Using action research to improve learning and formative assessment to conduct research

Eugenia Etkina,[1] Anna Karelina,[1] Sahana Murthy,[2] and Maria Ruibal-Villasenor[1]

[1]*Graduate School of Education, Rutgers University, New Brunswick, New Jersey 08904, USA*
[2]*Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

The paper reports on how educational research informed and supported both the process of refinement of introductory physics laboratory instruction and student development of scientific abilities. In particular we focus on how the action research approach paradigm combined with instructional approaches such as scaffolding and formative assessment can be used to design the learning environment, investigate student learning, revise curriculum materials, and conduct subsequent assessment. As the result of the above efforts we found improvement in students' scientific abilities over the course of three years. We suggest that the process used to improve the curriculum under study can be extended to many instructional innovations.

I. INTRODUCTION

This paper describes how the physics education research (PER) group at Rutgers University has developed and refined curricular materials for the laboratory component of "Physics for the Sciences," which is an introductory algebra-based course for science majors. Of equal importance, we report on how educational research informed, supported and was used to evaluate our progress. Over a period of several years we developed a student-centered curriculum, ISLE (Investigative Science Learning Environment) and the corresponding active learning environment to engage students in guided inquiry.[1] In parallel with the design and improvement of this course, we investigated student learning, and the practices and interventions that facilitate it.

The work on designing and implementing the ISLE curriculum is deeply intertwined with our research on knowledge construction and facilitation, as both the course and the group's insights related to science learning have developed concurrently and their influences have been mutual. The course has provided the opportunities to investigate student science learning in a real context, thus giving ecological validity to our findings. Besides, this course with its particular demands and goals, its academic context, and student population has shaped the group's research. Reciprocally, we have built "Physics for the Sciences" based on the results of our research and, furthermore, these results have guided us in the improvement of ISLE—a physics learning system where students construct their knowledge and, at the same time, assimilate the practices of the scientific community by emulating the work of scientists.[1,2]

In this light, this substantial part of the work of the PER group at Rutgers can be framed in the "action-research" tradition where educators create and improve teaching practices and curricula in repeated cycles of inquiry and interventions. In the PER community an example of adhering to the action-research paradigm in curriculum development is the process followed by the University of Washington's PER group which had an established and well-documents practice of developing physics curriculum based on research findings.[3,4]

This paper focuses on the relationship between modifications of curriculum materials and student learning. Specifically we investigated the relationship between student acquisition of scientific abilities in ISLE design labs and the structure of the lab and the scaffolding questions and prompts in the laboratory handouts.[5–7] Thus the work we present here is different from "traditional" studies of the relationship between research and learning which mostly focus on student learning of physics content.

With this study we seek to answer two types of questions. Questions 1–3 are generic questions that relate to any instructional innovation and questions 4–6 relate to the investigation of improvements of lab instruction.

Research questions:

(1) How does the "action-research" approach to the revision of the curriculum material work in practice?

(2) How might instructors manage the introduction of new curricular materials? Should the changes be made and implemented simultaneously or progressively?

(3) What happens to the quality of student work when the students are required to put more effort towards a particular aspect of learning?

(4) What can be done to improve the quality of students' experimental investigations?

(5) What are some of the steps that instructors may take to better support students' lab work?

(6) When does instructor help become inefficient or counterproductive?

II. THEORETICAL FRAMEWORK

How does one develop a successful instructional innovation? Following the design sequence proposed by Dick and Carey, we consider that deciding the goals of instruction should be the first step.[8] The next step is to determine the abilities and behaviors needed to achieve the chosen goals. We then establish the performance objectives and choose the assessment tools. Only at this point do we proceed with the development of instructional strategies and materials that will help us attain the goals. To ascertain how well the intervention works, we need to conduct formative assessments and revise the instructional materials accordingly. The process of trying and modifying strategies and materials usually

needs to be repeated several times in order to attain the desired results.

We find an analogous sequence of steps both in the action research-approach and in the formative assessment approach. Below we will show that the two can be combined together when we develop curriculum materials.

