PHYSICAL REVIEW SPECIAL TOPICS - PHYSICS EDUCATION RESEARCH **3**, 020106 (2007)

Acting like a physicist: Student approach study to experimental design

Anna Karelina and Eugenia Etkina
Graduate School of Education, Rutgers University, New Brunswick, New Jersey 08901, USA

(Received 16 April 2007; published 19 October 2007)

National studies of science education have unanimously concluded that preparing our students for the demands of the 21st century workplace is one of the major goals. This paper describes a study of student activities in introductory college physics labs, which were designed to help students acquire abilities that are valuable in the workplace. In these labs [called Investigative Science Learning Environment (ISLE) labs], students design their own experiments. Our previous studies have shown that students in these labs acquire scientific abilities such as the ability to design an experiment to solve a problem, the ability to collect and analyze data, the ability to evaluate assumptions and uncertainties, and the ability to communicate. These studies mostly concentrated on analyzing students' writing, evaluated by specially designed scientific ability rubrics. Recently, we started to study whether the ISLE labs make students not only write like scientists but also engage in discussions and act like scientists while doing the labs. For example, do students plan an experiment, validate assumptions, evaluate results, and revise the experiment if necessary? A brief report of some of our findings that came from monitoring students' activity during ISLE and nondesign labs was presented in the Physics Education Research Conference Proceedings. We found differences in student behavior and discussions that indicated that ISLE labs do in fact encourage a scientistlike approach to experimental design and promote high-quality discussions. This paper presents a full description of the study.

DOI: 10.1103/PhysRevSTPER.3.020106 PACS number(s): 01.40.Fk, 01.40.gb, 01.50.Qb

INTRODUCTION: THE NEEDS OF THE WORKPLACE AND THE GOALS OF SCIENCE EDUCATION

An editorial in the *Journal of Research in Science Teaching* written by Bybee and Fuchs[1] presents a summary of the 12 reports from business, industry, government agencies, and associated groups that provide recommendations for science and technology education and then suggests research questions that need to be answered in light of these new science education goals. The main thrust of all reports is to better prepare students for the 21st century workplace.

What knowledge and what abilities are needed to succeed in this 21st century workplace? This question has been addressed by individual research studies examining the need for various process abilities and for declarative knowledge of people in that workplace.[2–5] Duggan and Gott[6] studied the science used by employees in five science-based industries: a chemical plant specializing in cosmetics and pharmaceuticals, a biotechnology firm specializing in medical diagnostic kits, an environmental analysis lab, an engineering company manufacturing pumps for the petrochemical industry, and an arable farm. They found that most of the scientific conceptual understanding used by employees was learned on the job, and not in high school or university courses. They concluded: "A secure knowledge of procedural understanding appeared to be critical."

In addition to individual research studies like these, there have been a plethora of national studies and reports concerning desired outcomes of science education.[7–9] In the recent publication by the National Academy of Sciences,[10] *Taking science to school: Learning and teaching science in grades K-8*, the authors say that "…quality instruction should promote a sense of science as a process of building and improving knowledge and understanding. Students should have experiences in generating researchable questions, designing methods of answering them, conduction of data analysis, and debating interpretations of data."

According to all these studies and reports concerning science and engineering education (including the reports summarized in the *Journal of Research in Science Teaching* editorial[1]), students have to acquire not only conceptual and quantitative understanding of science principles, but also the ability to reason from the data, construct explanatory models, design experiments to test hypotheses, solve complex problems, and work with other people.[7,8,11,12] These requests place a heavy burden on the introductory physics courses for those students who will not take more physics in college (science majors, premeds, computer majors, etc.). In addition to learning the concepts and laws of physics in a course that moves very quickly, students need to acquire the abilities listed above. However, emphasizing scientific abilities could be beneficial for our students. They probably will not remember the details of Newton's third law or projectile motion while treating patients or studying the effects of chemicals, but all of them will need to make decisions based on evidence and use this evidence to test alternative explanations, deal with complex problems that do not have one right answer, and work with other people. Thus we suggest that it is possible to use the context of physics to help students develop the abilities that they will later use in their lives.

How the ISLE curriculum addresses the goals of science education

The Investigative Science Learning Environment[13] (ISLE) curriculum focuses explicitly on helping students develop abilities used in the practice of science. These abilities include the ability (a) to represent knowledge in multiple ways, (b) to design an experiment to investigate a phenomenon, to test a hypothesis, or to solve a problem, (c) to collect and analyze data, (d) to evaluate the effects of assumptions and uncertainties, (e) to communicate, and many others.[14] Students, who learn physics in the Investigative Science Learning Environment construct and test physics concepts by following a scientific investigation cycle. They start each conceptual unit by analyzing patterns in experimental data. They use multiple representations of the data to construct possible explanations or mathematical relations. In the next and crucial step, students test their constructed ideas using hypotheticodeductive reasoning.[15] Students predict the outcomes of new experiments based on their constructed ideas, perform the experiments, and possibly revise their ideas if the outcomes do not match the predictions. Finally, they apply these ideas to solve practical problems. In all these activities students work in groups, discussing their observations and explanations. ISLE large room meetings are interactive, using elements of peer instruction and a personal response system. For example, students discuss experiments in groups of two, come to a consensus and then share it with the rest of the class.[16] Recitation activities involve groups of students working on multiple representation problems, conceptual exercises, and context-rich problems.[17]

The ISLE laboratories are an integral component of the course. In labs students have an opportunity to collect and analyze data, to test explanations of the patterns in the data that they constructed in large room meetings, to apply tested concepts to solve practical problems, and to represent processes in multiple ways.[18] The ISLE laboratories are less prescriptive than traditional ones and have a noncookbook format. Students work in groups to design their own experiments. They then evaluate the results of the experiments and suggest improvements in their designs. They often need to devise at least two independent experimental methods to solve a problem. Write-ups given to the students for ISLE labs do not contain recipelike instructions on how to perform the experiments but instead guide students through various aspects of a typical experimental process.[19] In the write-ups (see an example in Appendix A 2), students are asked to think about different ways to achieve the goal of the experiment, draw a picture of the arrangement of the equipment that they plan to use, and describe the mathematical procedure they will apply. In addition, the write-ups guide them in recording additional assumptions they make, identifying sources of experimental uncertainty, and evaluating their effects. To facilitate student self-assessment, ISLE labs use scientific abilities rubrics that assist students in their work and help them write lab reports.[14] The rubrics contain descriptors of writing in response to each guiding question in the lab writeup. For example, when the students are asked to design two different experiments to determine an unknown quantity (an example of such an experiment is provided later in the paper), one of the rubrics that they use for guidance is as shown in Table I.[20]

