

Running Head: *ISLE* Transfer

Transfer of Scientific Abilities:

Building Habits of Mind

Eugenia Etkina*

Anna Karelina*

Maria Ruibal-Villasenor*

David Rosengrant◇

Rebecca Jordan*

Cindy E. Hmelo-Silver*

*Rutgers, The State University of New Jersey

◇Kennesaw State University

Correspondence to:

Eugenia Etkina

Graduate School of Education

Rutgers University

10 Seminary Place

New Brunswick, NJ 08901

etkina@rci.rutgers.edu phone: 1 732 932 7496 ext.8339

Abstract

Design activities when embedded in an inquiry cycle and appropriately scaffolded can promote the development and the transfer of the habits of mind 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 used by physicists to construct and evaluate scientific 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 cognitive abilities that prepare them to approach problems as scientists. A classroom study demonstrated that the students in the *ISLE* design lab performed equally well on traditional paper-and-pencil exams as *ISLE* students who did not engage in design activities. However, the design group significantly outperformed the non-design group while working on novel experimental tasks (in physics and biology) demonstrating the transfer of scientific “habits of mind”. This research demonstrates that a learning environment that seamlessly integrates cognitive apprenticeship, formative assessment, and preparation for future learning in a series of conceptual design tasks provides a rich context for learning that helps 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, Cocking, 1999; Brown, 1992; Derry, Seymour, Steinkuehler, Lee, & Siegel, 2004). It is particularly difficult when trying 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; rather as Bransford & Schwartz (1999) argue, it needs to prepare students for future learning (the PFL approach). In this paper we investigate whether engaging students in experimental design in such an environment affects student development of scientific habits of mind.

We explore experimental design because of the opportunity to involve students in genuine scientific practices that replicate the challenges of the real world and require innovative solutions to solve problems. Current recommended approaches to science instruction are based on a conception of science as inquiry and science learning as an apprenticeship in the practices of scientists (Duschl, Schweingruber, & Shouse, 2007). A central element of the work of scientists is the collaborative design of their own investigations. Important to this approach is that students are provided experimental problems which are appropriate at the level of student expertise and that require collaborative construction of their own solutions.

In order for students to develop innovative solutions, they must develop habits of mind that mirror authentic scientific practices (Duschl, Schweingruber, & Shouse, 2007; Ford & Forman, 2006). These habits of mind include 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 the ideas (Duschl et al., 2007). These habits are generative in nature and should promote interpretive knowing (Schwartz & Martin, 2004). Interpretive knowing is different from traditional understanding of knowing as replicative (information recalling) and applicative (using knowledge to solve problems). Interpretive knowing is the 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” (Broudy, 1977).

Participating in scientific practices can also provide students with resources for future learning, as becoming familiar with the disciplinary rules of the game would prepare them “to flexibly respond to a large set of complex problems instead of training them to use a narrow set of skills or strategies in limited circumstances” (Ford & Forman, 2006. p.26).

This study involves a system in which students have the opportunity to learn science by actively engaging in such practices (Etkina & Van Heuvelen, 2007). The Investigative Science Learning Environment (*ISLE*) (Etkina et al., 2007) is a complex, multifaceted intervention that synthesizes principles of cognitive apprenticeship with the methods of formative assessment and seeks to prepare students for future learning. In our interpretation of “preparation for future learning” we focus on the interpretive knowing (Bransford & Schwartz, 1999; Brody, 1977) that students use to approach novel 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, solving an experimental problem a scientist needs to decide what features of the problem are relevant and which can be ignored, how to represent the problem in different ways, including mathematical, how to use available equipment to collect necessary data, how to evaluate the quality of measurements, and how to make sense of the results.

Designing experiments often involves the students moving beyond direct application transfer (see discussion in Bransford & Schwartz, 1999). Instruction, therefore, should include the opportunity to develop interpretive knowing which should prepare them for future learning. Experimental design is an essential component of *ISLE* occurring mainly in the labs where students design their own experiments every week. 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 interpretive knowing and consequently prepare them for future learning. To test this hypothesis, we examine the effect of removing the design component of the *ISLE* system as we address 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 the designing their own experiments affect students' approaches toward experimental inquiry?
3. How does designing their own experiments affect students' development of experimental procedures, processes and methods?
4. Does the devising, designing, and conducting their own experiments affect students' acquisition of science concepts?
5. How does engaging in design affect students' ability to construct *interpretive knowing* and transfer it to new situations?

In the paper we will first describe the theoretical foundations of *ISLE* and design labs, provide details of the *ISLE* design labs, then 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. Design helps create rich contexts for learning (Harel & Papert, 1991). Through engaging in design, learners become more accountable for their learning through designing, 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). Design requires students to engage in metacognitive thinking and scientific discourse (Davidowitz & Rollnick, 2003; Driver, Asoko, Leach, Mortimer, & Scott, 1994; Germann, 1989; Gourgey, 1998; Hofstein, Shore, & Kipnis, 2004). First, metacognitive thinking such as planning, monitoring and evaluating is a requisite because without it, one cannot manage investigative tasks. Second, scientific discourse refers to 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, 1993).

There are multiple successful demonstrations that 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). Designing and carrying out investigations helps engage learners in important cognitive and social activities that promote the development of interpretive knowing. Perkins (1986) defended the view of knowledge as design: in this view knowledge is a structure that has to be built and adapted for specific purposes.

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

learning. But, design is a complex undertaking and requires scaffolding as might be provided through a cognitive apprenticeship in scientific practices.

Investigative Science Learning Environment

ISLE is a learning system that was developed to for large or small enrollment introductory non-physics 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 these science majors learn fundamental physics concepts but most importantly 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, 1998; Hmelo, Holton, & Kolodner, 2000; Hmelo-Silver, 2004; Kolodner, Crismond, Fasse, Gray, Holbrook, & Puntembakar, 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. recitation or a three-hour 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 to help students not only learn the concepts and laws of physics but most importantly 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 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 not to prove it and finally apply new knowledge to solve practical problems (Fig. 1). The *ISLE* cycle is a blend of Karplus' learning cycle (1977) and Lawson's science cycles (2002).¹

One of the most important *ISLE* features is that students learn collaboratively in large room meetings, problem solving sessions and laboratories. In this paper we will mostly focus on the *ISLE* laboratory environment (Etkina, Murthy & Zou, 2006).

Theoretical foundation for *ISLE* labs

The theoretical foundation for *ISLE* labs is based on a close integration of three pillars: Preparation for future learning, cognitive apprenticeship and formative assessment (Fig. 2). Based on these pillars, *ISLE* labs engage students in three types of activities. First, learners are engaged in generative activities such as design and reflection. Second, cognitive apprenticeship provides expert models to scaffold design and reflection. Third, we make this expertise visible and provide feedback to the students through formative assessment. In combination, these activities promote interpretive knowing and thus preparation for future learning.

Preparation for Future Learning and Interpretive Knowing

Preparation for future learning (PFL) (Bransford & Schwartz, 1999; Schwartz & Bransford, 1998; Schwartz, Bransford, & Sears, 2005) is one of the goals of *ISLE* labs. Here PFL refers to student's ability to approach new phenomena and problems similar to how a scientist would approach them. For example in one lab, students needed 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. A scientist, however, immediately realizes that one experiment would not be sufficient (different experiments that will produce similar results are needed). Further a scientist knows that the physics laws that she/he will use to analyze collected data are mathematical models of some ideal situation and they need to make sure that their experimental set up satisfies the mathematical model criteria. Finally a scientist keeps in mind that any result that she/he obtains has experimental uncertainties due to the instruments and procedures they use (Alberts, 2000). Moreover scientists would notice that this task has particular features about which they need to think while solving the problem. The goal of *ISLE* labs to help students develop such habits of mind and consequently to interpret experimental and conceptual problems in ways similar to that of scientists.

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 particularly ways of perceiving the world for certain purposes. 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. In previous studies of transfer within the PFL paradigm, researchers have focused on the development of some concrete piece of *interpretive knowing*. In one study, students learned to see the statistical properties of a situation by inventing mathematical ways to describe the reliability of a pitching machine; this prepared students to construct the idea of variability of data sets (Schwartz and Martin, 2004).

In another study, college students' looked for patterns in the kinds of information that people are able to recall and then they represented these patterns in graphs. These generative activities prepared students for constructing their understanding of psychological principles of memory after attending a lecture on the topic (Schwartz & Bransford, 1998). In this study, when we talk about PFL, we are not only referring to the development of some specific interpretive knowing that helps students pay attention to certain relevant features of a particular problem but to the students' development of a whole new, "scientific" way of looking at the world.