A. Action research

Action research is an inquiry tradition that researchers of different social sciences follow in order to understand complex situations and interactions, and to improve practice. In the action-research paradigm, the practitioners are at the same time the researchers. The knowledge product emerges as they collect data, and formulate and refine hypotheses. Therefore, they improve their practices and advance knowledge, bridging theory and practice at the same time. Kolb (1984) described action research as a learning cycle through which people construct knowledge by analyzing social contexts and learners performance, create theoretical models that reflect their experiences, and test these models in new situations.[9]

B. Formative assessment

In their seminal paper, Black and Wiliam showed that formative assessment is one of the most effective educational interventions.[10] Formative assessment is assessment of student learning in which the goal is to provide feedback for the immediate improvement of the teaching and learning process (pp. 7 and 8, Ref. [10]). In their later book, Black and colleagues[11] expanded ideas of formative assessment in a theoretical framework called AFL—assessment for learning, in which the primary goal of assessment is to promote student learning. Within this framework, instructors modify their teaching, and students modify their learning based on the feedback provided in the assessment activity. Without changes in the instructional process based on the student feedback, the process of formative assessment is not complete. If teachers do not revise their instruction based on student feedback, the improvement does not occur.

As we see from the above descriptions, the AFL framework and the action-research framework have the same pattern. Therefore we can integrate them together to simultaneously refine our understanding of student learning and improve that learning. In this paper, we will show how this theoretical framework can be applied to a specific example of scaffolding of the development of scientific abilities in introductory physics labs.

C. Scaffolding

If one of the goals of instruction is to help students think like scientists, then we need to recognize that the thinking processes and the procedures used by scientists are very demanding for students because they require the exercise of a large and complex set of methods and practices that we call scientific abilities. Facing these considerable difficulties, laboratory instructors can opt to make things easier by telling students what and how to do (this is done in traditional in-

structional laboratories). However this approach of presenting learners with an oversimplified version of scientific experimentation deters them from understanding and seeing the value of scientific reasoning and strategies.[12]

We believe that students can successfully engage in meaningful experimental investigations if we provide them with the necessary support and learning opportunities. This support comes in the form of scaffolding. Scaffolding is a temporary support provided by the instructor to assist learners in two different ways: first by allowing them to complete tasks that they otherwise would not be able to accomplish, and second by preparing them to learn successfully by themselves without external aids.[13–15] The support is then gradually withdrawn, so that the learners assume more responsibility and eventually become independent.

In our ISLE design labs students often encounter unfamiliar and complex tasks that require them to work in the zone of proximal development.[16] Vygotsky's zone of proximal development is "the distance between the level of an individual's independent problem solving and her or his problem-solving capabilities, under guidance or in collaboration with peers" (p. 86). In such situations, scaffolding becomes a necessary and crucial part of instruction. The second function of scaffolding applies to ISLE labs as well. The labs include rubrics, prompts, questions, and exercises that are meant to help students construct an understanding of the principles underlining scientific investigations, and therefore to prepare them for generating new knowledge and solving new problems. One of our major goals is that when students leave the physics lab, they retain this understanding and are able to activate their scientific abilities in different contexts. Hmelo-Silver distinguished between two types of scaffolding: black-box and glass-box scaffolding. Black box scaffolding facilitates learning by performing the task *for* the learner, without needing the learner to understand the process. In contrast, glass box scaffolding makes implicit processes explicit (allowing the learners to understand what support they receive and why they need it), so that ultimately they internalize the scaffolding and become independent learners.[17] The scientific abilities rubrics that we provide in ISLE labs are a clear example of glass-box scaffolding.

Holton and Clarke classified scaffolding by agent (expert, peer, or self-scaffolding) and by domain (conceptual or heuristic scaffolding).[18] To promote the ultimate goal of learners' independence, instruction must gradually shift the source of scaffolding from the teacher (or expert) to the learners. Learners' self-scaffolding is virtually equivalent to their own construction of knowledge, sustained by metacognitive thinking. Reassigning the charge of providing scaffolding from experts to learners opens the possibility of learners progressively taking responsibility and directing their own learning. This shift in the agent who delivers scaffolding is possible through peer scaffolding during collaborative work (also referred to as reciprocal scaffolding) as well as instructors' heuristic scaffolding. Heuristic scaffolding refers to the support that enhances learners' development of their own approaches and procedures for learning or problem solving, while the purpose of conceptual scaffolding is to provide assistance in the acquisition of new concepts. For example, when students are solving a problem involving Newton's