To study whether students develop scientific abilities in such an environment, we conducted two investigations in 2004–2005. We collected students' individual lab reports and scored the reports using the scientific ability rubrics. We found that students significantly improved their abilities to design an experiment to solve a problem, to develop a mathematical procedure to analyze data, and to evaluate experimental uncertainties and theoretical assumptions.[14] These outcomes were based on the analysis of student lab reports or, in other words, on their written work. To find whether ISLE labs make students not only write like scientists but also think and act like scientists, we monitored the way students spent their time during labs. The goals of this research project were to (a) document students' activities in the labs; (b) find whether student activities and discourse can be described through the lens of "scientific abilities;" (c) compare students' activities in the ISLE labs with students' activities in labs where they do not design their own experiments.

Recording and analyzing student behavior in labs

Other educational researchers have monitored students' activity in a classroom in studies of metacognition.[22–24] Our method of analysis for this study is based on the work of Lippmann and colleagues.[25,26] The work of Lippmann and colleagues is most relevant because they studied metacognition and its outcomes in laboratories in introductory physics courses. Their approach involves timing and coding student activities during the lab. Specifically, they focused on metacognition and sense-making. They videotaped students in the labs and then coded students' discussions about physics formulas or concepts, the group's experimental design, their data, or the laboratory question as sense-making. These discussions usually included some metacognitive statements such as "I don't get this," or "Whoa, I got a huge number," or "That might not be right, though…." These statements show evidence that students are monitoring their own reasoning, and these episodes were coded as metacognitive. Metacognitive episodes also included the evaluation of reasoning about plans, measurements, actions, data analysis, or data interpretation.

Lippmann and colleagues found that raters had low agreement with each other in coding for metacognitive episodes. However, their inter-rater reliability was high when they coded for sense-making episodes. Based on this, they developed a coding scheme for students' activities. They observed students in three different types of labs. In the first type, a typical cookbook laboratory, students were given detailed instructions to follow. This feature was the same for the second type, the cookbook+explanations laboratory, where the conceptual questions "explain why this happens," "predict what it will look like," and "do these trends make sense?" were added to the instructions on how to perform the experiments in the middle of the lab. Students had to answer them before and right after they performed the experiments. The third type (scientific community lab) was an open-ended laboratory where students had to design their own experiments to solve problems and defend their method and results in front of other students, mimicking the work of the scientific community. Lippmann *et al.* found that, in pure cookbook labs, students had significantly fewer episodes of sense-making than in the cookbook+explanation and scientific community labs where students designed their own experiments. The difference between design and cookbook+explanation labs was small. However, in her study, Lippmann notes that in the cookbook+explanation lab, a large amount of sense-making was observed only with excellent students who would easily go into a sense-making mode. "For many of the other stu-

TABLE I. An example of some scoring rubrics to assess the scientific abilities.

Ability to design and conduct an application experiment					
Scientific ability		Missing	Inadequate	Needs some improvement	Adequate
2	Is able to design a reliable experiment that solves the problem	The experiment does not solve the problem	The experiment attempts to solve the problem but due to the nature of the design the data will not lead to a reliable solution	The experiment attempts to solve the problem but due to the nature of the design there is a moderate chance the data will not lead to a reliable solution	The experiment solves the problem and has a high likelihood of producing data that will lead to a reliable solution
3	Is able to use available equipment to make measurements	At least one of the chosen measurements cannot be made with the available equipment	All of the chosen measurements can be made, but no details are given about how it is done	All of the chosen measurements can be made, but the details about how they are done are vague or incomplete	All of the chosen measurements can be made and all details about how they are done are provided and clear
4	Is able to make a judgment about the results of the experiment	No discussion is presented about the results of the experiment	A judgment is made about the results, but it is not reasonable or coherent	An acceptable judgment is made about the result, but the reasoning is flawed or incomplete	An acceptable judgment is made about the result, with clear reasoning. The effects of assumptions and experimental uncertainties are considered
5	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. A percentage difference is calculated in quantitative problems.

dents in this lab, this change did not activate sense-making. Instead they used their old strategies of lab survival: Ask the TA, ask other students, and read more in the manual." (Ref. 25, p. 4.)

DESCRIPTION OF THE STUDY

Our research project was conducted in the labs that were integrated in two large-enrollment (approximately 200 students in each) introductory algebra-based physics courses for science majors. These courses were offered on two campuses of the same north-eastern state university. Both courses had a lab as part of the course credit. The experimental course followed the ISLE curriculum with design labs,[18] and the control course had nondesign labs supplemented by additional reflective questions at the end. In both courses labs were 3 h long. In both courses, the teaching assistants (TAs) had about the same 1-h training time weekly to prepare for a new lab. In both courses, TAs had mixed levels of experience—ranging from first-year TAs to those who taught several times before. Below we describe the differences between the labs.

Control course

Nondesign lab

In these labs students performed experiments in groups of 3–4 by following well-written, clear, and concise guidelines (Appendix A 1 shows an excerpt from one of the write-ups) which instructed them on what and how to measure and how to record the data. The equipment and elaborate write-ups eliminated possible difficulties such as equipment-use malfunctions, wrong-assumption effects, and large uncertainty. In some labs students had to devise their own mathematical method to analyze data. At the end of the lab students had to answer conceptual and reflective questions. Teaching assistants provided immediate help to the students when they had a question. At the beginning of each lab, students had a short quiz and then a TA provided explanations. Students received a packet that included all of the labs at the beginning of the semester. They read the lab description at the beginning of the lab session and entered data in the blanks in the data tables as they went along. At the end of the lab they discussed reflection questions in the write-up and wrote the answers. However, when the TA checked the answers, he/she did not collect the writings, but discussed students' answers orally.

