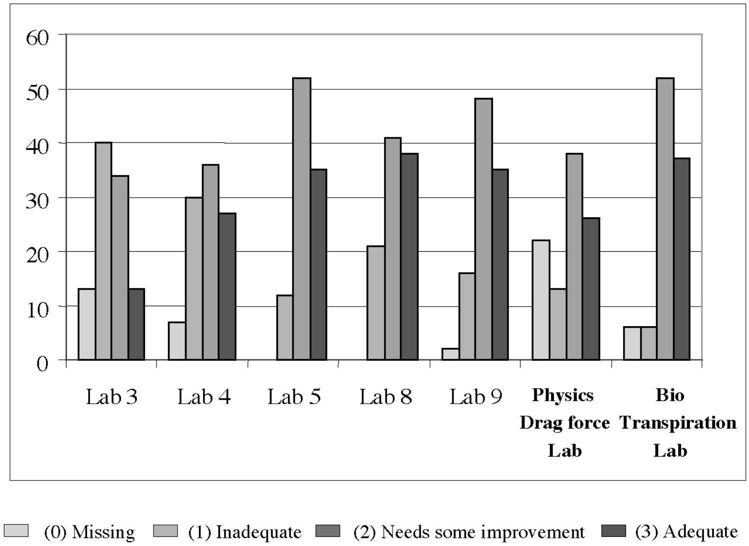


FIGURE 11 The percentage of design students whose lab reports received different rubric scores for their ability to consider assumptions in five regular semester design labs and two interpretive knowing test labs.

non-design student wrote step-by-step instructions emulating the style of the handouts he or she was used to, but this student did not provide a labeled sketch of a setup and force diagram, whereas the design student did. Also, the non-design student did not give any explanation or justification of the experimental procedure. The design student explained the physical processes of the experiment and gave a mathematical model of these processes. In addition, this student identified limitations and assumptions in this model and tried to evaluate how these assumptions may affect the result of the experiment. The non-design student provided a correct mathematical description of the physical process but did not consider inherent assumptions. This student used this model to evaluate the required drag coefficients of air-filled and helium-filled balloons, but these results did not incorporate uncertainties. Note that the non-design student emphasized the necessity to use multiple trials but did not demonstrate the understanding that this is necessary to evaluate and minimize uncertainty. Because the non-design student did not consider the uncertainty of the measurements, his or her judgment is not justified. It is impossible to say whether drag coefficients for air and helium balloons are different. In contrast, the design student made a judgment based on the uncertainty analysis and attempted to consider the effects of assumptions. In general, the lab report written by the student in the design group provided evidence of a more sophisticated approach to the same investigation compared to the report written by the student from the non-design group.

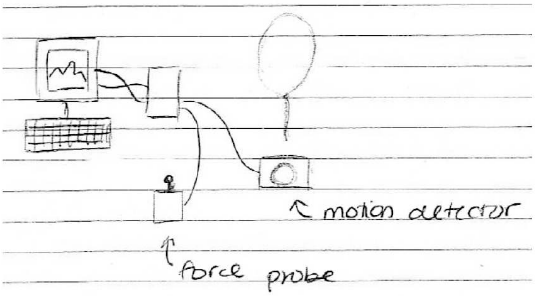
DISCUSSION

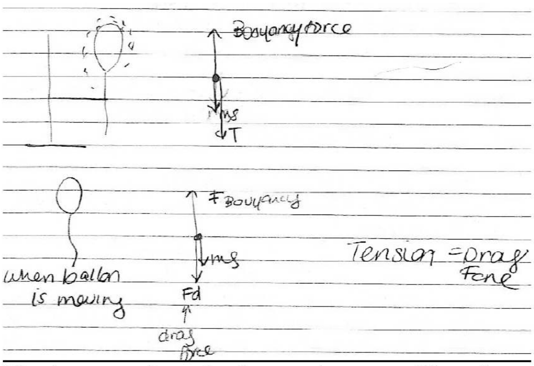
The results of this study support our hypothesis that the design element in the *ISLE* labs, which includes design itself, reflection, and self-assessment, enriches students' learning opportunities. Students who were in design lab sections