second law, we can ask them to draw a force diagram. This is heuristic scaffolding. When we ask them whether the magnitude of the normal force in this particular situation is greater or less than the magnitude of the gravitational force exerted on the system, this is an example of conceptual scaffolding. There are many different ways of providing heuristic scaffolding for learners, such as asking them generic questions to help focus on important aspects of the procedure, or questions that encourage students to reflect on their actions and choices. For example, Schoenfeld proposed that instructors should ask the following three questions at any time: *What (exactly) are you doing? Why are you doing it? How does it help you?*[19] Other researchers have recommended the use of sets of cards to help students recall appropriate metacognitive actions: *I thought about what I already knew; I made a plan to work it through; I thought about a different way to solve the problem.*[20] According to Holton and Clarke, heuristic scaffolding facilitates the learning of the approaches and procedures that promote independent knowledge construction and problem solving (p. 131, Ref. [18]).

In our labs, we set the goal to help students develop the heuristics of scientific practices. For this, we use heuristic scaffolding aimed at the development of scientific abilities. This scaffolding is generic and reflective by nature and contains three different modes:

(i) guiding prompts and questions on the handouts;

(ii) scientific abilities rubrics;

(iii) TAs' comments on written work, oral feedback and summary advice.

This study focuses on the revisions and gradual refinement of prompts and questions that we made during three years (2004–2006) based on formative assessment of student work done through scientific abilities rubrics. Such an approach allowed us to conduct an action-research study on student learning of scientific abilities, develop a set of instructional interventions to improve student acquisition of scientific abilities, and simultaneously reflect on the generic features of scaffolding that improve or deter student learning.

III. LABS: GOALS SELECTION, DEVELOPMENT, AND REFINEMENT

The labs described in this study are an integrated component of a lecture-lab-recitation course based on the ISLE curriculum. According to the course structure, students spend 3 h every week in the instructional lab. The goals of the course are to help students construct physics concepts and to develop scientific abilities. To be consistent with these global objectives, the goals of the lab component parallel the course goals, with an even greater emphasis on students' development of scientific abilities.[2,6,7]

To devise individual laboratory sessions we took into account the material that students learned during a particular week in whole-class meetings (the "lecture" component was referred to as whole-class meetings, as there was no traditional "lecture" in the course) and collaborative problem-solving recitations. We identified specific scientific abilities that fit best with the content. Then we determined explicitly and in detail, the learning goals (local goals) for each laboratory session (these are both content and scientific abilities goals). After that, we selected the experimental tasks that were best suited to help students achieve the goals and wrote lab handouts. The handouts have clear heuristic glass-box scaffolding: prompts and questions that focus students' attention on different aspects of scientific inquiry. We also chose the relevant rubrics that helped students plan, monitor, and evaluate their work. As the semester progressed this last component faded and finally disappeared. When students performed the lab, they wrote lab reports in which they described their experiments, answered the handout questions, and self-assessed their work based on selected rubrics. They handed in their reports at the end of the 3 h lab session.

After students handed in their lab reports, we photocopied students' lab reports. At the end of the semester, we scored them according to the same rubrics that students used when working in the labs. We analyzed the scores and determined which were the abilities that students were not developing optimally. We then revised the handouts to provide better scaffolding where needed. We repeated the process every year, giving new students the revised handouts, scoring their reports to assess our work, and modifying the handouts consequently.

IV. DESCRIPTION OF THE STUDY

The study was conducted during three consecutive years (2004–2006) of the implementation of an introductory algebra-based physics course that followed the ISLE curriculum. The course spanned two semesters, and comprised 185 students who were science majors (examples of the majors are: biology, chemistry, environmental science, and exercise science). There were two 55 min lectures, one 80 min recitation, and a 3 h lab per week. Although we did not pretest students we can assume that the student population in the course was relatively stable over those years—the course prerequisites as well as advising strategies did not change. Rutgers' incoming student population did not change over those years either, based on the data collected by the Office of Institutional Research and Academic Planning (incoming freshman SAT scores changed by less than 1% over the last 10 years).

The researchers (the authors of this paper) were associated with the development and the teaching of the course at different stages of the study (for example: MR-V and AK were lab instructors in 2005 and 2006, EE trained lab instructors during 2005 and 2006, SM was course instructor during 2004). All authors were involved with developing and revising labs at various stages of the study.