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 (Brown, Collins, & Duguid, 1989; Collins, Brown & Newman, 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, nature of AIDS, etc). In addition, other forms of modeling are provided as instructors demonstrate hypothetico-deductive reasoning, construct representations, devise mathematical procedure, 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 the means of prompts and questions given to the students on the lab handouts, (c) instructors' feedback, and (d) 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 focuses students' attention on the important aspects of the design process and simultaneously help make their thinking about the salient elements of design visible. Because we are introducing students to the practices of real science, we need to provide them with 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 and answer the reflection questions given in laboratory handouts and when they write their individual lab reports. *ISLE* labs have embedded reflection and connection-building activities because they are critical for transfer, inform both direct application and PFL perspectives (Bransford & Schwartz, 1999).

In addition we provide scaffolding for students' inquiry activities by introducing the processes of physics in a structured and simplified manner. The *ISLE* cycle (Fig. 1) helps communicate the process of scientific inquiry in physics and it gives structure to the series of the 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 to conceptualize lab experiments as different variations of these three. This procedural facilitation is essential since 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, but also it helps them to perceive their learning as an integrated process, and to understand more 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, et al., 2006a). These rubrics serve simultaneously two purposes. They provide the modeling and coaching aspect of the cognitive apprenticeship and 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. 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 when solving experimental. Scientific abilities are the habits of mind that scientists used when they approach new problems; they are cognitive tools that scientists “know with”. These abilities help interpret the situation in ways that are characteristic of science, thus they are components of interpretive knowing. Helping students develop these abilities will mean moving them closer to developing scientific interpretive knowing. We prefer the term scientific abilities instead of science-process skills to underscore that these are not automatic skills, but are instead processes that students need to use reflectively and critically (Salomon & Perkins, 1989).

Scientific abilities 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 was constructed based on the analysis of the history of the practice of physics (Holton & Brush, 2001; Lawson, 2000;

Lawson, 2003), research on the goals of science curricula and taxonomies of cognitive skills (Bybee & DeBoer, 1994; Marzano, 2001); recommendations of cognitive scientists (Schunn & Anderson, 2001) 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, for 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. Rubrics serve as a tool that helps students develop these abilities.

The scientific abilities rubrics serve as a tool for procedural facilitation and feedback (an example of a rubric is in Fig. 3)² as they help novices to complete inquiry tasks, which require complex and unfamiliar scientific abilities. At the beginning of the semester students are told which ones are the appropriate rubrics for self assessment for each particular experimental task, towards the end of the semester this scaffolding fades as they need to decide themselves which rubrics from the available set 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 in the next lab holds a discussion about students' successes and weaknesses. Using the rubrics helps prepare students to learn from the instructor explanations and subsequent lab discussion (Bransford & Schwartz, 1999; Schwartz, and Martin, 2004). At the same time it provides 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 summary, the three theoretical pillars make *ISLE* laboratories 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' construction of their knowledge (Harel & Papert, 1991).

An Investigation of the Role of Design

To investigate the causal importance of design and reflection, we created an alternative *ISLE* non-design lab to address whether design affects students' development of scientific abilities. In this non-design alternative, students generate a smaller set of cognitive products. They do not generate interpretations of data, hypotheses, predictions, or experimental procedures. They do not use rubrics to self assess themselves and do not have to engage in constant metacognitive reflection. We hypothesize that this will result in an impoverished laboratory experience and diminish subsequent learning and transfer compared with students who participate in the *ISLE* design labs.

Participants

There were 186 students enrolled in the *ISLE* physics class (the number of students attending different activities varied through the semester). There were 72 male students and 114 females who signed up for different sections of the course. During the semester, students were not informed about the study. At the end, 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 an introductory physics course for science majors. There were two 55-min. lectures or large room meetings, one 80-min problem solving session, and a 3-hour 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. There were eight problem-solving sections and eight lab sections. Students signed up for lab sections and problem solving sessions separately, thus students that were in the same lab section were not in the same problem solving section. There were three TAs in the lab sections; one TA taught four sections and the other two taught two sections each. All three TAs were highly skilled in interactive teaching with their teaching experience ranging from 4 to 7 years of teaching.

For the purpose of the study the lab sections were split into two groups: design group (four lab sections) and non-design group (four lab sections).³ Two of the TAs taught one design and one non-design section and the third TA taught two of each. To ensure that the design and non-design lab sections were equivalent, we administered Lawson's test of hypothetico-deductive reasoning (Lawson, 1978) in the first lab session. Coletta and Philips (2005) found that student's learning gains are strongly correlated with their scores on this test. Students in the two groups did not differ (Design mean=11.04, SD= 3.83; control mean 10.95 SD= 2.96, $F(1, 185) = 0.06, p = 0.81$).

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, while the material covered in the labs, 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 they had already discussed in large-room meetings and problem-solving sessions. In one-third of the labs they designed an experiment to investigate a new phenomenon and find patterns. The scaffolding was provided through lab handout questions that focused their attention on the elements of the scientific process: representing the situation, deciding on the experiment, analyzing experimental uncertainties, etc.

Although scaffolding was provided, students struggled first in coming up with possible designs, and then in actually implementing them and evaluating the results. They had to use different representations such as force diagrams, energy bar charts to help them 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 received a certain value through the experiment and calculation, they could not ask the TA 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 were asked to compose their written lab report to describe what they did and to answer the questions in the lab handout (see Appendix 1 for an example). Students could consult relevant self-assessment rubrics to improve their reports as they were writing or at the end.

The TAs did not help students design experiments and when students had difficulties, they asked questions and provided hints but did not answer their questions directly. Usually a lab had two experiments: one that the students had to design from scratch choosing the equipment, putting it together and devising the mathematical procedure and the other one where the

experimental set up was provided but they had to invent their own procedure. They wrote the reports during the 3-hour lab period and handed them in before they left.

After each experiment students were asked to reflect on the purpose of the experiment and its place in an overall scientific process. Reflection questions focused student attention on contrasting cases such as the difference between an observational experiment and testing experiment (Etkina et al., 2002). 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 a 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 that students did after each lab contained reading passages with reflection questions. Students analyzed stories about historical developments of several scientific theories and applications such as the nature of AIDS, prophylaxis, and pulsars. They had to identify the elements of scientific inquiry that are present when scientists answer new questions or apply knowledge. The purpose of the passages was to help students reflect on the common elements of a scientific investigation.

Non-Design Labs. Students in the non-design labs used the same equipment as 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 but they did not need to think about theoretical assumptions as these were provided to them in the text. These labs were not cookbook labs; we call them non-design labs. These labs had homework as well —mostly physics problems that prepared students to do the next lab. The TAs taught the labs differently.

They provided an overview of the material at the beginning of the lab and then later if students had questions, they answered these questions. In Appendix 1 we provide examples of lab handouts for design and non-design groups.

Procedure for Transfer Tasks

To assess how students transfer scientific abilities to an unfamiliar physics content in the same functional context (according to the classification by Barnett and Ceci, 2002), we developed a lab task where both groups designed an experiment and wrote a lab report. In contrast to regular labs, which students performed during semester, this particular task was identical for the design and the non-design groups. The task involved drag force in fluid dynamics. This physics content was not covered in the course. To minimize the spreading of information among the students we developed four similar versions (Appendix 2). 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 task during the lab (3 hours) on week 13 of the semester. Prior to this, they had performed 10 labs. The lab sections were spread from Wednesday to Friday. Four design sections had labs on Wednesday and Thursday morning; non-design sections had the lab on Thursday afternoon and Friday morning. Thus it is possible that students in design sections shared information with the students in non-design sections. The drag force lab was attended by 89 students in design sections and by the same number of students in non-design sections.

The second transfer experiment involved a biology task that was given as the final lab exam for the course in week 14. 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 to 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 to complete for students with very little plant physiology background; (b) students can use multiple measures to determine transpiration rates which gave them some room for inventiveness, evaluation and decision making; and (c) students are more willing to accept a biology assignment as a final exam for their physics lab if they perceive that there is a physical basis (evaporation and osmosis) underlying the biological process of transpiration.

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

Assessment of Physics Knowledge and Scientific Abilities

In Table 1 we list the data sources that we used to answer the research questions. Below 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 of the time spent by a group of students on different activities (Karelina & Etkina, 2007).