Experimental course

ISLE design lab

As we mentioned above, in ISLE laboratories students designed their own experiments. They worked in groups of 3–4. Lab write-ups did not contain instructions on how to perform the experiments; instead, they had questions guiding students through various aspects of a typical experimental process (Appendix A 2 shows an ISLE lab write-up). At the end of each experiment, students answered reflective questions that focused on different aspects of the procedure that they invented. In addition, students used scientific abilities rubrics for guidance and self-assessment.[14] TAs served as facilitators. Students received a lab packet when the semester started and had an opportunity to read the task before coming to the lab. Students' individual lab reports were graded based on the rubrics (about 3–4 rubric items per experiment); the rubrics encouraged students to spend considerable time writing.

ISLE practical

The practical exam was a lab where students needed to design experiments without the scaffolding present in regular ISLE labs (see Appendix A 3). The tasks were posted on the class website about a week prior to the exam. Student work was graded based on their reports. The practical exams occurred twice each semester. The data in this study are from the second practical, which occurred during week 8 of the semester.

ISLE biolab

For the purposes of the study, we devised a lab where students had to design an experiment to solve a biology-related problem (see Appendix A 4). The biology content of the problem was unfamiliar to the students. The lab had no scaffolding questions and instructors did not provide any help or assistance to the students. This lab was a part of the study and not a part of the course; eight students volunteered to perform the lab after the course was completed; they did not write a lab report. Consequently, the lab was not graded. The lab had no effect on students' grades.

Sample

We monitored the behavior of 14 different groups of three to four students: four ISLE labs, four ISLE practical exams, one ISLE biolab, and five different nondesign labs. Each observation lasted for the entire lab. The groups were chosen randomly from both courses. The observations were conducted in different lab sections, thus there was a mixture of TAs in terms of the years of teaching experience in both courses: there were TAs who have taught before and there were TAs who taught for the first time. For the biolab part of the study, we observed one of the two groups of students who performed it. The students who volunteered for the biolab were better than average students with course grades of A's and B's.

CODING AND LIMITATIONS

We assumed the populations in the two courses to be roughly similar. We based our assumption on the fact that the courses had the same prerequisites and covered the same material. However, we did not have a pretest to compare student populations; thus this assumption is one of the limitations of the study.

We also assumed that the presence of the observer did not affect student behavior. An observer sat beside the members of a lab group timing and recording all student activities and conversations. Students have not met the observer before the lab and TAs introduced the observer as a member of the education research group. Neither TA nor students were aware of study questions. After the lab was over, the field notes were rewritten and a complete transcript of each lab session was constructed. The analysis of the first transcripts revealed certain patterns in student activities that led to devising codes for student activities.

As a starting point for the coding, we used the tree-coding scheme for student activities described in Ref. [25,26]: 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.

We used Lippmann's three-item coding scheme while analyzing the observations for the first lab groups. We observed that the students in design labs spent considerable time writing. Thus, we modified the coding scheme to include the writing. Later, while observing nondesign labs, we found that students spent considerable time getting help from the TAs. So we included an additional code "TA's help" (see Table II). We analyzed subsequent observations using this modified coding scheme. We found that all behaviors fit into one of the five coding categories.

Although having only one observer was a limitation of our study, we found this new five-item coding scheme to be

TABLE II. General codes.

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

reliable. We verified it with a second observer. To do the verification, we had two trained observers code the same group of students performing a lab without any discussion of the observations (we used four new lab groups that were not included in this study). We analyzed coding of every minute of a 170-min lab. We found that raters agree in about 84% of cases before the discussion and 100% after the discussion. Most of the original disagreement came from the coding "TA help." One of the observers coded all interactions between the students and the TA as help, the other one only the interactions when TA actually helped students with the design, mathematics, assumptions, etc. As it was the latter observer who coded the labs presented in this study, we can assert that the time coded as TA help was actually spent on the help related to the lab.

Making-sense codes

We focused on sense-making because it represents verbalization of the students' cognitive processes. We classified the content of sense-making discussions further (see Table III) according to activities matching the descriptions of relevant scientific abilities (the list of abilities is available in Ref. 21).

Examples of sense-making discussions

Below are examples of sense-making episodes involving discussions related to the effects of assumptions in the experimental and mathematical procedure.

ISLE Lab: Effect of assumptions

S1: I think we can ignore the friction.

S2: But we cannot ignore it. We should take it into account.

S1: No, it is too small.

TA: How can you check this?

S1: Let's measure the friction.

S2: How?

S1: Do you remember that lab where we measured it? We can tilt the track and measure the angle when the car starts sliding. *They tilt and observe that the car slides immediately at an extremely small angle, which they cannot measure.*

S2&S1: So, we can ignore the friction!

Nondesign lab: Effect of assumptions

S1: What temperature of ice should we plug into the equation?

S2: 0 °C

S1: How can you be sure that it is zero degrees?

S2: It should be. It is always 0 °C.

S3: No. Ice can have much lower temperature.

S1&S2: Let's ask the TA if we should take zero degree temperature.

TA: Yes, you can assume that it is 0 °C.

Notice here that the beginnings of the discussions were very similar; students were exchanging their unfounded

TABLE III. Sense-making codes.

D	Design	Discussing experimental design and setup, planning the experiment, etc.
M	Mathematical model	Choosing the mathematical model and the physical quantities to be measured
A	Assumptions	Discussing the effects of assumptions inherent in the mathematical model and model's limitations
U	Uncertainties	Discussing sources and calculating values of experimental uncertainties
Min	Minimizing	Discussing the ways to minimize uncertainties and the effects of the assumptions on the outcome
R	Revising	Discussing reasons for the discrepancy and the ways to improve the experimental design to get the discrepancy less than the uncertainty
O	Unexpected observations	Discussing the reasons for obtaining unexpected data
VA	Validating assumptions	Discussing how to justify assumptions

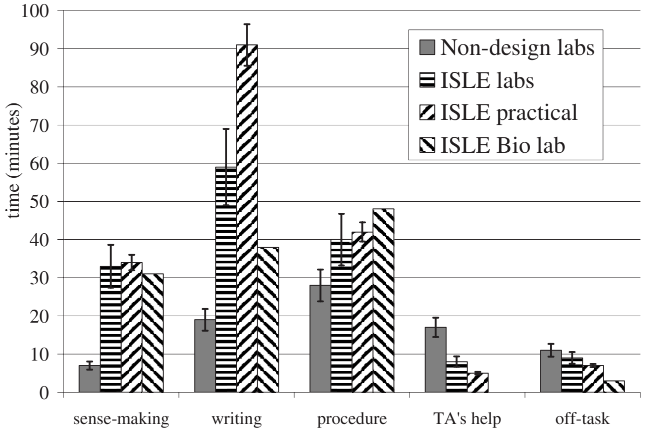


FIG. 1. Time spent on different activities (in minutes). The data are averaged over the sample. The error bars show the standard deviations of the data sets for the observed groups.

opinions and their discussion brings them nowhere. However, the TAs' intrusions were very different. In the ISLE lab the TA triggered the next level of discussion by suggesting that students check their ideas. In the second episode, the TA answered the question. This made any further sense-making unnecessary.