A. Data sources

For this study we collected the following data: observations of student behavior in the labs, analysis of lab reports, and scores of lab reports based on selected rubrics. Instructors recorded observations of student behavior in the labs during the school year by making notes after each lab. As described in the previous section, the results of students' lab report scores were important for the redesign of the lab ma-
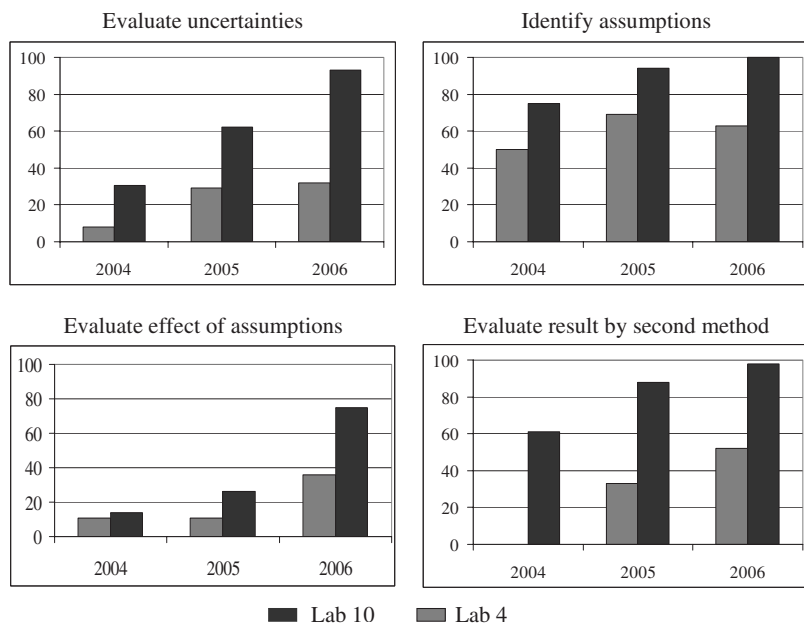
FIG. 1. Students' performance in labs 4 and 10. Bars represent the percentage of the students whose lab reports received high scores: either "adequate" (score of 3) or "needs improvement" (score of 2). We do not have data on the rubric "ability to evaluate the result by second method" in lab 4 in year 2004.

terials. We present the rubric scores of students' lab reports for each of the three years, as the part of this study.

The reliability of the rubric scores of the lab reports was established in the following way: two trained scorers chose four to seven lab reports for each lab for initial scoring. These reports were rescored until the scorers achieved 100% agreement. Then, they proceeded to score about 20% of lab reports in the sample. After the achieved agreement was above 90%, the scorers split the rest of the lab reports and scored them separately.

In this study the rubrics for the following scientific abilities were chosen: ability to evaluate uncertainties, ability to identify assumptions, ability to evaluate effect of assumptions, ability to evaluate result by a second method. See Appendix A for a description of the above rubrics and[2] for a description of all scientific abilities rubrics.

B. Lab development, data collection, and lab revisions

In the labs, students working in groups of three or four designed their own experiments guided by heuristic scaffolding questions. In the first version of the lab handouts this scaffolding contained:

(i) goals of the lab,

(ii) guiding questions that focused students' attention on what to think about when designing the experiment, and

(iii) reflection questions that asked students to connect their lab experiences to everyday experiences.

In addition, students were advised to use the rubrics for self-assessment. The rubrics were a form of expert heuristic scaffolding. At the same time, they encouraged reciprocal scaffolding between students working collaboratively on an experiment. The nature of experiments was such that it required a significant cooperation of group members, as none of the students individually could solve the problems. This naturally yielded to reciprocal scaffolding. However, this reciprocal scaffolding was usually conceptual as students were

mostly focused on how to solve the experimental problem from the content point of view.

The first full implementation of the design labs was in the fall of 2004. We observed students in the labs and made copies of their lab reports. After the semester was over, we compiled the observations and scored the reports for the labs at the beginning and at the end of the semester. We found that from lab 4 to lab 10 students improved significantly on their ability to design an experiment, to devise a mathematical procedure and to communicate the details of the procedure,[2] however, they did not demonstrate similar improvements in their ability to evaluate experimental uncertainties and assumptions (Fig. 1, results for 2004). In addition, a careful examination of each lab showed specific difficulties that students could not overcome. At the end of the spring semester we conducted a survey regarding students' expectations and goals of the labs. The analysis of students' responses showed that they understood that the labs were helping them develop experimentation abilities but did not see the connection between the labs and their future work.[21]