A trained observer sat with a group of students for a 3-hour lab. 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: making-sense, logistic, and off-task. According to Lippmann, during sense making episodes students are talking to each other, working on figuring

out the answer, and holding a coherent conversation. During the logistic mode students gather equipment, operate equipment, collect data, read, and write. An off task mode involves the time intervals when students are not directly engaged in the lab task.

In addition to Lippmann's 3-item coding scheme we used a code for writing and for TA help (see table 2). Observations for two years conducted prior to this study of more than 30 lab groups showed that all behaviors fit into one of the five coding categories. To establish the inter-rater reliability of the coding scheme for this study, two observers independently coded four of the first lab observations (20% of the observations). They achieved an 84% agreement on the codes before the discussion and 100% after the discussion. After the reliability was established, one observer observed student lab groups. The observer timed and recorded one design group and one non-design group each week, observing 20 three-hour labs.

Observations of student behavior during transfer labs. During the transfer labs we observed one group per lab section: in total collecting observations of four student groups from the design lab sections and four student groups from 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 TA) using the same scientific abilities rubrics that the students used during the labs for self-assessment. For each lab all three scorers independently scored 2-3 students' lab reports using the chosen rubrics. Then they discussed the discrepancies in the scores to make sure that the details of the particular labs are taken into account. Then they scored additional 7-10 randomly chosen lab reports (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 a 100% agreement after the second scoring. Then each rater scored an additional 15-17 reports.

Physic knowledge. Regular exams consisted of multiple choice and open-ended questions. Fifty-five minute midterms had 11 questions. Eight of them were multiple-choice (equally split between conceptual and quantitative) and three were open-ended problems. A 3-hour final exam had 18 questions, 12 of them were multiple choice (nine quantitative and three conceptual) and 6 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 in charge.

Rubric scores for the transfer labs. During the physics transfer lab and biology transfer lab students in each lab section worked in the same group of three or four as they did during the semester and submitted individual reports for grading. The four design sections had both transfer 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 the reliability of the scores as described above. The rubrics chosen for scoring were for the following 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; the ability to communicate.

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 3. To inspect for differences among groups for each of the coding schemes listed in Table 2, we conducted a two-way analysis of variance (ANOVA) between design and non-design groups

across the coding scheme. Differences between groups (i.e, main effect) were noted ($F(1, 108) = 45.16, p < 0.001$). In addition, differences in lab activity (i.e., treatment) were also noted ($F(5,108) = 40.12, p < 0.001$); too few observations in the reading category (listed in Table 2) caused us to eliminate this category from the analysis. Inspecting the data, these differences were most evident between sense-making and overall time spent in lab (Table 3). In general, the design students spent greater time sense-making in lab and this difference became more evident toward the end of the semester with the non-design students significantly decreasing the amount of time they spent sense-making with each lab (Fig. 5). While the design students on average spent close to three hours in lab throughout the semester, the non-design students spent much less time and this decreased, on average, toward the end of the semester.

By the end of the semester, on average, the design students were spending much more time in the lab and on sense-making (Table 4) compared to the non-design students. Although the lab was 3-hour long, non-design students chose to leave early.

How does Design Affect Students' Approaches Toward Experimental Inquiry?

The observations showed that there was a large difference in the behavior of design and non-design students during the physics transfer lab and the bio transfer lab. The pattern described above, which emerged during regular labs, persisted during the transfer labs although during the transfer labs, both groups had to design experiments with no scaffolding. Significant differences were noted between groups and activities in the physics transfer lab (Two-way ANOVA, among groups; $F(1,42) = 14.33, p < 0.001$; among treatment: $F(6, 42) = 130.39 p < 0.001$). In the biology transfer lab, differences between groups were not significant but differences among treatment were (among treatment were significant but differences among group was not (Two-

way ANOVA, between group $F(1, 42) = 3.70, p = 0.061$; among treatment $F(6, 42) = 141.80, p < 0.001$).

Inspecting more closely, the physics transfer lab took more time for design students. The design groups spent over 40 minutes more time in the lab room than non-design students. The difference between the duration of the physics transfer lab was 162 ± 17 (SD) min design versus 120 ± 25 min non-design. The difference in bio transfer lab duration was also noted: 176 ± 26 min design versus 153 ± 26 non-design.

Design students spent considerable time doing sense making in both labs; non-design students spent little time in both labs. Figure 6 shows differences in sense-making discussions. The sense-making lasted a mean of 52 ($SD=7$) minutes and a mean of 42 ($SD=7$) minutes in design groups in the physics and bio labs respectively; whereas the means were 15 ($SD=3$) and 19 ($SD=3$) minutes in non-design groups. The time that students spent on other activities was more similar for both groups.

How does Design Affect Students' Development of Scientific Abilities?

In Figure 7 we present the scores of the students in design sections on the relevant abilities at the beginning and the end of the semester. Students in design sections improved their performance on the abilities chosen for assessment.

Does Design Impede Students' Acquisition of Normative Science Concepts?

When considering the normative science concepts that were tested via multiple-choice and free-response questions and problems, students performed similarly in the design and non-design groups on both midterms and final exam, shown in Table 5. (ANOVA; Midterm exam 1, $F(1,182) = 0.25, p = 0.616$; Midterm exam 2, $F(1, 190) = 1.31, p = 0.253$; Final exam, $F(1, 190) = 0.45, p = 0.502$; After 3 contrasts Sequential Bonferonni correction critical value all

$p > 0.017$).

How does Design Affect Students' Ability to Construct Interpretive Knowing and Transfer it to New Situations?

Design students demonstrated significantly better scientific abilities than the non-design students did as shown in Figures 8 to 11. We aggregated the data for rubric scores 2 and 3 as they represent different levels of proficiency level as opposed to 0 (missing) and 1 (inadequate). A significant number of design students got scores of 2 or 3 for evaluating assumptions in their lab reports (Figure 8a). They were able to identify relevant and significant assumptions of the theoretical model that they used, whereas only a few non-design students did ($\chi^2(3, N=178) = 68, p < 0.001$ for the physics transfer lab and $\chi^2(3, N=181) = 120, p < 0.001$ for the biology transfer lab). About half of design students evaluated the effects of assumptions on the result or validated them in both labs. No students in non-design section made an attempt to do this (Figure 8a).

During the semester, non-design students were to learn how 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 it. Only few of these students, however, used this ability in the independent experimental investigation (Figure 8b). More than 50% of design students evaluated the effect of experimental uncertainties in this lab. The difference in the number of students who evaluated uncertainties is statistically significant with $\chi^2(3, N=178) = 30, p < 0.001$ for the physics transfer lab and $\chi^2(3, N=181) = 94, p < 0.001$ for the bio transfer lab.

A high score on the rubric “evaluating the results by means of an independent method” is possible only when a student discusses the discrepancy between the results of two methods and

possible reasons of this discrepancy considering assumptions and uncertainties. We found that design students demonstrated a higher ability to evaluate the result (Fig. 8c). We can see that 72% of design students received score 2 or 3 for their reports in the physics transfer lab, i.e. discussed the reasons for the discrepancy. In non-design sections only 43% students did ($\chi^2(3, N=178) = 16, p < 0.001$). In bio transfer lab 79% of the design students discussed the difference versus 37% non-design ($\chi^2(3, N = 181) = 42.25, p < 0.001$).

One of the main scientific abilities we want students to develop is an ability to communicate their ideas. This includes an ability to draw diagrams and pictures, describe details of the procedure, and to explain the methods. The analysis of lab reports showed that more than 60% of design students drew a picture while only 8% of non-design students did. Figure 8d shows the results of the scoring of the reports using the communication rubric. The difference in their scores is statistically significant ($\chi^2(3, N = 178) = 30, p < 0.001$) for the physics lab and [$\chi^2(3, N = 181) = 41.65, p < 0.001$] for the bio lab.

The analysis of the students' reports for the physics transfer 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). In mechanics such tools are force diagrams. Students from different sections demonstrated a different quality of drawing force diagrams 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. mislabeled or not labeled force vectors, wrong directions, extra incorrect vectors present, or vectors missing), while only 2% of design students

made a mistake in force diagrams. This difference is statistically significant ($\chi^2(3, N = 178) = 18, p < 0.001$).

In addition, we analyzed another ability – the ability to construct consistent representations (force diagram versus mathematics, a picture versus a free body diagram, etc). 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 ($\chi^2(2, N = 178) = 7.8, p = 0.02$).