FINDINGS OF THE STUDY

Duration: Design lab versus a nondesign lab

In spite of the fact that each lab in the control course had more experimental tasks, it took students about half the time to complete them compared to ISLE labs (average of 80 min versus 160 min). During ISLE labs, students spent a great deal of time planning, discussing, and writing detailed lab reports (see Fig. 1).

Writing took about 60 min in a typical ISLE lab versus 20 min in a nondesign lab. During the practical exam, writing took even more time (90 min) because the practical exam grades were based mostly on the quality of the lab report. The biolab was not graded and the lab report was not a requirement, so in the biolab the category of "writing" includes mostly recording data, calculating, and reading the explanatory text.

The experiments themselves (procedures) took more time in ISLE labs (40–50 min versus 28 min for the traditional labs). This difference probably occurred because ISLE students needed to revise and improve the experiment as they proceeded. In ISLE labs the TAs' explanations were minimal. Discussions similar to those shown above were considered sense-making for both nondesign and ISLE labs.

Sense-making and scientific abilities

Figure 1 shows that there was a remarkable consistency of ISLE students engaging in sense-making. In the regular ISLE labs, during a practical test, and during a biolab, students spent about the same time (about 33 min) making sense of what they had to do (Table IV). Thus, averaging over all ISLE labs including the practical and biolab, sense-making lasted about 20% of the time. In nondesign labs, sense making lasted for 5–8 min, i.e., 8% of the total time that students were in the lab (see Fig. 2). This difference is statistically significant with the level of significance $p=3.35\times10^{-5}$ and the variance ratio $F=41.0963$.

TABLE IV. Time spent on sense-making discussions (in minutes).

	Lab 1	Lab 2	Lab 3	Lab 4	Lab 5
Nondesign	7	8	9	4	6
ISLE labs	31	16	41	42	
ISLE practical	33	36	33	33	
ISLE biolab	31				

Other differences include the relatively large percentage of the lab time that students in nondesign labs spent on the procedure compared to sense making (almost three times longer) while nondesign students spent comparable fractions of the lab time on the procedure and sense-making.

To provide details of students' discussions during sense-making episodes and to record the sequence of different activities, we used time lines (see Fig. 3). The set of gray and black strips with the smallest increment of 1 min shows what students did during the lab, when, and for how long. Numbers beneath the time lines show how many minutes passed after the beginning of the lab. The gray color on the sense-making part of the time line marks instances when student sense-making was prompted by the questions in the lab write-up or by a TA. The black color indicates when sense-making was spontaneous. Figure 3 shows, for example, that ISLE students spent the first 5 min on off-task activities and then the next 3 min engaged in sense-making that was coded D, which means they designed and planned the experiment.

One can see from Fig. 3 that episodes of sense-making were often followed by writing. This happened because students first discussed their answers to the questions in the lab

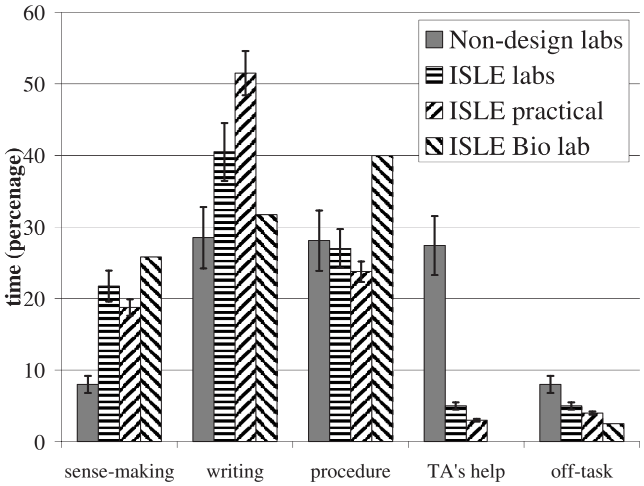


FIG. 2. Percentage of time spent on different activities. The data are averaged over the sample. The error bars show the standard deviations of the data sets for the observed groups.

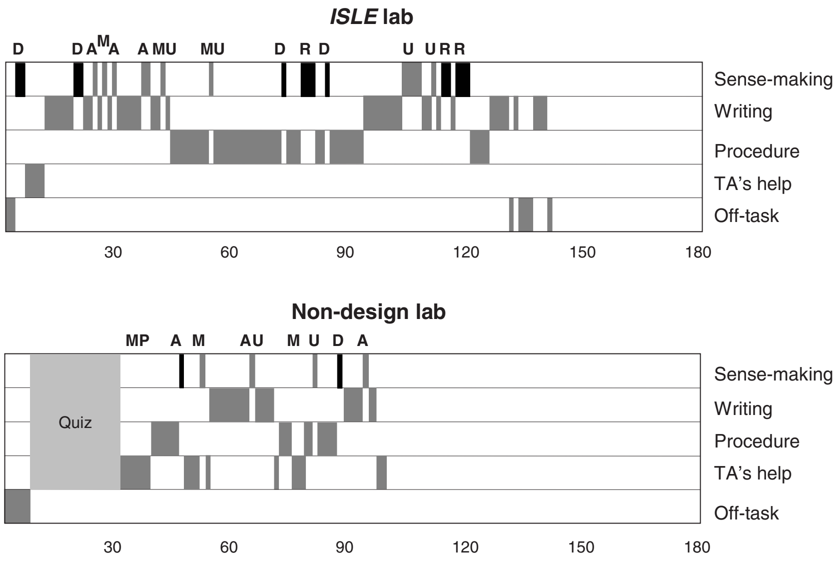


FIG. 3. Time lines for randomly selected groups of students in different types of labs. Black color bars indicate episodes of student sense-making that were not prompted by a write-up or a TA. Gray sense-making episodes indicate that student activity was prompted by a TA's question or by a question in the lab write-up. Sense-making subcoding reflects the scientific abilities described in the text.

write-up and then wrote the results of their discussion. The cases when sense-making and procedure followed each other correspond to situations when students first discussed what to do and then implemented their plan.