Based on scoring of the reports, observations, and the survey, we revised the labs. The revisions went along several directions: we provided more heuristic scaffolding in addition to the questions in the lab handouts, focusing student attention on experimental uncertainties and validation of assumptions in every lab; we enhanced the reflection questions connecting them to students future jobs; we revised the experiments, and made the handouts more student friendly. Specifically, we added special documents describing in detail how to evaluate experimental uncertainties. We also specified a set of rubrics for self-assessment for each lab, we added a question "why did we do this lab" asking students to connect the abilities that they used in the lab to their future work and we added interesting short stories in each lab handout that showed students the importance of scientific abilities in the everyday world. These stories were very short, related to current physics and biology applications and had the ma-

TABLE I. Steps aimed at improving different abilities in the course.

	2005	2006
Ability to evaluate uncertainties	Text on uncertainty analysis Added scaffolding in each lab	Homework exercises to labs 2, 4, 5, and 9
Ability to identify assumptions	Added scaffolding in each lab	
Ability to evaluate effect of assumptions	Text on evaluation of assumptions	Homework exercise to lab 4
Ability to evaluate result with second method	Text on uncertainty analysis	Homework exercise to lab 4
General	Specified rubrics for each lab task. Added "why did we do this lab?" Added short stories to make the labs lessdry. Decreased the number of experiments. Revised reflective questions.	Removed short stories. Added homework with historical passages. Explicitly stated learning goals of each lab.

jor goals of increasing student motivation. In addition we made changes in each lab that were specific to the content of that particular lab. For example, in several labs we decreased the number of experiments that students had to design to give them more time to reflect and revise their experimental work and writing.

The following year (2005) we repeated the procedure. We found an increase in students' ability to evaluate the uncertainties (Fig. 1, data for 2005) but did not find much improvement in the ability to evaluate the effects of assumptions. Thus for 2006 we added more exercises to help students learn how to evaluate the uncertainty and how to evaluate the assumptions. We also made another change in the lab handouts—we removed the interesting stories as we found them distracting. The last change that was implemented was the lab homework. We realized that students faced a multitude of decisions in the lab when they had to design their own experiments. They could not allocate the time and resources needed to ponder about important aspects of scientific investigations such as the differences in types of scientific experiments.[5] In addition, homework would give students practice in their incipient scientific abilities such us evaluating uncertainty or determining the effects that their assumptions might have on the final results. The lab homework was divided in two categories: (i) practice problems that gave students an opportunity to practice the scientific abilities that we found to be the most difficult to develop, such as estimating the uncertainty in experimental results, and (ii) analysis of historical research (for instance the discovery of prophylactics) or scientific investigations of poignant interest (for research on how the simian immunodeficiency virus could have crossed the species barrier and infected humans). For the second type of exercises students had to read passages describing the investigations. These passages averaged two pages and were adapted from various sources. After reading the passage, students had to answer several questions about the type of experiments described, the sequence of experiments that constituted the entire investigation, and the reasoning that guided the inquiry. These exercises focused students' attention on the dynamics of scientific investigations (how research is motivated and how it progresses and evolves). By reflecting on the passages students were supposed to develop read-out strategies[22] for the elements of scientific research. If students were able to iden-

tify these elements and to recognize the different aspects of the scientific modus operandi, the students would acquire a more advanced and nuanced understanding of what a scientific investigation entails. These deeper insights could affect their own scientific inquiry positively, because it became more deliberate and better rooted in the practices of the scientific community. It is important to note how these homework assignments were different from the interesting stories that we removed from the lab handouts. While the primary goal of "interesting stories" was to increase student motivation, the historical passages had the purpose of showing students how a specific piece of scientific knowledge was constructed over the years. They were much longer than the interesting stories, and ended with questions aimed at the understanding of the process involved in the discovery.

At the end of the 2006 we had the following elements in the lab handouts:

(i) learning goals of the lab that specified what abilities will be the focus of that particular lab;

(ii) rubrics that students need to focus on for self-assessment;

(iii) actual lab tasks—questions that students had to answer during the lab that helped them build a particular ability;

(iv) special lab task-related exercises that focused on the elements of the scientific abilities;

(v) lab reflection questions that were included right in the experimental design part of the lab and at the end of the lab;

(vi) special exercises usually done as homework that allowed students to read and analyze a summary of a historical scientific development (unrelated to the lab task) that illustrated the actual application of a particular ability. Students had to answer questions that helped then reflect about the role of the abilities under analysis.