We also analyzed the lab reports written by design students during regular semester labs and their lab report written for the physics and bio transfer labs to find how the abilities develop during the semester and how students apply them in novel situations. Here we show how students acquire the ability to consider assumptions, and how they apply it in independent investigation without scaffolding. Figure 9 shows the students' lab reports scores for ability to consider assumptions in the theoretical model for six regular labs and two transfer labs. The number of students who identified assumptions in their lab reports increases during the semester and reaches about 80% in the labs at the end of the semester.

To illustrate the differences in student lab reports in the design and non-design groups in both transfer labs, we provide the examples of two representative lab reports from both groups annotating them using the rubrics (Figure 10). These two excerpts demonstrate the differences in the students' approach. The non-design student writes step-by-step instructions copying the handout style, but does not provide a labeled sketch of a set-up and force diagram, while the design student does. Also the non-design student does not give any explanation or justification of the experimental procedure. The design student explains the physical processes of the experiment and gives a mathematical model of these processes. In addition, this student

identifies limitations and assumptions in this model and tries to evaluate how these assumptions may affect the result of the experiment. The non-design student provides a correct mathematical description of the physical process but does not consider inherent assumptions. This student uses this model to evaluate the required drag coefficients of air filled and helium filled balloons, but these results do not incorporate uncertainties. Note here that the non-design student emphasizes the necessity to use multiple trials but does not demonstrate the understanding that this is necessary to evaluate and minimize the uncertainty. This student seems to simply repeat the rote of the regular labs. Since the non-design student does not consider the uncertainty of the measurements, his/her judgment is not justified. It is impossible to say whether drag coefficients for air and helium balloons are different. On the contrary, the design student makes a judgment based on the uncertainty analysis and attempts to consider the effects of assumptions. In general, the design lab report demonstrated a more sophisticated approach to the same investigation when compared to the non-design report.

Discussion

The results of the study support our hypothesis that the design element from *ISLE* labs enriches students' learning opportunities. Students who were in design lab sections approached new tasks in biology and physics in more scientifically productive ways than the students who did not experience design labs during the semester. In particular, increased student attention to the measurement, assumptions in the mathematical procedure, evaluation of the results, and the ability to communicate were evident. These interpretive ways of knowing were most likely encouraged by the design and undertaking of experiments and the embedded reflection.

We contend that it is the three theoretical pillars: preparation for future learning, cognitive apprenticeship and formative assessment of the *ISLE* design labs that account for group

differences. First, 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 be successful in it by following several strategies. The structure of the tasks and the instructors make scientific thinking visible to the students. The complex abilities such an ability to test a hypothesis 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. The *ISLE* design labs provided scaffolding to learners' work and progress. We were able to slowly remove scaffolding as students became more independent and began to develop scientific habits of mind. These labs integrate formative assessment by students' ongoing assessment of their progress and by continuous feedback and adjustment of instruction in order to respond to students' needs. The design labs prepare for future learning as instructional practices support students in the development of the types of knowledge that are more useful when handling new situations.

More concretely, we search for the answers to the following five questions: How does the need to design their own experiments affect the types of activities in which learners engage? How does the designing their own experiments affect students' approaches toward 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? How does engaging in design affect students' ability to construct *interpretive knowing* and transfer it to new situations?

With respect to the first question, we found that lab activities elicit 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 it was significantly larger than the time for non-design students. Both groups of design and non-design students started the semester spending about the same 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 in reasoning exchanges. Previous studies such as the one by Hmelo, Nagarajan, and Day (2002) have shown that authentically complex tasks compel students to engage in monitoring and reflecting. Our previous studies show that in introductory labs in traditionally taught courses students spend about 15 min on sense-making on average (Karelina & Etkina, 2007) which is much less than 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 what is required in labs, 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.

The second question that we sought to answer was how design of experiments affects the way students approach experimental inquiry. Design students spent significantly more time on sense-making than non-design students. This seems to indicate that the design labs support 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 we give students a task that requires sense-making, students do not engage in it. The design students may continue with sense-making even when there are 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 for their group mates while non-design students got used to following directions without questioning them. Thus when they encounter a task with no directions, 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 they improved significantly throughout the semester, but not all to the same degree. We discovered that after ten 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 the ability to evaluate their experimental results by means of an independent method.

At the same time, we found that students' learning of normative physics content did not suffer when they designed their own experiments taking into account that they spend more time writing their lab reports and thinking about scientific procedures versus solving physics problems. More significantly, their learning of physics did not suffer even though in some occasions their devised procedures were not optimal or their experimental results were incorrect. We make this statement because 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 the exams. Therefore, we conclude that with the design experiments, students learn more than without them. They achieved similarly on exams than those students who are not assigned design

tasks and, in addition, they develop more productive scientific habits of mind. Learning the scientific habits of mind is as important as learning science concepts (Duschl, Schweingruber, & Shouse, 2007).

Finally, we also found that when students engaged in the design of experiments, their capacity to construct interpretive knowing and to transfer this knowing in the attempt to solve new problems increases significantly. By analyzing students' lab reports of their physics and biology "transfer" investigations, we determined that students in the design group have acquired and applied some scientific habits of mind necessary to generate relevant, new knowledge. These habits of mind refer to cognitive steps, procedural moves and decisions required by the highly complex process of designing an experiment. According to Baker and Dunbar (1996), scientists possess a variety of rich experimental schemas that consist of the general properties in designing different types of experiments. 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 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

In this study we regard transfer as the enculturation in the scientific community of

practice with their characteristic ways of framing and interpreting problem situations. For example, when scientists encounter a problem that involves some type of measurement, they automatically think of different kinds of measurement uncertainties inherent in this particular measurement. This ability to think about certain aspects of a problem is characteristic for different disciplines and this is what we call interpretive knowing. Therefore helping students see these aspects and attend to them 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.

We hypothesize that interpretive knowing and the concept of cognitive resources (Hammer, Elby, Sherr & Redish, 2005) are connected. One of the goals of science education is to help students learn to activate the 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. Possibly, that subsequent activation of the same group of resources helps form a habit of using them together. We can interpret the results of our study in a way that students who were enrolled in design labs learned to activate some of the resources that scientists would activate when faced with the same problem without any prompts. We found that the multi-faceted intervention in the design labs affords students a new way of perceiving the experimental tasks.

Much of the previous research on transfer has focused on specific kinds of knowledge or procedures rather than habits of mind important in scientific inquiry. Unlike the prior research, we have demonstrated that these habits of mind transfer across scientific domains. But this level of transfer does not come easily. Rather we argue that it is the synthesis of the three pillars on the

basis of which *ISLE* labs were developed that is important in providing the kinds of experiences that prepare students for future learning and transfer. The cognitive apprenticeship helps provide the support that is needed for learners to engage in the complex tasks of designing, conducting, and interpreting experiments. The formative assessments scaffolds students in thinking about what was important for design and promotes student meta-cognition as students 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 construct interpretive knowing that enables them to look at lab tasks with scientific eyes; therefore they are prepared for future learning. In addition to the three pillars, we provide another example of the affordances of design tasks to provide a rich context for learning. Unlike constructionist theory that emphasizes constructing a concrete artifact (e.g, Kafai & Resnick, 1996), we have demonstrated that designing a conceptual object supports the development of habits of mind, as Bereiter and Scardamalia (2006) suggest is essential to creating a knowledge building society.

Author Notes

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 in representing the data, and Noah Finkelstein for productive discussions. We are indebted to Alan Van Heuvelen for writing non-design labs, providing advice during the project, and providing comments on this paper. We also thank NSF for its continuous support of our work. The project was supported by NSF grant DRL 0241078. Conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

Footnotes:

¹ For details see in Etkina and Van Heuvelen (2007) and Van Heuvelen and Etkina (2006).

² A complete set of rubrics can be found at <http://paer.rutgers.edu/scientificabilities>

³ Problem solving sections did not correspond to the lab sections; students from one lab section could be in different problem solving sections and vice versa.

References

- Alberts, B. (2000). Some thoughts of a scientist on inquiry. In J. Minstrell and E. van Zee (Eds.) *Inquiring in Inquiry Learning and Teaching in Science* (pp.3-13), Washington, DC: AAAS.
- 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*. La Jolla, CA: Lawrence Erlbaum Associates.
- Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, 128(4), 612-637.
- Barron, B. (1998). Doing with understanding: Lessons from research on problem- and project-based learning. *Journal of the Learning Sciences*, 7(3-4), 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(8), 797-817.
- Bereiter, C., & Scardamalia, M. (2006). Education for the knowledge age: Design-centered models of teaching and instruction. In P. A. Alexander & P. H. Winnw (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), pp. 7-74.
- Bransford, J., Brown, A., & Cooking, R. (Eds.). (1999). *How people learn: Brain, mind, experience, and school*. Washington DC: National Academy Press.