Lab report of a student from a non-design lab group *(Task: version 2)*	Commentary
1) Find the circumference of the balloon. 2) Using circumference find the radius of the balloon. 3) Find the mass of the balloon by taping the string to the electronic scale. (Note: you will get a negative value. Take the absolute value) m = 0.65g 4) Find the drag force $F_D = 0.65g\times 10m/s^2$	Clear step-by-step instructions, copying the handout's style. No explanation or justification of the method. No picture or FBD.
5) Determine the velocity of the balloon when air resistance and gravitational force are equal: a) place the motion detector on a stand b) place the sensor face downward c) place the helium balloon on the floor d) release the balloon as the motion detector collects data	Detailed instructions
e) on the position-time graph find constant slope segment f) the slope of the graph is the velocity V_1 = .8975m/s g) repeat twice more h) find the average velocity V_{avg} = .854m/s	Multiple trials with averaging without evaluating uncertainty, repeating the rote of the regular labs
6) Use the following equation to determine Reynolds number… You should get a value larger than 10. - use the equation to solve for drag coefficient …Cd=0.51	A correct mathematical description but it does not consider inherent assumptions.
- now repeat this procedure for air filled balloon. Make sure to drop the balloon from the level of the motion detector… - air filled balloon - Cd= 0.61 Drag coefficient for air and helium are indeed different.	Since the results do not incorporate uncertainties, judgment is not justified and not reasonable.

FIGURE 12 Examples of students' lab reports.

(Continued)

<table>
<tr><th>Lab report of a student from a design lab group
(Task: version 4)</th><th>Commentary</th></tr>
<tr><td>Part I. We need to know which equation to use based on the Reynolds number…To find the velocity we will have a motion sensor above the helium balloon. The balloon will be released and the motion sensor will measure its upward velocity.</td><td>Clear description of the experiment Explanations and justification of chosen methods with clear force diagram.</td></tr>
<tr><td>
</td><td>Here is the picture of the set-up.

The position-time graph is attached</td></tr>
<tr><td>The velocity was taken 3 times and averaged to allow for random uncertainty....
When the balloon is let go the velocity increases until it reaches terminal velocity, here the net force is zero and acceleration is zero.
When balloon is at rest the net force on it is equal to zero too.</td><td>Multiple trials with averaging to evaluate uncertainty</td></tr>
<tr><td>
</td><td>Labeled sketch, chart and force diagram are consistent

Here are two force diagrams for balloon at rest and at terminal velocity.</td></tr>
<tr><td>The buoyant force is always the same. Therefore the drag force is equal to the force of the string</td><td></td></tr>
</table>

FIGURE 12 *(Continued).*

<table>
<tr>
<td>attaching the balloon to the scale… Cd = 0.43
Assumptions: balloon travels in straight path, balloon is point particle, cross-section is circle, cross-section is level.
<table>
<tr><td>Uncertainty</td><td>Relative uncertainty</td></tr>
<tr><td>L= 21.9cm±0.05cm
V(random uncertainty)
=1.248m/s±0.1
F_D(scale) = 1.04g±0.05g</td><td>0.2%

8% - the largest
5%</td></tr>
</table>

Part II. Prediction (of the speed of the air balloon falling to the ground)
When the air balloon falls it reaches terminal velocity. Drag force equals the force of the earth. …
We can use the equation … to get the velocity: V= 0.438±0.021m/s
We will have a motion sensor aimed down and drop a balloon below it. It will record the velocity of the air balloon before it hits the ground.
Assumptions: 1. Balloon achieves terminal velocity – otherwise Fe≠Fd; 2. Re>10 – otherwise Fd equation is wrong 3. Cd is the same for air and helium – otherwise calculated velocity will be wrong.
V was measured and averaged over 3 trials (1.476, 1.02, 1.153). V=1.216±0.228m/s
The values do not overlap and therefore are not equal. Some assumptions must have been incorrect.</td>
<td>Explanations and justification of chosen methods with a clear force diagram.
Additional assumptions are identified.
Uncertainties are evaluated: diameter, scale, motion detector, and random uncertainty of the velocity.

The final result incorporates uncertainty
A picture is here.
All important assumptions are identified.
It is shown which values are affected by assumptions
The result incorporates uncertainty.
Judgment is based on the uncertainty analysis. There is attempt to consider the effects of assumptions.</td>
</tr>
</table>

FIGURE 12 *(Continued)*.

approached new tasks in biology and physics in more scientifically productive ways than students who did not experience design labs during the semester. In particular, we can say that students in design lab sections were attentive to the measurement, considered assumptions in the mathematical procedure, evaluated the results, and communicated much better than students from non-design sections. As the students demonstrated these abilities without prompts or scaffolding and in completely new content, we can say that they developed some of the scientific habits of mind. From the results we infer that for the students to recognize that the transfer tasks required these abilities, they had to interpret those tasks, and the context in which the tasks had to be performed, accordingly. These interpretive ways of knowing were most likely encouraged by the designing and undertaking of their own experiments and by the embedded reflection during the semester.