In addition, the lab instructors provided extensive written and oral feedback to the students. The students also had an opportunity to revise their lab reports after they were graded. After these additions, we repeated the research procedure. Table I shows the steps we took in different years, aimed at the improvement of various abilities.

C. Specific heat lab

Here we describe in detail the process of improvement of the materials for one lab where students had to conduct a

TABLE II. Analysis of students' difficulties in 2004 and the steps we took in 2005 in response. The right column describes the change as related to a specific ability (labeled SA) with a corresponding rubric or a general issue that is not assessed by rubrics.

Students' difficulties	Our steps	Issue
Students did not have time to complete a detailed investigation for the second lab task (specific heat).	We deleted the first lab task. We suggested how to reduce time students spent on writing.	General (time management)
More then 30% of students did not design two independent experiments to determine the specific heat.	We emphasized the importance of the designing two experiments..	SA: Evaluating result by second method.
Students only said that the assumptions will affect the result, but they did not specify how: "If any assumption is false it will change the final outcome of the experiment and our value will be wrong."	We encouraged students to describe the effects of assumptions in detail and provided them with an example.	SA: Identify, evaluate and validate assumptions (increased scaffolding).
Almost 70% of students did not evaluate uncertainties or did it inadequately.	We helped students with the routine steps of evaluating uncertainties.	SA: Evaluate uncertainty
Students often described how uncertainties could be minimized at a superficial level: "Uncertainty lies in the temperature reading and the mass reading. Minimize by using more exact equipment."	We encouraged students to minimize uncertainties in the experiment.	SA: Minimize uncertainties
Only 25% of students were able to evaluate uncertainties adequately and propagate them to the final value.	We emphasized the importance of including uncertainty in the final value.	SA: Evaluate uncertainties
After obtaining the result for specific heat through the second independent method, most students understood that their main assumption that the system's energy was not lost during the experiment was wrong. In most cases students arrived at the correct conclusions on what should be done to improve the results: "We conclude that this relationship is applicable if you do quick transfer of one substance to another." "Based on the results of two experiments they are out of range because our assumptions were wrong. There is a lot heat lost in the transfer of the object…" However they usually did not repeat their experiment using an improved design.	We encouraged students to make a real improvement in the experimental design and repeat the experiment to get a satisfactory result.	SA: Evaluating result by second method

calorimetry experiment (Lab 10 of the semester). We present the original text of the task, the results of scoring of the lab using the scientific abilities rubrics, and the subsequent revisions of the materials. Appendix B contains the three versions of this lab. We have chosen this particular laboratory to show the revision process in detail because of the following reasons:

(i) To complete this lab, students need to apply many different scientific abilities.

(ii) The outcomes of this lab are extremely sensitive to the procedure of the experiment. In all calorimetry experiments where thermal energy escapes the system resulting in a considerable temperature drift which leads to significant systematic error. Therefore, the ways in which the assumptions, implicit in the procedure, may affect the results are especially important to consider. To complete the task successfully students have to be aware of assumptions they make, whether the assumptions are valid, and how the assumptions may affect the result.

(iii) The lab is at the end of the semester; thus we can observe and analyze a cumulative effect of the lab curriculum.

Using information from lab reports written during Fall 2004, and observations made by the lab instructors, we were able to identify the aspects of the lab that needed improvement. In Table II we show the intended goals that students failed to achieve and the changes that we introduced in the lab handout to help students better reach those goals. Students' main problems were with the ability to calculate the uncertainty in their results, and with the ability to determine the way in which the assumptions, implicit in the mathematical procedure, could have affected the results.[23]

In 2005, observations of student behavior in the labs showed that this time students had time to work on the two proposed experiments for specific heat. However, trouble appeared when they compared the results for the specific heat of an unspecified metal obtained from these two experiments and found that the two values were different. As they did not expect to make improvements and repeat the procedure, students were frustrated and disappointed. Another problem was that students were not able to measure the equilibrium temperature with the alcohol thermometers that we made available to them. For example, one group of students wrote: "*Description of the experiment*: We will then quickly put object in the water. We will observe the equilibrium temperature and record it. What went wrong was that equilibrium temperature was read incorrectly."