- Bransford, J. D. & Schwartz, D. T. (1999). Rethinking transfer: A simple proposal with multiple implications. In *Review of Research in Education* edited by A. Iran-Nejad and P.D. Pearson, Washington DC: AERA, 24, pp. 61-100.
- 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, 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 and E. van Zee (Eds.) *Inquiring in Inquiry Learning and Teaching in Science* (pp.3-13), Washington, DC: AAAS.
- 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, NY: Macmillan.
- Coletta, V. P. & Phillips, J. A. (2005). Interpreting FCI scores: Normalized gain, preinstruction scores, & scientific reasoning ability. *American Journal Physics*, 73(12), 1172-1182.
- Collins, A., Brown, J. S. & Holum, A. (1991). Cognitive apprenticeship: Making thinking visible. *American Educator*, 15(3). 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: LEA.
- Cowie, B., & Bell, B. (1999). A model of formative assessment in science education. *Assessment in Education*, 6, 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, 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., Mortimer, E., & Scott, P. (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 Academy Press.
- Etkina, E., Murthy, S., & Zou, X. (2006). Using introductory labs to engage students in experimental design. *American Journal of Physics*, 74 (11), 979-986.
- Etkina, E., Van Heuvelen, A., White-Brahmia, S., Brookes, D.T., Gentile, M., Murthy, S. Rosengrant, D., and Warren, A. (2006) Developing and assessing student scientific abilities. *Physical Review. Special Topics, Physics Education Research*. 2, 020103.

- Etkina, E., & Van Heuvelen, A. (2007). Investigative Science Learning Environment – A Science Process Approach to Learning Physics. In E. F. Redish and P. Cooney (Eds.), *PER-based reforms in calculus-based physics*. College Park, MD: AAPT.
- Etkina, E., Van Heuvelen, A., Brookes, D. T. & Mills, D. (2002). Role of experiments in physics instruction – A process approach. *The Physics Teacher*, 40(6), 351-355.
- Ford, M. J. & Forman, E. A. (2006). Redefining disciplinary learning in classroom contexts. *Review of Research in Education*, 30, 1-32.
- Gallagher, S. A., Stepien, W. J., Sher, B. J., & Workman, D. (1995). Implementing problem based learning in science classrooms. *School Science and Mathematics*, 95, 136-146.
- Gee, J. P. (1999). *An introduction to discourse analysis: Theory and method*. New York, NY: 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.
- 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. 1-52). Greenwich, CT: Information Age Publishing.
- Harel, I., & Papert, S. (1991). *Constructionism: research reports and essays, 1985-1990*. Norwood, NY: Ablex Publishing Corporation.

- Hmelo, C. E., Holton, D. L., & Kolodner, J. L. (2000). Designing to learn about complex systems. *The Journal of the Learning Sciences*, 9(3), 47-298.
- Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review*, 16, 235-266.
- 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, 219-243.
- 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, 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: Lawrence Erlbaum Associates.
- Karelina, A. & Etkina, E. (2007). Acting like a physicist: Study of Student approach to experimental design. *Physics Review, Strand Physics Education Research*, 3.
- Karplus, R. (1977). Science teaching and the development of reasoning. *Journal of Research in Science Teaching*, 14, 169-175.
- Kindfield, A. (1993). Biology diagrams: Tools to think with. *The Journal of the Learning Sciences*, 3, 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), 1-31.

- 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. *The Journal of the Learning Sciences*, 12(4), 497-547.
- 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(9), 949-968.
- 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(6), 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*. New York: Kluwer Academic Publishers.
- Lawson, A. E. (2003). The nature and development of hypothetico-predictive argumentation with implications for science teaching. *International Journal of Science Education*, 25(11), 1387-1408.
- Lippmann, R., & the Physics Education Research Group. (2002). Analyzing students' use of metacognition during laboratory activities. *AREA Meeting (New Orleans, 2002)*. 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. *EARLI symposium (2005)*. Retrieved February 16, 2007 from <http://www.anst.uu.se/rekun676/meta.pdf>.

- Marton, F. (2006). Sameness and difference in transfer. *The Journal of the Learning Sciences*, 15(4), 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(7), 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/frameworks/s_framework_05/toc.html
- Perkins, D. N. (1986). *Knowledge as design*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Salomon, G. & Perkins, D. N. (1989). Transfer: Rethinking mechanisms. *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, and R. Glaser (Eds.), *Thinking and Learning Skills: Research and Open Questions*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Schunn, C. D., & Anderson, J. R. (1999). The generality/specificity of expertise in scientific reasoning. *Cognitive Science*, 23(3), 337-370.
- Schwartz, D. L., & Bransford, J. D. (1998). A Time for Telling. *Cognition & Instruction*, 16, 475-522.
- 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 Publishing.

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*, 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 & Instruction, 24*, 171-209.

Appendix 1:

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 force 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? Make a comparison table.*

Handout for the same experiment in 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 four times. Take the average of the four 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 $N = 4$ 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 2: Four versions of the physics transfer task.

Lab task: Investigation of the behavior of the balloon

Equipment available: *a balloon filled with helium, a balloon filled with air, meter stick, measuring tape, stop watch, 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.

Table 1

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 the designing their own experiments affect students' approaches toward experimental inquiry?	Observations of student behavior during the physics transfer lab and bio transfer lab (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 the devising, designing and conducting their own experiments affect students' acquisition of science concepts?	All students' scores on regular exams that included multiple choices and open-ended questions (2 midterms and one final exam).
How does engaging in design affect students' ability to construct <i>interpretive knowing</i> and transfer it to new situations?	Student rubric scores for the physics transfer lab and bio transfer lab.

Table 2

Codes for observations of lab behaviors

Code	Description
Making sense	Discussions about physics concepts, experimental design, the data, and the questions in the handout.
Writing	Describing the experiment, recording data, calculating the values, and explaining the results
Procedure	Gathering equipment, mounting set-up, and taking data.
TA help	Listening to a TA who was explaining and answering their questions (only for non-design groups).
Off-task	Any activity that did not relate to the laboratory task.

Table 3

Total time in minutes spent by students on different activities

Lab	Design							Non-design						
	Sm	Wr	Pro	Rd	TA	Oft	Tot	Sm	Wr	Pro	Rd	TA	Oft	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 = TA help; Oft = off task;

Tot = total.

Table 4

The average time in minutes that students spent on different activities in the labs

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

Table 5.

Exam scores

	Fall Exam1			Fall Exam2			Fall Final Exam		
	M.C.	F.R.	Overall	M.C.	F.R.	Overall	M.C.	F.R.	Overall
	(80)	(60)	(140)	(60)	(80)	(140)	(120)	(120)	(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.

Figure 1. ISLE Cycle: Students' activities that emulate the processes of scientific knowledge construction.

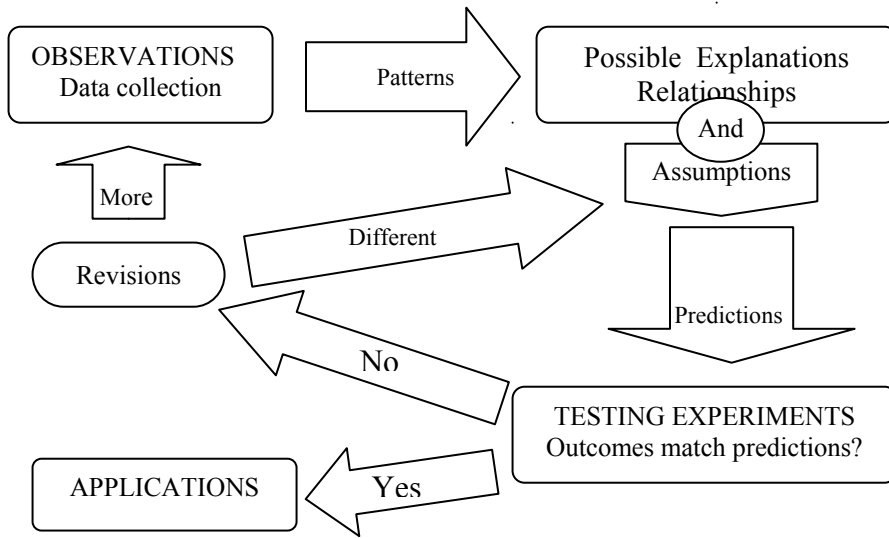


Figure 2. Three theoretical pillars of *ISLE*.