We contend that cognitive apprenticeship and formative assessment form the foundation of the *ISLE* design labs that account for group differences in terms of interpretive knowing. Through cognitive apprenticeship in the design labs, students became acquainted with the processes, procedures, and methods of scientific practice. They had to design their own experiments in every lab. Although design is a complex task, we helped students to be successful by following several strategies. The structure of the tasks together with the added scaffolding provided by the instructors made scientific thinking visible to the students. Complex abilities—such as the ability to collect and analyze data—were broken into smaller sub-abilities, and scaffolding questions, rubrics, and reflective questions helped students master the elements of this thinking gradually. In addition, students received continuous coaching, making their learning guided, supported, supervised, and managed. We were able to gradually remove scaffolding and coaching as students became more independent and began to develop scientific habits of mind. These labs integrated formative assessment by students' ongoing assessment of their progress and by continuous feedback and adjustment of instruction in order to respond to students' needs.

In our study we sought to address the following five questions: How does the need to design their own experiments affect the types of activities in which learners engage? How does designing their own experiments affect students' approaches to experimental inquiry? How does designing their own experiments affect students' development of experimental procedures, processes, and methods? Does devising, designing, and conducting their own experiments affect students' acquisition of science concepts? and How does engaging in design affect students' development of scientific abilities as evidenced in their performance on transfer tasks?

With respect to the first question, we found that lab activities during the semester elicited more thoughtful responses when students engaged in design tasks, as the amount of time that design students spent on sense-making remained constant throughout the semester and was significantly greater than the corresponding time spent by non-design students. Both groups of students started the semester spending about the same amount of time on sense-making, but around the third week of the semester non-design students began to dedicate less than half the time that their counterparts did to reasoning exchanges. Previous studies, such as work by Hmelo et al. (2002), have shown that authentically complex tasks compel students to engage in monitoring and reflecting. Our previous studies showed that in introductory labs offered through traditionally taught courses, students spend 15 min on average on sense-making (Karelina & Etkina, 2007), which is much less than what we observed in the design labs in this study. It is possible that student sense-making time decreases as the semester progresses because the nature of the tasks does not require such activity, and students who initially seem to be eager to engage in sense-making stop doing it. An alternative explanation is that as students become familiar with lab requirements, they are able to execute them with less effort as some of these become ritualized (Kolodner et al., 2003). However, if that were the case, then we would expect decreases for both groups.

In addition, we found that when students engaged in sense-making, they focused on different issues. Whereas design students spent relatively more time discussing the experiment, assumptions in the mathematical procedure, and revisions of their work, the non-design students focused their efforts mostly on the mathematical procedure. Why might we find such a difference? One possible explanation is that when students designed their own experiments, they were often unsure of the correctness of their actions. This created a basis for metacognitive thinking, such as planning and evaluating the methods and the results. The students realized the necessity of these activities that are naturally embedded into designing the experiment. When performing the non-design experiments, students had to follow the directions and might not have felt an urge to evaluate their actions. It is possible that in such situations, students considered evaluating results, assumptions, and uncertainties as an ungrounded and useless activity. This could have decreased students' motivation. Furthermore, even if students learned how to evaluate the results and their methods, they did not know when and why they should apply this knowledge.

The second question that we sought to answer was how the design of experiments affects the way in which students approach experimental inquiry. Design students spent significantly more time during the semester on sense-making than non-design students. This seems to indicate that the design labs supported the development of students' initial tendencies to engage in experimental inquiry. We conjecture that all students start a physics course expecting to spend time sense-making in the labs. It is the prescriptive structure of the traditional tasks that might discourage sense-making, resulting in students getting into a habit of not spending time on monitoring and reflecting. This habit is so strong that, even if given a task that requires sense-making, students do not engage in it. The design students may have continued with sense-making even once the scaffolding was faded and there were no prompting questions in the lab handout because they acquired such a habit. Another explanation is that students in design sections got used to justifying their actions and procedures to their group mates, whereas non-design students got used to following directions without questioning them. Thus, when they encountered a task with no directions (in the interpretive knowing test labs), they did not spend time arguing about it but performed the experiments that first came to mind.

With respect to our third question about the effects of design tasks on students' development of scientific abilities, we found that students' abilities improved significantly throughout the semester but not all to the same degree. We discovered that after 10 weeks the students designing their own experiments advanced their ability to identify and evaluate sources of uncertainty, their ability to minimize uncertainties, their ability to identify the implicit assumptions in their procedures and estimate the effect of these assumptions, and their ability to evaluate their experimental results by means of an independent method.