The detailed analysis of the sense-making episodes reveals differences between different labs. In the nondesign labs students engaged in sense-making for a very short time. Few of their statements could be coded as related to scientific abilities (discussing the mathematical model, experimental uncertainties, theoretical assumptions, evaluating results, etc.). The TA often explained the experiment design and the mathematical model.

Another important difference was how often students switched to sense-making mode without prompting by TAs or the questions in the lab write-up. Such episodes of self-triggered sense making are indicated on the time lines by black bars (Fig. 3). Since experimental design requires independent decisions, it is not surprising that these episodes happened much more often in ISLE labs than in nondesign labs. Indeed, one can see that the self-triggered sense-making usually happened during periods when discussion and procedure episodes follow each other, i.e., during periods of planning, executing, and revising an experiment.

Role of the TA

The TAs' role seems to be important in the students' lab activities. The examples of sense-making discussions show that a TA can trigger sense-making or, on the contrary, end it. For example, during the TA meetings in the ISLE course, TAs were specifically instructed not to answer students' questions directly. It requires considerable effort, experience, and content knowledge for a TA to trigger sense-making. Not all of the ISLE TAs were able to do this at each interaction with students. However, our findings show that the role of the lab structure and instructions is usually larger than the TAs' effect. If the TAs had the most important effect, then the spread of the time data between different TA sections in the same environment would be larger than the spread between the types of labs. However, although the TAs in all sections were different, the spread of the time averages for different activities is much smaller than the differences between the types of labs, thus suggesting that the time distribution and duration of activities depend more on the type of the lab than on the quality of the TA. One of the explanations might be that most of the questions students may have during the lab are explained in the nondesign manual and do not require much discussion. On the contrary, in ISLE labs students have so many questions that even the most efficient TA is not able to respond to all of them.

Outcome of sense-making

Sense-making discussions happened mostly in two situations: (1) when students were answering write-up questions; and (2) when students were having difficulty or doubts about the experimental procedure. After situations of type 1, students usually proceeded to writing since they had to provide a written answer. After situations of type 2, students would usually proceed to carry out the experiment. A detailed analysis of the time lines reveals that, in ISLE labs, sense-making discussions in type 2 situations were followed by procedural changes, i.e., attempts to improve and revise the experiment or carry out the next steps [Fig. 4(a)]. In nondesign labs, about 70% of such sense-making discussions led to asking a TA who provided an immediate answer [Fig. 4(b)]. We observed only one episode when students asked a TA which parameters to plug into a formula to analyze data and the TA made them derive the formula.

Thus in ISLE labs students pose questions and answer them themselves, whereas in nondesign labs students seldom pose their own questions and tend to search for answers from external authorities.

Conceptual questions: ISLE labs versus nondesign labs

Nondesign labs had conceptual questions at the end. The goal of the questions was to encourage students to analyze the physical processes involved in the lab experiments.

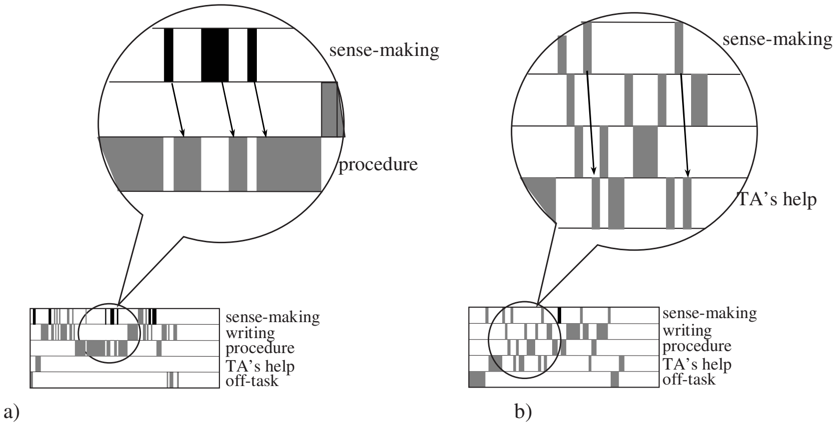


FIG. 4. Sense-making outcomes: (a) in ISLE labs a discussion leads to a decision and to an execution; (b) in nondesign labs sense-making leads to a TA's help, with the TA answering and explaining.

These questions look very similar to some guiding questions in the ISLE labs. For example, in the calorimetry lab, students in an ISLE lab had to answer the following questions when designing an experiment: "How would you minimize uncertainties?" At the end of a parallel nondesign lab, students had to answer: "Are there any ways to reduce the error in this minilab?" Although the questions were similar, their place and role in the lab were different. In the ISLE lab students had to answer this question while they were planning the experiment—before they started to perform it. Eventually, students learned that thoughtful consideration of these questions would help them perform a successful experiment. In contrast, superficial answers could result in inconsistent data and in repetition of the experiment. Below are examples of the student discussions of this question in the ISLE lab.

S1: We should heat it to the boiling point, but do not let it boil, so we will not lose water by evaporating.

TA: Why do you need to heat it to the boiling point?

S1: The larger the temperature difference, the smaller the uncertainty.

Students discussing why the measurements of the specific heat are different:

S1: I think that we skewed our first experiment (with hot water). I guess the calorimeter is leaking, so the temperature changed too fast. It was 95 °C in the beaker but I think that after we poured it into the calorimeter the temperature became much smaller.

S2: Let's repeat the experiment with warm water. Somewhat 60 °C. Maybe it will help.

S3: Right. And do it faster.

We see that students made a great deal of cognitive effort to answer the questions and returned to them several times during the lab.

In the nondesign lab students had to answer the conceptual question about reducing the experimental error after they had already performed the experiment. The whole lab activity did not depend on answering this question. Students did not have to choose better ways to carry out the experiment because all the necessary instructions were given to them by a detailed write-up. As a result the answers to the questions were irrelevant to the students' activities during the lab. Consider an example of students' answering this question in the nondesign lab.