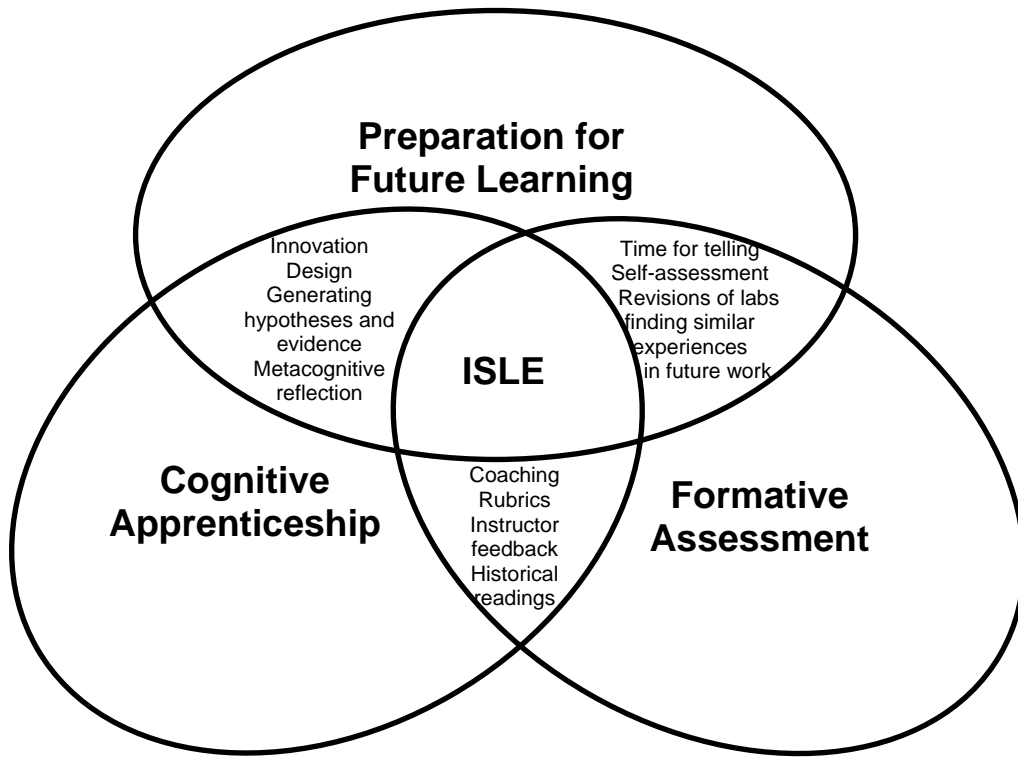


Figure 3. An example of a rubric for one sub-ability.

<i>Scientific Ability</i>	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 4. Why *ISLE* labs are generative: Different types of products.

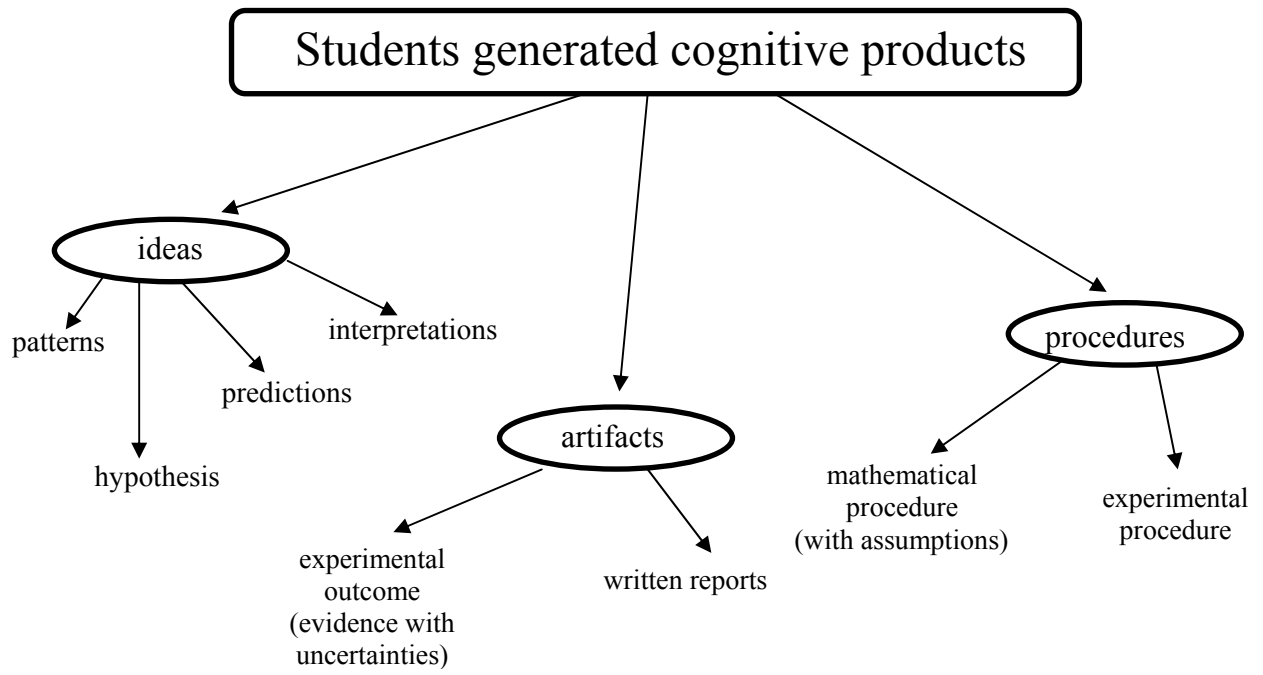


Figure 5. Time students spent on sense making discussions in regular design and non-design labs during the semester.

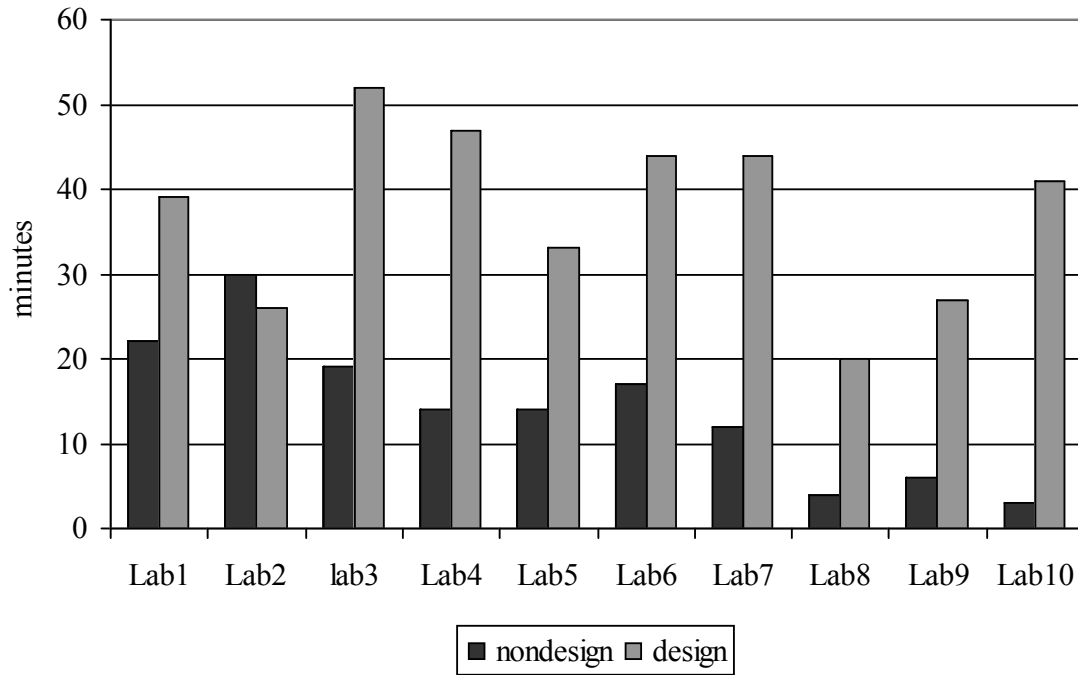


Figure 6. The time spent on different activities by teams of students during the physics and bio transfer labs (N=4 groups).