Previous research (Kuhn & Pease, 2008) has shown that it is possible even for elementary students to develop inquiry abilities, such as formulating questions, interpreting evidence, drawing conclusions, and representing and communicating findings. However, these authors found that without special instruction directed at the acquisition of inquiry skills, students do not progress in this regard, as those abilities are by no means intuitive. Similarly, we found that by simply completing step-by-step lab tasks, the students in the non-design group did not incorporate scientific abilities into their investigative resources to the level that those in the design group did. Kuhn and Pease attributed the effectiveness of their inquiry instruction to two features: (a) the extended, deep engagement of learners in problems that required the abilities; and (b) the use of gradually fading scaffolding. This is consistent with our findings. In this same direction point the results of Roth and Roychoudhury (1993) and Roth (1994), who investigated an instructional environment in which students successfully developed what they called *science process skills*. These studies indicate that both the activities assigned and the instructional context are critical for the learning of these complex skills. Learners need to work in rich environments solving meaningful problems and have the guidance of a teacher who introduces them to the practices of the scientific community.

At the same time, we found that students' learning of normative physics content did not suffer when students designed their own experiments, taking into account that they spent more time writing their lab reports and thinking about scientific procedures versus solving physics problems. More significantly, students' physics learning did not suffer, even though in some cases their devised procedures were not optimal or their experimental results were incorrect. There were several instances during the semester when students learned particular content in the labs only and then had to solve problems related to this content on exams. Therefore, we conclude that students who engage in experimental design and reflection learn more than those who do not. The design students in this study achieved similar scores on exams as those students who were not assigned design tasks, and they developed more productive scientific habits of mind. We contend that learning the scientific habits of mind is as important as learning science concepts (Duschl et al., 2007).

One might question why students in the design group did not perform better on the exams than those in the non-design group, taking into account that the former spent more time in the laboratories. As we found, however, most of the extra time that students spent in the labs was dedicated to the experimental design, not the physics concepts that were being assessed by paper-and-pencil exams. In addition, there is the possibility that the cognitive load of the design distracted some of the students from focusing on the physics concepts, but the results argue against this explanation. We contend that the careful scaffolding provided in the *ISLE* design labs allowed students to learn both the physics content and the scientific abilities (Hmelo-Silver, Duncan, & Chinn, 2007).

Finally, we also found that when students engaged in the design of experiments, they not only developed scientific abilities but used them without prompts and scaffolding on transfer tasks. Students from the design group wrote lab reports that received much higher scores on all scientific abilities, perhaps meaning that these students acquired the schema (Baker & Dunbar, 1996) of a simple lab design experiment: think about the physics of the situation, try to represent your ideas consistently to solve the problem, assess your assumptions, evaluate your uncertainties, make sense of your results, and clearly communicate them to a person who will read your report.

In conclusion, we have shown that the development of students' scientific abilities is fostered through design labs. Students generated written samples reflecting understanding of the process of experimental design in physics, and this knowledge was successfully applied to new problems.

CONCLUSIONS

According to Baker and Dunbar (1996), scientists possess a variety of rich experimental schemas. For example, when scientists encounter a problem that involves some type of measurement, they tend to think of different kinds of measurement uncertainties inherent in this particular measurement. This ability to think about certain aspects of a problem is what we call *interpretive knowing*. Helping students see these aspects and attend to them without prompts is an important aspect of building their habits of mind. These habits of mind can be used in any situation, not only in the physics classroom. In this study, we regarded enculturation in the scientific community of practice (Barab & Duffy, 2000) with a characteristic way of approaching and framing problem situations as transfer (Schwartz et al., 2005). As we mentioned previously, interpretive knowing cannot be seen directly; it can only be inferred from observing and analyzing students' approaches to dealing with problem situations.

Based on our findings, we suggest that the idea of interpretive knowing is essentially connected to the concept of cognitive resources (Hammer, Elby, Scherr, & Redish, 2005). One of the goals of science education is to help students learn to activate community-accepted cognitive resources when solving science problems.

These resources are not the factual solutions but the approaches that scientists use when faced with similar problems. This activation of resources by a person can be prompted by the context (when the task directly or indirectly leads the students) or can occur deliberately when the person is aware of what needs to be done in a particular situation. That subsequent activation of the same group of resources may possibly help form a habit of using the resources together. We can then interpret the results of our study in the following way: Students who were enrolled in design labs learned to activate without any prompts some of the resources that scientists would activate when faced with the same problem. We argue that the multifaceted intervention in the design labs afforded students a new way of perceiving the experimental tasks.

Research on transfer has commonly focused on specific kinds of knowledge or procedures rather than on habits of mind important in scientific inquiry. Besides this branch of transfer research, there have been important studies on the transfer of abilities, such as the work by Engle (2006) on the transfer of the ability to construct scientific explanations or Zohar and Nemet's (2002) study on the transfer of argumentation skills. Our study contributes to this work because it demonstrates, in a way, the transfer of scientific abilities across scientific domains. This level of transfer does not come easily. Rather, we argue that it is the synthesis of the three theoretical pillars that provide a foundation for *ISLE* labs that is important. Cognitive apprenticeship helps provide the support that is needed for learners to engage in the complex tasks of designing, conducting, and interpreting experiments. Formative assessments scaffold students in thinking about what is important for design and promote students' metacognition that they need to compare their own work with standards provided by the rubrics. Finally, when students engage in these complex tasks, they must activate their prior knowledge, differentiate their ideas, and look at lab tasks with scientific eyes. Only those students who develop interpretive knowing can be successful in those activities. In conclusion, we can say that one does not need to construct a concrete artifact to learn (as suggested by Kafai & Resnick, 1996), but rather one needs to design a conceptual object. This type of design supports the development of the habits of mind, which, as Bereiter and Scardamalia (2006) indicated, are essential to creating a knowledge-building society.