Nondesign lab write-up: Name the main source of the experimental error.

Students: The heat is leaking.

Cookbook lab write-up: How can you reduce the experimental error?

Students: We can take the ideal calorimeter, which does not let heat out.

TA, checking the answers: Yes, it is correct.

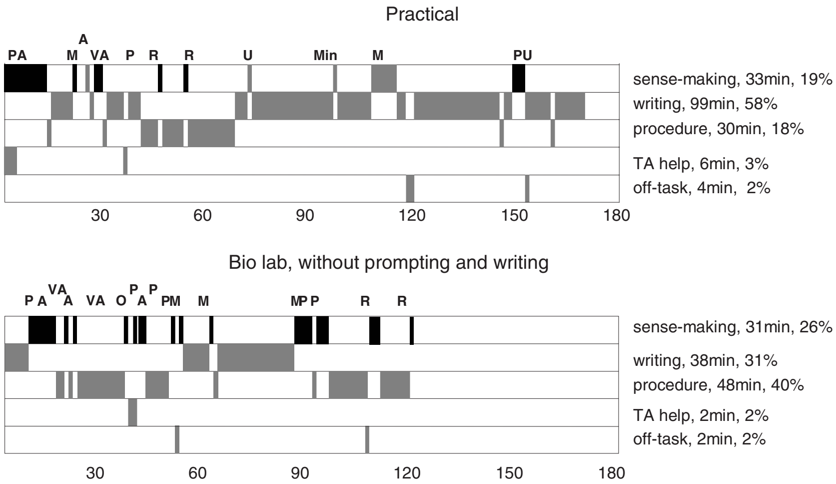
Here we see that students realized that "heat leaking" is crucial for the calorimetry experiment but they did not try to search for realistic ways to minimize the effect. They gave a superficial answer that was not useful as a practical improvement to the experiment. The TA accepted this answer so students did not return to this question. The total time students spent answering the question was about a minute.

Thus, we see that the location of questions in the labs and the role they play leads to completely different treatments by the students. The integration of the questions into the lab activity makes students treat them as important steps to improve the performance. Otherwise students do not consider the conceptual questions to be related to the real experiment and answer the questions quite superficially.

Independent thinking

Our observations show that ISLE lab guidelines help students approach experimental work in ways that resemble the work of scientists. However, a large part of student behavior was a response to the guiding questions in the lab write-ups. It is important to know whether students can approach experimental design like scientists without the prompts in the lab write-up. Do students acquire the ability to pose productive questions and search for answers without prompting? To answer these questions we monitored students' behavior during practical exams in the mid-semester and the biolab at the end of the semester. On the practical exam students were encouraged to use scientific ability rubrics[14,18] as the guidelines. During the biolab there was no scaffolding and no guidelines at all.

The observations (Figs. 1 and 5) show that students spent as much time in the sense-making mode during the practical exam and the biolab as during a regular ISLE lab. One can see that students alternated between sense-making and writing many times during the practical. Could this sense-making


Practical

Bio lab, without prompting and writing

FIG. 5. Time lines for the practical exam and for the biolab.

be caused by the rubrics that prompted students to focus on specific scientific abilities? During the biolab there was no prompting. Thus, we speculate that students' sense-making was triggered exclusively by their own metacognition.

DISCUSSION

In this study we investigated student behavior in labs where they design their own experiments and in labs where students need to follow instructions to perform prechosen experiments. Although both types of lab had similar questions prompting students to think about the experimental uncertainties and theoretical assumptions, students responded differently to the prompts. Students in ISLE labs engaged in behaviors that are much closer to the behaviors of scientists than did students in the traditional labs. We also found that ISLE students carry these behaviors into situations where there is very little or no scaffolding. There are several major differences between student's behaviors in the two types of labs.

(1) ISLE students spend more total time in the lab even when they have to design the same experiments as nondesign students. The major differences in the times spent were in sense-making and in writing.

(2) ISLE students not only spend more time in the lab but they also distribute this time differently among the activities. Their attention is almost evenly spent on sense-making and procedure, while nondesign students spend almost three times more time on the procedure than on sense-making.

(3) ISLE students are remarkably consistent in spending their time on different aspects of the lab whether they have guiding questions (labs) or not (practical, biolab). Although we observed only one group in the biolab, we are currently repeating this experiment with many more groups and finding that the results are very similar.

(4) After sense-making discussions, students in nondesign labs seek TA help more often than ISLE students, who tend to resolve these discussions either by conducting an experiment or by writing their answers.

(5) Students treat lab reflection questions differently depending on the role of these questions in the lab. Reflection questions, that do not affect the procedure that students use to solve the problem, get answered superficially.

Below we present some explanations for these findings.

We speculate that student behavior can be explained if we consider their lab experiences during the semester. The ISLE labs provide scaffolding and guidance by focusing students' attention on the repetitive elements of an experimental procedure. This scaffolding is provided via guiding write-up questions, TA questions, and self-assessment rubrics. The questions in the labs encourage them to think through the procedure before they start performing the experiment. Thus, students get used to spending time on careful planning and paying attention to details while they are designing an experiment. The TA's behavior as a part of the lab environment supplements the lab write-ups. Possibly, if ISLE TAs were not specifically instructed not to answer students' questions directly, the students' behavior would be different. The rubrics emphasize the need to consider and evaluate assumptions inherent in the mathematical model that they will use. Thus, when they devise a mathematical procedure, they think of validating and minimizing the assumptions. The lab guiding questions and the rubrics ask the students to consider the experimental uncertainty and make students design two independent experiments when determining the value of a physical quantity. Thus, they get accustomed to this procedure and repeat it when working on their own. They spent a great deal of time writing lab reports to communicate the details of the experiment because their work was graded based on a written report. Does it mean that students understand the purpose of these elements of experimental investigation and will use them when working independently? Our study does not answer this question, but currently we are conducting a new project that will allow us to have more definite answers.