TABLE III. Analysis of students' difficulties in 2005 and the steps we took in 2006 in response. The right column describes the change as related to a specific ability (labeled SA) with a corresponding rubric or a general issue that is not assessed by rubrics.

Students' difficulties	Our steps	Issue
38% of students did not evaluate uncertainties or they did it but incorrectly.	From the beginning of the semester we added homework exercises to improve students' ability to take into account uncertainties and assumptions.	SA: Evaluate uncertainty. Evaluate and validate assumptions.
Students did not expect to repeat experiments and felt frustrated and disappointed.	We warned students beforehand about the possibility of repeating the experiment.	General (students' expectations)
Students were not able to measure the equilibrium temperature with alcohol thermometers.	We substituted digital thermometers for alcohol thermometers thus allowing students to observe temperature change in time.	General (equipment issue)
Many students readily took the example assumption as something that they had to write.	We deleted the example.	SA: Identify and evaluate assumptions.
Many students did not know how to validate assumptions: "Our assumptions are valid because otherwise our value for c would be inaccurate."	We gave more detailed instructions on how to determine whether an assumption is valid.	SA: Evaluate and validate assumptions

The scoring at the end of the 2005 Fall semester showed an improvement on all of the abilities, still many students had difficulties evaluating uncertainties and considering assumptions. Many students readily repeated the assumptions given as the example in the handout without developing any further ideas. This made us speculate that giving learners this type of support prevents them from advancing toward the goals or, at least, it slows down their progress. It was also apparent that many students did not know how to validate assumptions. Remarks as the following were common: "Our assumptions are valid because otherwise our value for c would be inaccurate." On the other hand, the reports demonstrated a substantial improvement with respect the previous year as some students were able to determine how assumptions might affect the results "If the assumptions were incorrect heat would escape in experiment 1 and heat would enter in experiment 2. This would make our specific heat appear smaller in experiment 1 and would appear larger in experiment 2…"

Some students revised their experiments, and thereby improved them: "The two results do not match. This is most likely due to a failure to have the object in one of its baths for the appropriate time… New results $c=589.6 \text{ J}/\,^{\circ}\text{C kg}$… We did not have time to repeat the second experiment. However, the first experiment was improved, and moving the object faster and measuring faster allowed for better results."

To respond to students' difficulties and frustration we adjusted the handout again seeking to address students' persistent difficulties. In Table III we show the students' difficulties that we encountered during the 2005 lab implementation as well as after analyzing students' reports and the corresponding changes that we made to tackle them in 2006.

In summary, by analyzing lab reports (scoring them with the rubrics) and observing students' behavior in the labs we assessed students' learning, in particular, their acquisition of scientific abilities. We identified the areas in need of improvement and we modified the lab handout to address those weaknesses by providing better scaffolding. We repeated the same process twice and we evaluated the adequacy of the changes by comparing students' performance (as shown in their reports) in different years. Table IV presents the number and kind of changes that we introduced during the years 2005 and 2006 on the handout for the calorimetry lab.

TABLE IV. Number of steps aimed on improving different abilities in lab 10.

	2005	2006
Ability to evaluate uncertainties	3 (2 for evaluating and 1 for minimizing)	0
Ability to identify assumptions	1 (increased scaffolding)	2 (reduced scaffolding)
Ability to evaluate effect of assumptions	1	2
Ability to evaluate result with second method	2	0
General	3	3

TABLE V. Chi-square for the differences in scores in lab 4 between consecutive years (df=3). Note: a single asterisk represents significant difference and triple asterisks represent highly significant differences.

	From 2004 to 2005	From 2005 to 2006	When did change occur?
Ability to evaluate uncertainties	$\chi^2=19.3$*** $p<0.001$	$\chi^2=7.9$* $p=0.049$	First year mostly
Ability to identify assumptions	$\chi^2=5.8$ $p=0.12$	$\chi^2=10.4$* $p=0.02$	Second year slightly
Ability to evaluate the effects of assumptions	$\chi^2=4$ $p=0.26$	$\chi^2=10.1$* $p=0.02$	Second year slightly
Ability to evaluate the result with second method	N/A	$\chi^2=10.4$* $p=0.02$	N/A

V. RESULTS

Prior to this project, our group had identified 48 different scientific abilities that apply to various aspects of scientific investigations, and had developed corresponding rubrics that describe four distinctive levels of competence for each of the abilities.[2] However for the sake of clarity in our exposition and due to space considerations, we have chosen to report on only the four scientific abilities that have been found most difficult to acquire: specifically, the ability to estimate the uncertainty in the result, the ability to identify the assumptions implicit in the mathematical procedure, the ability to evaluate how the assumptions may affect the result, and the ability to evaluate the result by means of an independent method.