ACKNOWLEDGMENTS

We thank John Bransford, Jose Mestre, Joe Redish, and Larry Suter for their advice in the design of the project; Michael Gentile for teaching the labs in the course; Ravit Duncan for advice on representing the data; Noah Finkelstein for productive discussions; Tara Bartiromo for editing the article; and all of the students in Physics 193/94, without whom this study would not be possible. We are indebted to Alan Van Heuvelen for writing non-design labs, providing advice during the project, and providing comments on this article. We thank the anonymous reviewers, Associate Editor Barry Fishman, and Co-Editor-in-Chief Yasmin Kafai for their helpful feedback on an earlier manuscript. We also thank the National Science Foundation for its continuous support of our work. The project was supported by National Science Foundation Grant DRL 0241078. Conclusions or recommendations expressed in this material are ours alone and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Alberts, B. (2000). Some thoughts of a scientist on inquiry. In J. Minstrell & E. van Zee (Eds.), *Inquiring in inquiry learning and teaching in science* (pp. 3–13). Washington, DC: American Academy for the Advancement of Science.

Apedoe, X., & Ford, M. (2009). The empirical attitude, material practice and design activities. *Science & Education.* Retrieved July 3, 2009, from http://dx.doi.org/10.1007/s11191-009-9185-7

Baker, L. M., & Dunbar, K. (1996). Constrains on the experimental design process in real world science. In G. W. Cottrell (Ed.), *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society* (pp. 399–440). La Jolla, CA: Erlbaum.

Barab, S. A., & Duffy, T. (2000). From practice fields to communities of practice. In D. Jonassen & S. M. Land (Eds.), *Theoretical foundations of learning environments* (pp. 25–56). Mahwah, NJ: Erlbaum.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128,* 612–637.

Barron, B., Schwartz, D. L., Vye, N. J., Moore, A., Petrosino, A., Zech, L., et al. (1998). Doing with understanding: Lessons from research on problem- and project-based learning. *Journal of the Learning Sciences, 7,* 271–311.

Bell, P., & Linn, M. (2000). Scientific arguments as learning artifacts: Designing for learning from the Web with KIE. *International Journal of Science Education, 22,* 797–817.

Bereiter, C. (1995). A dispositional view of transfer. In A. McKeough, J. Lupart, & A. Marini (Eds.), *Teaching for mastery: Fostering generalization in learning* (pp. 21–34). Mahwah, NJ: Erlbaum.

Bereiter, C., & Scardamalia, M. (2006). Education for the knowledge age: Design-centered models of teaching and instruction. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (2nd ed., pp. 695–713). Mahwah, NJ: Erlbaum.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7–74.

Bransford, J., Brown, A., & Cooking, R. (Eds.). (1999). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academies Press.

Bransford, J. D., & Schwartz, D. T. (1999). Rethinking transfer: A simple proposal with multiple implications. In A. Iran-Nejad & P. D. Pearson (Eds.), *Review of research in education* (Vol. 24, pp. 61–100). Washington DC: American Educational Research Association.

Broudy, H. S. (1977). Types of knowledge and purposes of education. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the acquisition of knowledge* (pp. 1–17). Hillside, NJ: Erlbaum.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences, 2*(2), 141–178.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32–42.

Bybee, R. W. (2000). Teaching science as inquiry. In J. Minstrell & E. van Zee (Eds.), *Inquiring in inquiry learning and teaching in science* (pp. 20–46). Washington, DC: American Academy for the Advancement of Science.

Bybee, R. W., & DeBoer, G. E. (1994). Research on goals for the science curriculum. In D. L. Gabel (Ed.), *Handbook of research on science teaching and learning* (pp. 357–387). New York: Macmillan.

Campione, J. C., Shapiro, A. M., & Brown, A. L. (1995). Forms of transfer in a community of learners: Flexible learning and understanding. In A. McKeough, J. Lupart, & A. Marini (Eds.), *Teaching for transfer: Fostering generalization in learning* (pp. 35–68). Mahwah, NJ: Erlbaum.