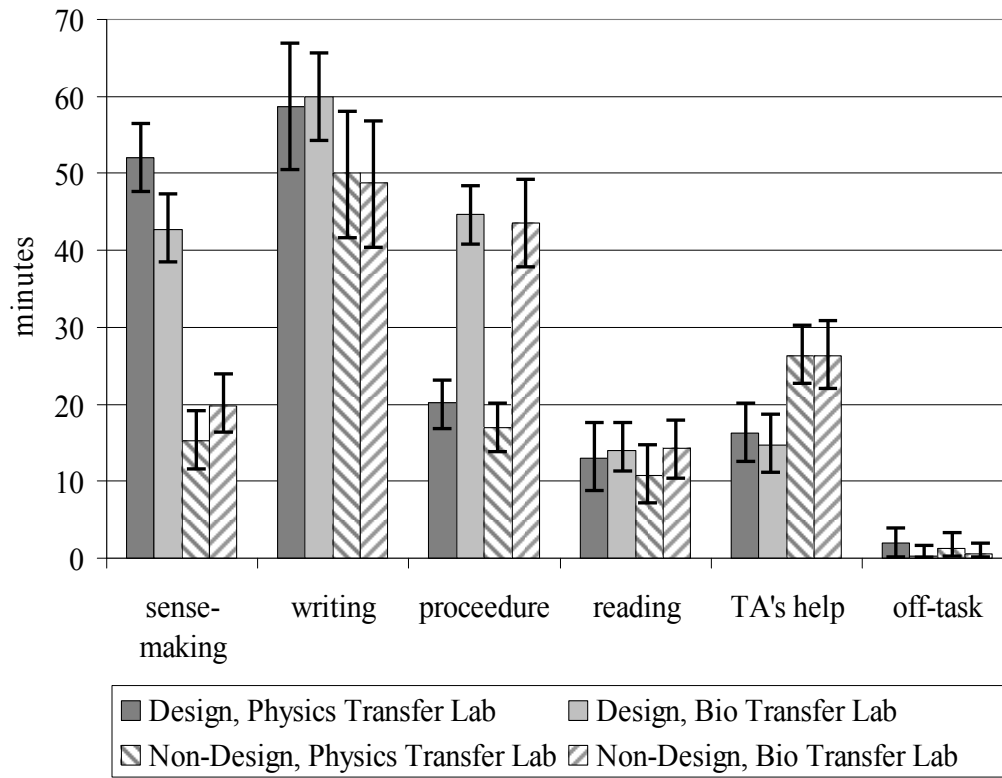
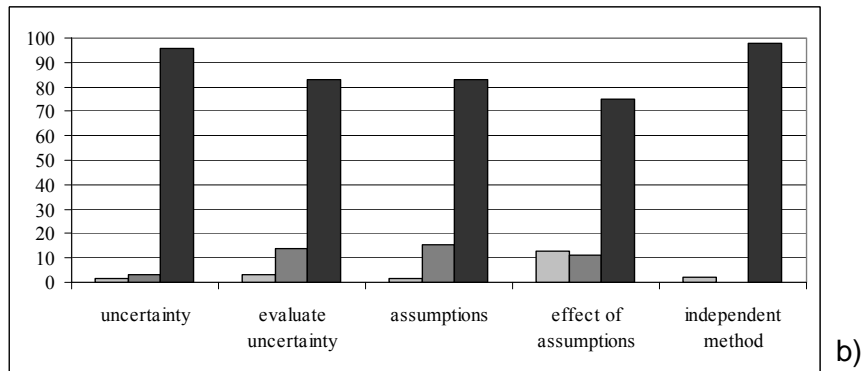
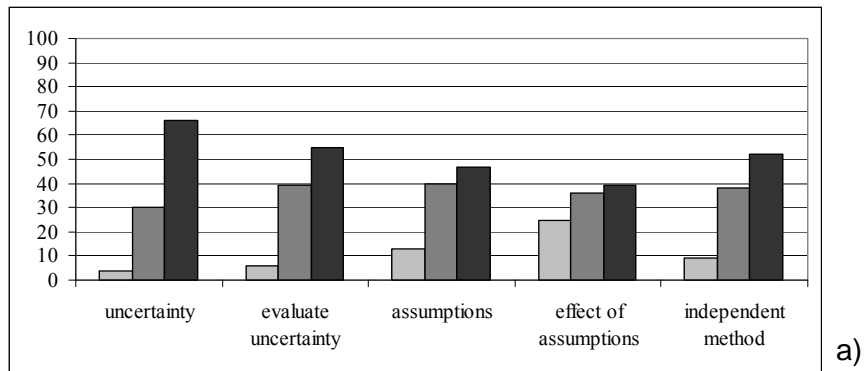
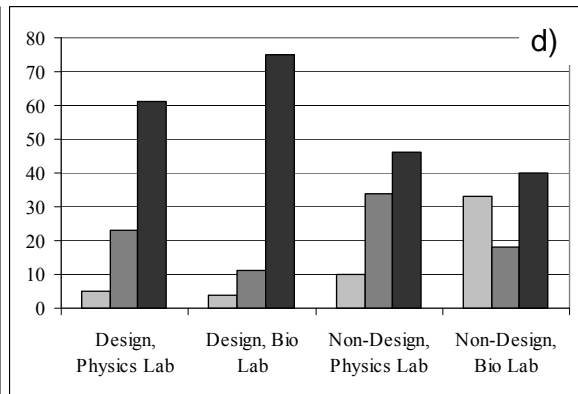
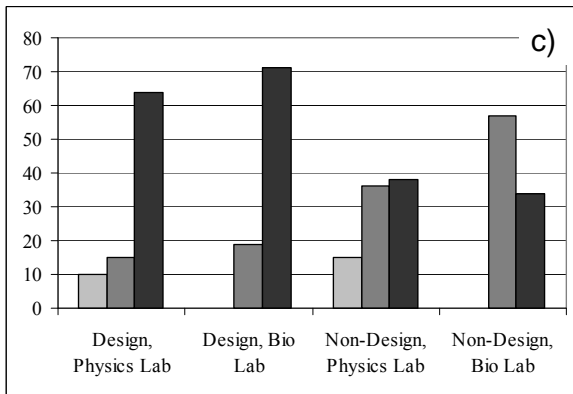
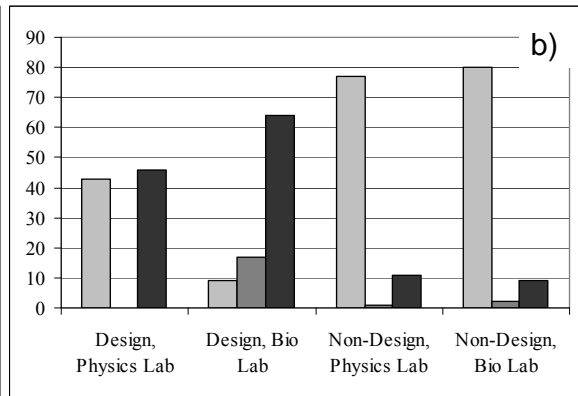
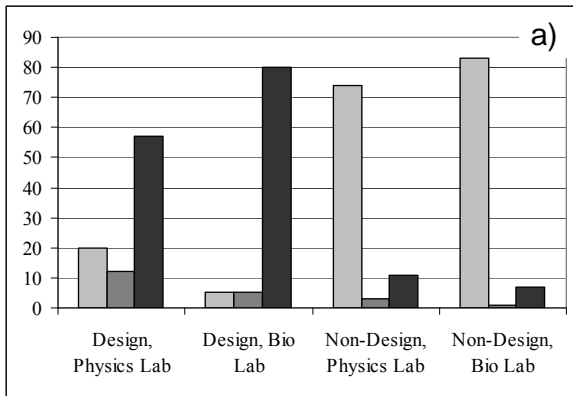


Figure 7. Design students' performance on scientific abilities a) at the beginning of the semester and b) at the end of the semester. The differently shaded bars show the percentage of students who received the scores of 0 – missing, 1 – inadequate, 2- needs some improvement, and 3 – adequate. The data for scores 2 and 3 are aggregated.