The traditional labs, even supplemented with conceptual and reflection questions, did not engage students in similar activities. They did not spend time choosing the strategy, validating results, and improving their design. The additional conceptual questions did not make students think deeply about processes during experiment. This finding seems to contradict the results of Lippmann, who indicated that sense-making in scientific community labs and in cookbook +explanations labs took similar amounts of time. However, if we examine the labs more deeply we will find that the findings are rather similar. In Lippmann's cookbook +explanations labs, students had to answer conceptual and reflective questions while they were doing the experiments or before they did them—for example, they had to predict the direction of the induced current and then explain how the current was being induced by writing "a paragraph as if you were explaining this to someone who didn't understand this but was in physics." This approach resembles ISLE labs with their guiding write-up questions more than nondesign labs. Thus, we can say that conceptual and reflective questions promote deep discussions only when students can see that their answers to these questions affect how they proceed in the lab.

Our findings have many limitations. First, we did not have a pretest to ensure that the populations of the students in the two courses are similar. Second, we observed a small number of groups in both types of lab. Third, we did not videotape the labs. Thus, when we went back and checked the coding we only had a written record of the episodes, not the actual behaviors. However, even with these limitations, our findings provide a basis for more rigorous studies that we are currently conducting.

As we discussed in the Introduction, students "should have experiences in generating researchable questions, designing methods of answering them, conduction of data analysis, and debating interpretations of data."[10] Such activities resemble the activities that will be important in the knowledge-based workplace of the future. And these are exactly the activities in which students were frequently engaged in the ISLE labs. However, those who have tried to implement labs without detailed directions for the students know how difficult this is. Students do not like the frustration associated with the design process, especially at the beginning. They want clear directions and clean experiments. Our study provides support for the argument that design labs are worth the effort.

A brief report of some of our findings that came form monitoring student activity during ISLE and nondesign labs was presented in the Physics Education Research Conference Proceedings.[27]

ACKNOWLEDGMENTS

We thank Baki Brahmia for providing us an opportunity to conduct this study, Alan Van Heuvelen and David Brookes for helpful comments on this manuscript, and the National Science Foundation for supporting this project (Grants No. DUE-0241078 and No. REC 0529065). We also thank Sahana Murthy, Michael Gentile, David Brookes, Richard Fiorillo, Gabe Alaba, Maria Ruibal Villasenor, and Alan Van Heuvelen for their help in devising and implementing the ISLE labs.

APPENDIX: LAB WRITEUPS

1. Nondesign lab: Determine the heat capacity of aluminum

(1) Use the following table to record your data [table is omitted here].

(2) Fill the calorimeter about half full with tap water.

(3) Record the mass of the calorimeter and water ($m_{water+cal}$).

(4) Record the initial temperature of the water (T_{water}).

(5) Place a hot piece of aluminum, which you can obtain from the boiling water bath, in the calorimeter and cover it immediately. Notice that you need to record the temperature of the aluminum piece for later calculations.

(6) Shake the water, making sure you move the piece of aluminum so that the water-aluminum system may come to thermal equilibrium.

(7) After 3 min record the final temperature (T_{equ}).

(8) Obtain the mass of the calorimeter, water, and aluminum ($m_{\mathrm{Al}+water+cal}$).

(9) Devise a method to determine the specific heat of aluminum, and compare it with the specific heat of aluminum in the textbook. Find the percentage difference.

Questions

(1) What assumptions did you have to make to derive the formula for the heat capacity of aluminum?

(2) Are there any ways to reduce the error in this minilab?

(3) Suppose you put a block of iron and a block of Styrofoam in the freezer and allow them to stay for a little while. If you gripped both blocks which one would feel warmer? Explain why.

(4) What are significant sources of error and how would each source of error affect the result?

2. ISLE lab: Specific heat of unknown object

Design two independent experiments to determine the specific heat of the given object. The material comprising the object is not known.

Equipment. Water, ice, beaker, hot plate, Styrofoam container with a lid, weighing balance, and thermometer. First, recall (you don't have to write anything!) why it is important to design two experiments to determine a quantity.

Play with the equipment to find how you can use it to achieve the goal of the experiment. Come up with as many designs as possible. Write brief outlines for each design. Working with your lab partners, choose the best two designs. Indicate the criteria that you used to decide which designs were the "best."

For each method, write the following in your lab report.

(a) Give a verbal description and draw a labeled sketch of the design you chose. Include the quantities you will measure.

(b) State the mathematical procedure you will use.

(c) List all assumptions you have made in your procedure. For example, one of the assumptions in your procedure may be that no energy is lost from the calorimeter to the air. Determine if this assumption is valid in your experiment, and if not, evaluate whether the specific heat you determined using the procedure with assumptions is greater than or smaller than the actual value.

(d) List sources of experimental uncertainty. Decide what is the largest source of uncertainty. Use the weakest link rule to estimate the uncertainty in your result. How would you minimize uncertainties?

(e) Perform the experiment. Make sure you take steps to minimize experimental uncertainties. Record your measurements in an appropriate format.

(f) Calculate the specific heat, based on your procedure and measurements. Include experimental uncertainty in each value of specific heat that you determine.

(g) After you have done both experiments, compare the two outcomes. Discuss if they are close to each other within your experimental uncertainty. If not, explain what might have gone wrong—perhaps one of your assumptions was not valid. If your results are not close to each other within experimental uncertainty, perform the experiment again, taking steps to improve your design.

3. ISLE practical

(I) 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 successive latching position. Your task is to determine the elastic potential energy stored in the launcher in each of these launching positions, using the generalized work-energy principle.

(II) Getting the Hot Wheels car to successfully make a loop-the-loop. Your task is to determine the least-energy launching position so that the car will make it around the loop without losing contact with the loop—on the first try (do not use a trial and error method). If you use the next-lower-energy setting, the car should not make it around the loop. You may use the results you obtained from the previous experiment.

4. ISLE biolab

[As students did not study humidity or transpiration, we provided them with information about both that we downloaded from the internet. Here we do not show all of the text—one can Google both and see what is available.]

Design two experiments to determine transpiration rate using stem cuttings from a single species of plant.

Available equipment. Water, beaker holding plant cuttings, Parafilm, tubing, ring stand, graduated pipet, timers, humidity sensor, cup, cup with hole, scissors, and two droppers.

What do you need to know more about to be more successful in determining the transpiration rate?

The following resources were given to the students (they could also use the internet for more information about transpiration and humidity).