Collins, A., Brown, J. S., & Holum, A. (1991). Cognitive apprenticeship: Making thinking visible. *American Educator, 15,* 38–46. Retrieved February 16, 2007, from http://www.kenton.k12.ky.us/DL/General/Readings/Collins_CogApp.pdf

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning and instruction: Essays in honor of Robert Glaser* (pp. 453–494). Hillsdale, NJ: Erlbaum.

Cowie, B., & Bell, B. (1999). A model of formative assessment in science education. *Assessment in Education, 6*(1), 101–116.

Davidowitz, B., & Rollnick, M. (2003). Enabling metacognition in the laboratory: A case study of four second year university chemistry students. *Research in Science Education, 33*(1), 43–69.

Derry, S. J., Seymour, J., Steinkuehler, C., Lee, J., & Siegel, M. (2004). From ambitious vision to partially satisfying reality: An evolving socio-technical design supporting community and collaborative learning in teacher education. In S. A. Barab, R. Kling, & J. Gray (Eds.), *Designing for virtual communities in the service of learning* (pp. 256–295). Cambridge, MA: Cambridge University Press.

Driver, R., Asoko, H., Leach, J., Scott, P., & Mortimer, E. (1994). Constructing scientific knowledge in the classroom. *Educational Researcher, 23*(7), 5–12.

Duschl, R. A., Schweingruber, H. A., & Shouse, A. W. (2007). *Taking science to school: Learning and teaching science in grades K-8.* Washington, DC: National Academies Press.

Engle, R. A. (2006). Framing interactions to foster generative learning: A situative explanation of transfer in a community of learners classroom. *Journal of the Learning Sciences, 15,* 451–498.

Etkina, E., Murthy, S., & Zou, X. (2006). Using introductory labs to engage students in experimental design. *American Journal of Physics, 74,* 979–986.

Etkina, E., & Van Heuvelen, A. (2007). Investigative Science Learning Environment—A science process approach to learning physics. In E. F. Redish & P. Cooney (Eds.), *PER-based reforms in calculus-based physics* (Vol. 1, pp. 1–48). College Park, MD: American Association of Physics Teachers. Retrieved May 8, 2008, from http://www.per-central.org/document/ServeFile.cfm?ID=4988

Etkina, E., Van Heuvelen, A., Brookes, D. T., & Mills, D. (2002). Role of experiments in physics instruction—A process approach. *The Physics Teacher, 40,* 351–355.

Etkina, E., Van Heuvelen, A., White-Brahmia, S., Brookes, D. T., Gentile, M., Murthy, S., et al. (2006). Developing and assessing student scientific abilities. *Physical Review Special Topics, Physics Education Research, 2,* 020103.

Ford, M. J., & Forman, E. A. (2006). Redefining disciplinary learning in classroom contexts. In J. Green & A. Luke (Eds.), *Review of research in education* (Vol. 30, pp. 1–32). Washington, DC: American Educational Research Association.

Gallagher, S. A., Stepien, W. J., Sher, B. J., & Workman, D. (1995). Implementing problem based learning in science classrooms. *School Science and Mathematics, 95*(3), 136–146.

Gee, J. P. (1999). *An introduction to discourse analysis: Theory and method.* New York: Routledge.

Germann, P. J. (1989). Directed-inquiry approach to learning science process skills: Treatment effects and aptitude-treatment interactions. *Journal of Research in Science Teaching, 26*(3), 237–250.

Goodwin, C. (1994). Professional vision. *American Anthropologist, 96,* 606–633.

Gourgey, A. F. (1998). Metacognition in basic skills instruction. *Instructional Science, 26,* 81–96.

Hammer, D., Elby, A., Scherr, R. E., & Redish, E. F. (2005). Resources, framing, and transfer. In J. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective* (pp. 89–120). Greenwich, CT: Information Age.

Harel, I., & Papert, S. (1991). *Constructionism: Research reports and essays, 1985-1990.* Norwood, NJ: Ablex.

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

Hmelo, C. E., Nagarajan, A., & Day, R. S. (2002). It's harder than we thought it would be: A comparative case study of expert-novice experimentation. *Science Education, 86*(2), 219–243.

Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review, 16,* 235–266.

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.

Hofstein, A., Shore, R., & Kipnis, M. (2004). Providing high school chemistry students with opportunities to develop learning skills in an inquiry-type laboratory: A case study. *International Journal of Science Education, 26*(1), 47–62.

Holton, G., & Brush, S. G. (2001). *Physics the human adventure: From Copernicus to Einstein and beyond.* New Brunswick, NJ: Rutgers University Press.

Kafai, Y., & Resnick, M. (Eds.). (1996). *Constructionism in practice: Designing, thinking and learning in a digital world.* Mahwah, NJ: Erlbaum.