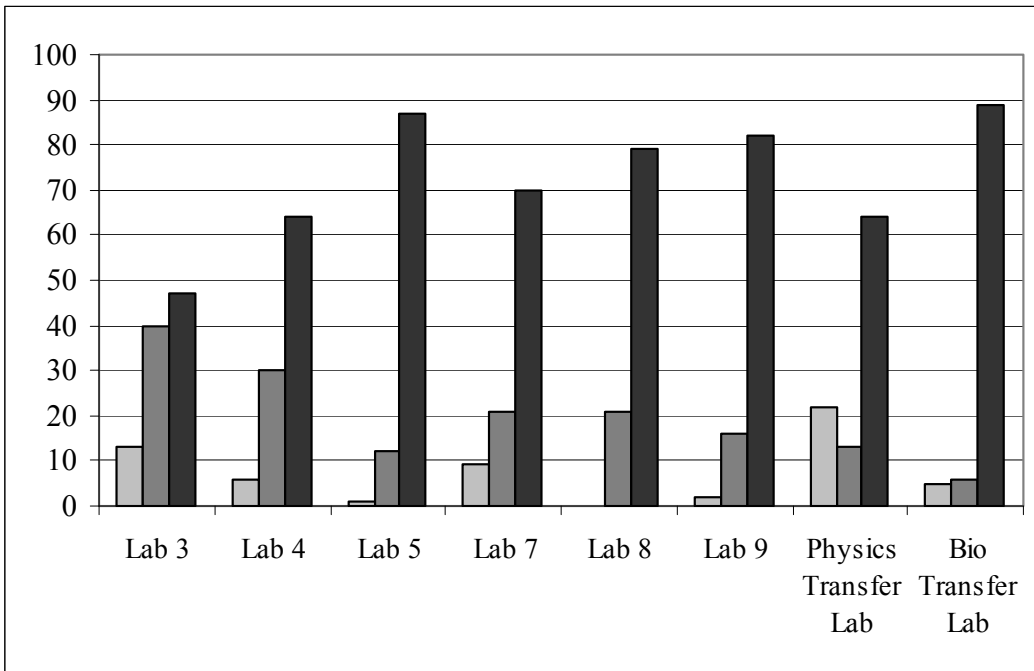
missing
 inadequate
 Needs some improvement and adequate

Figure 8. 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 result by means of an independent method; d) communicate ideas during physics and bio transfer labs.



missing
 inadequate
 Needs some improvement and adequate

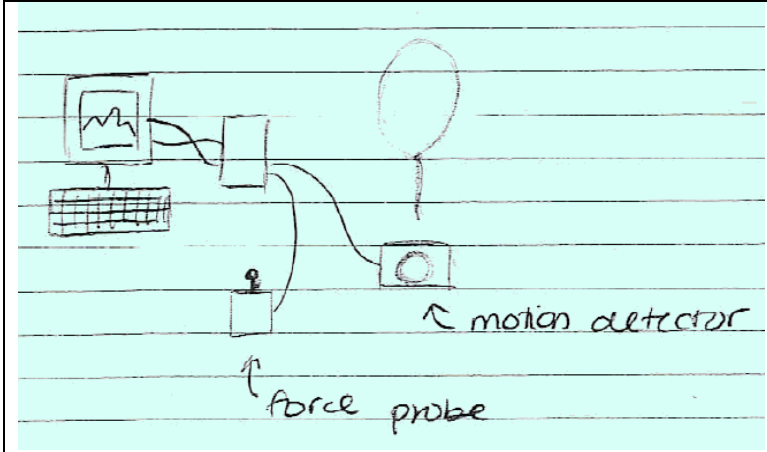
Figure 9. The percentage of design students whose lab reports received different rubric scores for their ability to consider assumptions in the six regular semester design labs and two transfer labs.



missing
 inadequate
 Needs some improvement and adequate

Figure 10. Example of students' lab reports.

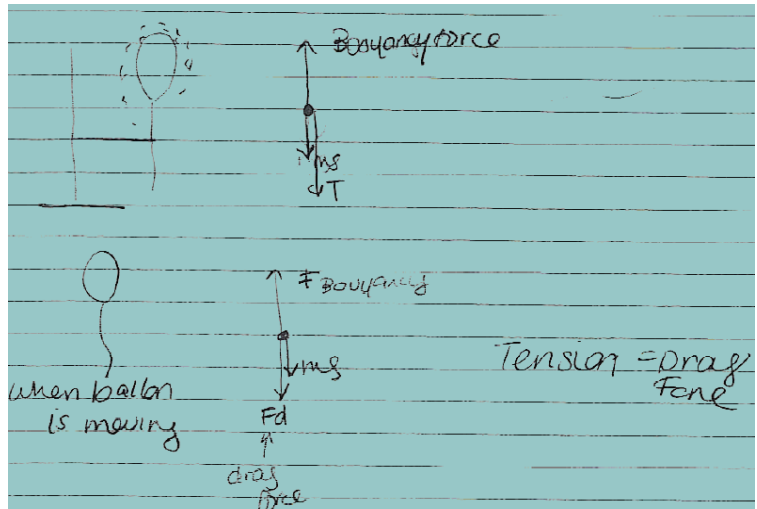
Lab report of a student from a non-design lab group <i>(Task: version 2)</i>	Commentary
<p>1) Find the circumference of the balloon.</p> <p>2) Using circumference find the radius of the balloon.</p> <p>3) Find the mass of the balloon by taping the string to the electronic scale. (Note: you will get a negative value. Take the absolute value) $m = 0.65\text{g}$</p> <p>4) Find the drag force $F_D = 0.65\text{g} \times 10\text{m/s}^2$</p> <p>5) Determine the velocity of the balloon when air resistance and gravitational force are equal:</p> <ol style="list-style-type: none"> place the motion detector on a stand place the sensor face downward place the helium balloon on the floor release the balloon as the motion detector collects data on the position-time graph find constant slope segment the slope of the graph is the velocity $V_1 = .8975\text{m/s}$ repeat twice more find the average velocity $V_{\text{avg}} = .854\text{m/s}$ <p>6) Use the following equation to determine Reynolds number... You should get a value larger than 10.</p> <ul style="list-style-type: none"> - use the equation to solve for drag coefficient ...$C_d=0.51$ - now repeat this procedure for air filled balloon. Make sure to drop the balloon from the level of the motion detector... - air filled balloon - $C_d= 0.61$ <p>Drag coefficient for air and helium are indeed different.</p>	<p>Clear step-by-step instructions, copying the handout's style.</p> <p>No explanation or justification of the method.</p> <p>No picture or FBD.</p> <p>Detailed instructions</p> <p>Multiple trials with averaging without evaluating uncertainty, repeating the rote of the regular labs</p> <p>A correct mathematical description but it does not consider inherent assumptions.</p> <p>Since the results do not incorporate uncertainties, judgment is not justified and not reasonable.</p>
Lab report of a student from a design lab group <i>(Task: version 4)</i>	Commentary
<p>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.</p>	<p>Clear description of the experiment</p> <p>Explanations and justification of chosen methods with clear force diagram.</p> <p>Here is the picture of the set-up.</p> <p>The position-time graph is attached</p>



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.



The buoyant force is always the same. Therefore the drag force is equal to the force of the string attaching the balloon to the scale... $C_d = 0.43$

Assumptions: balloon travels in straight path, balloon is point particle, cross-section is circle, cross-section is level.

Uncertainty	Relative uncertainty
$L = 21.9\text{cm} \pm 0.05\text{cm}$	0.2%
$V(\text{random uncertainty}) = 1.248\text{m/s} \pm 0.1$	8% - the largest
$F_D(\text{scale}) = 1.04\text{g} \pm 0.05\text{g}$	5%

Part II. Prediction (of the speed of the air balloon falling to the ground)

When the air balloon falls it reaches terminal velocity.

Multiple trials with averaging to evaluate uncertainty

Labeled sketch, chart and force diagram are consistent

Here are two force diagrams for balloon at rest and at terminal velocity.

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

<p>Drag force equals the force of the earth. ...</p> <p>We can use the equation ... to get the velocity: $V = 0.438 \pm 0.021 \text{ m/s}$</p> <p>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.</p> <p>Assumptions: 1. Balloon achieves terminal velocity – otherwise $F_e \neq F_d$; 2. $Re > 10$ – otherwise F_d equation is wrong 3. C_d is the same for air and helium – otherwise calculated velocity will be wrong.</p> <p>V was measured and averaged over 3 trials (1.476, 1.02, 1.153). $V = 1.216 \pm 0.228 \text{ m/s}$</p> <p>The values do not overlap and therefore are not equal.</p> <p>Some assumptions must have been incorrect.</p>	<p>assumptions</p> <p>The result incorporates uncertainty.</p> <p>Judgment is based on the uncertainty analysis. There is attempt to consider the effects of assumptions.</p>
---	---