Excerpt from resources for help

Transpiration is the process by which water evaporates from the leaves (and sometimes stems) of plants. As this water evaporates, an osmotic gradient pulls water up from the roots through tubelike structures called xylem. Water filling these structures keeps nonwoody portions of the plants from wilting. The critical sites for this evaporation are the stomata. These are openings controlled by the plants to help them regulate water loss. These sites are also critical for gas exchange (allowing necessary CO_2 to enter and O_2 to exit). Plants vary in the number of stomata they contain….

What is humidity and how do we measure it? Humidity is something we hear about daily in weather reports. Humidity is to blame for that muggy, steam-room feeling you experience on certain summer days.

Humidity can be measured in several ways, but *relative humidity* is the most common. In order to understand relative humidity, it is helpful to understand first absolute humidity….

---

[1] R. W. Bybee and B. Fuchs, "Preparing the 21st century workforce: A new reform in science and technology education," J. Res. Sci. Teach. **43**, 349 (2006).

[2] P. Chin, H. Munby, N. L. Hutchinson, J. Taylor, and F. Clark, in *Reconsidering Science Learning*, edited by E. Scanlon, P. Murphy, J. Thomas, A. Whitelegg, and E. Whitelegg (Routledge Falmer, London, 2004), p. 118.

[3] M. Coles, in *Proceedings of the Eighth Symposium of IOSTE*, edited by K. Colhoun, R. Panwar, and S. Shrum (University of Alberta Press, Edmonton, 1997), Vol. 1, p. 292.

[4] R. Gott, S. Duggan, and P. Johnson, "What do practicing applied scientists do and what are the implications for science education?," Res. Sci. Technol. Educ. **17**, 97 (1999).

[5] E. Lottero-Perdue and N. W. Brickhouse, "Learning on the job: The acquisition of scientific competence," Sci. Educ. **86**, 756 (2002).

[6] S. Duggan and R. Gott, "What sort of science education do we really need?," Int. J. Sci. Educ. **24**, 661 (2002).

[7] *Improving Undergraduate Instruction in Science, Technology, Engineering, and Mathematics: Report of a Workshop*, edited by R. A. McCray, R. L. DeHaan, and J. A. Schuck (National Academy Press, Washington, DC, 2003), http://darwin.nap.edu/books/0309089298/html

[8] *Committee on the Engineer of 2020, Phase II, Committee on Engineering Education, National Academy of Engineering* (National Academy Press, Washington, DC, 2005), http://nap.edu/catolog/11338.html

[9] National Research Council,*National Science Education Standards* (National Academy Press, Washington, DC, 1996).

[10] *Taking Science to School: Learning and Teaching Science in Grades K–8*, edited by R. A. Duschl, H. A. Schweingruber, and A. W. Shouse (National Academy Press, Washington, DC, 2007).

[11] R. Czujko, in *The Changing Role of Physics Departments in Modern Universities*, edited by E. F. Redish and J. S. Rigden, AIP Conf. Proc. No. 399 (AIP, Woodbury, NY, 1997), p. 213.

[12] J. D. Bransford, A. L. Brown, and R. R. Cocking, *How People Learn: Brain, Mind, Experience, and School* (National Academy

Press, Washington, DC, 1999).

[13] E. Etkina and A. Van Heuvelen, Investigative Science Learning Environment: Using the processes of science and cognitive strategies to learn physics, in *Proceedings of the 2001 Physics Education Research Conference*, edited by S. Franklin, J. Marx, and K. Cummings, (Rochester, NY, 2001), p. 17.

[14] E. Etkina, A. Van Heuvelen, S. White-Brahmia, D. Brookes, M. Gentile, S. Murthy, D. Rosengrant, and A. Warren, "Scientific abilities and their assessment," Phys. Rev. ST Phys. Educ. Res. **2**, 020103 (2006).

[15] A. Lawson, "What Does Galileo's Discovery of Jupiter's Moons Tell Us About the Process of Scientific Discovery?," Sci. Educ. **11**, 1 (2002).

[16] E. Mazur, *Peer Instruction: A User's Manual* (Prentice-Hall, Upper Saddle River, NJ, 1997).

[17] P. Heller and M. Hollabaugh, "Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups," Am. J. Phys. **60**, 637 (1992).

[18] E. Etkina, S. Murthy, and X. Zou, "Using introductory labs to engage students in experimental design," Am. J. Phys. **74**, 979 (2006).

[19] S. Murthy and E. Etkina, in *Physics Education Research Conference*, edited by J. Marx and S. Franklin, AIP Conf. Proc. No. 790 (AIP, Melville, NY, 2004), p. 133.

[20] A complete set of rubrics is available at http://paer.rutgers.edu/scientificabilities

[21] E. Etkina, A. Van Heuvelen, D. Brookes, S. Murthy, A. Karelina, and M. Villasenor, in *Proceedings of a National STEM Assessment Conference*, Washington, DC, October 19–21, 2006, edited by Donald G. Deeds and Bruce W. Callen, National Science Foundation and Drury University, 2008.

[22] V. Otero, The process of learning about static electricity and the role of the computer simulator, Ph.D. Thesis (San Diego University, 2001), http://education.colorado.edu/faculty/oterov/Otero_Dissertation.htm

[23] P. Adey and M. Shayer, An Exploration of Long-Term Far-Transfer Effects Following an Extended Intervention Program in the High School Science Curriculum, Cognition and instruction 11, 1 (1993).

[24] A. F. Artzt and E. Armour-Thomas, Development of a Cognitive-Metacognitive Framework for Protocol Analysis of Mathematical Problem Solving in Small Groups, Cognition and instruction 9, 137 (1992).

[25] R. Lippmann and the Physics Education Research Group, Analyzing students' use of metacognition during laboratory activities, in *AREA Meeting* (New Orleans, 2002), available at http://www.physics.umd.edu/perg/papers/lippmann/meta_lab.pdf

[26] R. Lippman Kung, A. Danielson, and C. Linder, Metacognition in the students laboratory: Is increased metacognition necessarily better, in *EARLI symposium* (2005), available at http://www.anst.uu.se/rekun676/meta.pdf

[27] A. Karelina and E. Etkina, in *Physics Education Research Conference*, edited by Laura McCullough, Leon Hsu and Paula Heron. AIP Conf. Proc. No. 883 (AIP, Melville, NY, 2007), p. 93.