Karelina, A., & Etkina, E. (2007). Acting like a physicist: Student approach study to experimental design. *Physical Review Special Topics, Physics Education Research, 3,* 020106.

Karplus, R. (1977). Science teaching and the development of reasoning. *Journal of Research in Science Teaching, 14*(2), 169–175.

Kindfield, A. (1993). Biology diagrams: Tools to think with. *Journal of the Learning Sciences, 3*(1), 1–36.

Kolodner, J. L. (2002). Facilitating the learning of design practices: Lessons learned from an inquiry into science education. *Journal of Industrial Teacher Education, 39*(3), 9–40.

Kolodner, J. L., Crismond, D., Fasse, B., Gray, J., Holbrook, J., & Puntembakar, S. (2003). Putting a student-centered learning by design curriculum into practice: Lessons learned. *Journal of the Learning Sciences, 12,* 495–547.

Kozma, R. B., & Russell, J. (1997). Multimedia and understanding: Expert and novice responses to different chemical representations of chemical phenomena. *Journal of Research in Science Teaching, 34,* 949–968.

Kuhn, D., & Pease, M. (2008). What needs to develop in the development of inquiry skills? *Cognition and Instruction, 26,* 512–559.

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

Lawson, A. E. (2000). How do humans acquire knowledge? And what does that imply about the nature of knowledge? *Science & Education, 9,* 577–598.

Lawson, A. E. (2002). The learning cycle. In R. G. Fuller (Ed.), *A love of discovery*: *Science education—The second career of Robert Karplus* (pp. 51–62). New York: Kluwer Academic.

Lawson, A. E. (2003). The nature and development of hypothetico-predictive argumentation with implications for science teaching. *International Journal of Science Education, 25,* 1387–1408.

Lippmann, R., & the Physics Education Research Group. (2002). *Analyzing students' use of metacognition during laboratory activities.* Retrieved February 16, 2007, from http://www.physics.umd.edu/perg/papers/lippmann/meta_lab.pdf

Lippmann Kung, R., Danielson, A., & Linder, C. (2005). *Metacognition in the students' laboratory: Is increased metacognition necessarily better?* Retrieved February 16, 2007, from http://www.anst.uu.se/rekun676/meta.pdf

Marton, F. (2006). Sameness and difference in transfer. *Journal of the Learning Sciences, 15,* 499–535.

Marzano, R. J. (2001). *Designing a new taxonomy of educational objectives.* Thousand Oaks, CA: Corwin Press.

Merrill, M. D. (2002). A pebble-in-the-pond model for instructional design. *Performance Improvement, 41,* 39–44. Retrieved March 12, 2008, from http://www.ispi.org/pdf/Merrill.pdf

National Assessment Governing Board. (2005). *Science framework for the 2005 National Assessment of Educational Progress.* Retrieved November 9, 2007, from http://www.nagb.org/publications/frameworks/s_framework_05/toc.html

Perkins, D. N. (1986). *Knowledge as design.* Hillsdale, NJ: Erlbaum.

Roth, W.-M. (1994). Experimenting in a constructivist high school physics laboratory. *Journal of Research in Science Teaching, 31*(2), 197–223.

Roth, W.-M., & Roychoudhury, A. (1993). The development of science process skills in authentic contexts. *Journal of Research in Science Teaching, 30*(2), 127–152.

Salomon, G., & Perkins, D. N. (1989). Rocky roads to transfer: Rethinking mechanisms of a neglected phenomenon. *Educational Psychologist, 24*(2), 113–142.

Scardamalia, M., & Bereiter, C. (1985). Fostering the development of self-regulation in children's knowledge processing. In S. F. Chipman, J. W. Segal, & R. Glaser (Eds.), *Thinking and learning skills: Research and open questions* (Vol. 2, pp. 65–80). Hillsdale, NJ: Erlbaum.

Schunn, C. D., & Anderson, J. R. (1999). The generality/specificity of expertise in scientific reasoning. *Cognitive Science, 23,* 337–370.

Schwartz, D. L., Bransford, J. D., & Sears, D. A. (2005). Efficiency and innovation in transfer. In J. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective* (pp. 1–52). Greenwich, CT: Information Age.

Schwartz, D. L., Brophy, S., Lin, X. D., & Bransford, J. D. (1999). Software for managing complex learning: An example from an educational psychology course. *Educational Technology Research and Development, 47*(2), 39–59.

Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging student production in statistics instruction. *Cognition and Instruction, 22*(2), 129–184.

Van Heuvelen, A., & Etkina, E. (2006). *Active learning guide.* San Francisco, CA: Addison Wesley.

Wilensky, U., & Reisman, K. (2006). Thinking like a wolf, a sheep or a firefly: Learning biology through constructing and testing computational theories—An embodied modeling approach. *Cognition and Instruction, 24*(2), 171–209.

Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching, 39*(1), 35–62.

Appendix A
Handout in a Design Lab for One of the Experiments During the Semester

Application experiment: The energy stored in the Hot Wheels launcher

The Hot Wheels car launcher has a plastic block that can be pulled back to latch at four different positions. As it is pulled back, it stretches a rubber band—a greater stretch for each of the four latching positions. Your task is to use the generalized work-energy principle to determine the elastic potential energy stored in the launcher in each of these launching positions.

Available equipment: Hot Wheels car, Hot Wheels track, Hot Wheels car launcher, meter stick, two-meter stick, ruler, masking tape, timer, scale to measure mass, spring scale.

Write the following in your lab report:

a) Start by making a rough plan for how you will solve the problem. Make sure that you use two methods to determine the energy. Write a brief outline of your procedure including a labeled sketch.
b) In the outline of your procedure, identify the physical quantities you will measure and describe how you will measure each quantity.
c) Construct free-body diagrams and energy and/or momentum bar charts wherever appropriate.
d) Devise the mathematical procedure you will need in order to solve the problem. Decide what your assumptions are and how they might affect the outcome.
e) Perform the experiment and record the data in an appropriate manner. Determine the energies.
f) Use your knowledge of experimental uncertainties to estimate the range within which you know the value of each energy.
g) Which rubrics should be used to evaluate your work? Please use them.
h) *What are the common features between this physics experiment and the estimation of the age of the Iceman in the homework reading? Make a comparison table.*

Appendix B
Handout for the Same Experiment as in Appendix A but for the Non-Design Lab

Energy stored in the Hot Wheels launcher: The Hot Wheels car launcher has a plastic block that can be pulled back to latch at four different positions. As it is pulled back, it stretches a rubber band a greater distance for each of the four latching positions. Your first task is to determine the elastic potential energy stored in the launcher in each of these four launching positions.

Procedure: Launch the car vertically into the air starting at one of the launching positions. When released, the car flies up into the air. By measuring the maximum height the car reaches, you should be able to decide the original elastic energy stored in the Hot Wheels launcher.

a) Measure the mass of the Hot Wheels car.

b) Hold the Hot Wheels car launcher so that it is oriented almost vertical—so the car does not fall out when placed in the launcher. Experiment a little with shooting the car almost vertically up into the air.
c) When ready to make quantitative measurements, place a meter stick beside the launcher and note the position on the meter stick of the front of the car when the car is ready to be launched.
d) Hold the launcher firmly and release it. Observe the highest position of the car. Subtract its initial position from this highest position to find the total vertical distance the car traveled.
e) Repeat this measurement three times. Take the average of the three vertical distance measurements and calculate the standard deviation of the measurements. Note: The standard deviation is calculated using the equation below:

$$\text{s.d.} = \sqrt{\frac{\Sigma(X_i - \bar{X})^2}{N-1}}$$

where X_i are the values of the four readings, $\bar{X}$ is the average of these four values, and $N = 3$ is the number of values being averaged.

f) Calculate the fractional uncertainty in the vertical distance measurement ($\Delta h/h$).
g) Repeat the measurements for the other three launching positions.
h) *Analysis:* Construct a work-energy bar chart for the process starting with the car resting on the stretched launcher and ending when the car is at its maximum elevation.
i) Apply the generalized work-energy equation for the process.
j) Insert your measurement numbers and determine the initial elastic energy of the launcher.
k) Calculate the fractional uncertainty of the elastic potential energy for each launching position—equal to the fractional uncertainty of the vertical distance traveled times the elastic energy for that launching position.

Appendix C
Four Versions of the Physics Interpretive Knowing Test Lab

Investigation of the behavior of a balloon

Equipment available: a balloon filled with helium, a balloon filled with air, meter stick, measuring tape, stopwatch, motion detector, electronic mass measuring scale that can be used to measure forces, computer, additional resources.

Version 1: You hold an air balloon and a helium balloon. Design experiments to determine which physical model best explains their motion if you release them: the model with no air friction, the model with viscous flow, or the model with turbulent flow.

Version 2: Design an experiment to determine whether a helium-filled balloon and an air-filled balloon have the same drag coefficients.

Version 3: Design and perform an experiment to determine the drag coefficient of the air balloon. Use this result to predict the speed of the helium balloon just before it reaches the ceiling. Then design and perform an experiment to determine this speed. Is the result consistent with your prediction?

Version 4: Design and perform an experiment to determine the drag coefficient of the helium balloon. Use this result to predict the speed of the air balloon just before it reaches the ground. Then design and perform an experiment to determine this speed. Is the result consistent with your prediction?

In your report describe the experiment, your analysis, and judgment so that a person who did not see you perform the experiment could understand what you did and follow your reasoning.