Figure 1 presents the percentage of students who received high scores on different scientific abilities in labs 4 and 10 as assessed by the rubrics over the course of three years. We can find three patterns on the graphs: (1) students' performance on scientific abilities improved during the course of a single semester, every year; (2) students competence at the end of a semester improved steadily during these three years, and (3) for some of the abilities the changes made in the first three labs resulted in the improved performance on lab 4.

We used the chi-square test to evaluate the difference in students' performance between consecutive years both at the beginning (lab 4) and at the end (lab 10) of the fall semester. For lab 4 the difference in student achievement was nonexistent or very small between successive years for most of abilities except the ability to evaluate uncertainty (see Table V). For this last ability there is a high significant difference between years 2004 and 2005. For the ability to evaluate the result by means of an independent method we do not have data available for 2004.

Specific heat lab

In this section we report with more detail on the level of competence—in terms of scientific abilities—demonstrated in students' reports for the calorimetry lab. The task for this lab was to find the specific heat of an unknown material. The rubrics scores of students' lab reports for the four abilities that are most difficult to attain improved significantly year after year. Figure 2 clearly reflects this considerable improvement. The darker top portion of the bars represents the percentage of students who received perfect scores. This percentage increased through time for all four scientific abilities. For a detailed statistical analysis of the results see Table VI which shows that this increase is statistically significant. This table displays the χ^2 values for the change in scientific abilities scores between consecutive years.

Students demonstrated in their reports a higher competence level on each of the abilities year after year. This competence was not built up simultaneously for all the abilities and the changes followed a different pattern for each of the abilities. We presume from the data that the improvement demonstrated in each of the abilities at certain points in time reflects the focus of our efforts (Table VI).

In the following paragraphs we present an account of how the modifications that we made each year on the lab hand-
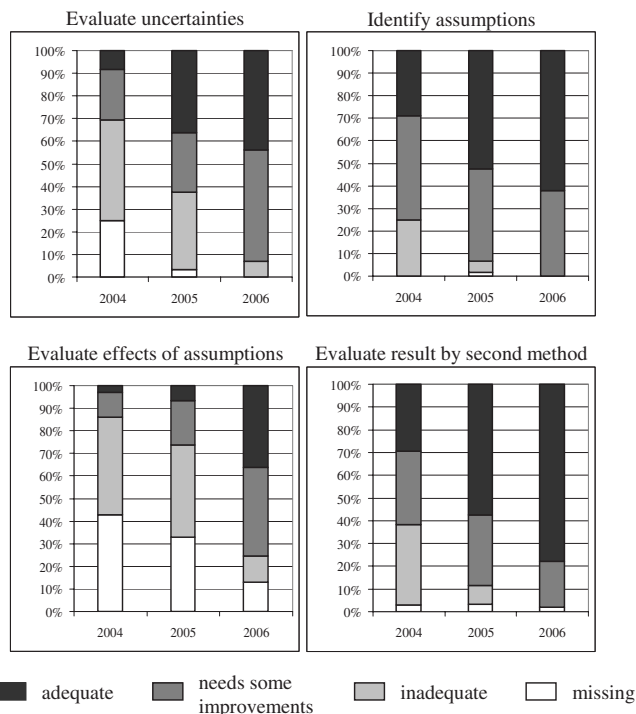

FIG. 2. Student scientific abilities scores for the calorimetry lab during 2004–2006. Bars represent the percentage of the students whose lab reports received scores shown at the bottom of the figure. Notice the decrease in the height of the white bars and the increase in the black ones.

TABLE VI. Chi-square for the differences in scores in the lab on calorimetry (df=3).

	From 2004 to 2005	From 2005 to 2006	When did changes occur?
Ability to evaluate uncertainties	$\chi^2=16.9$*** $p<0.001$	$\chi^2=18.3$*** $p<0.001$	Both years
Ability to identify assumptions	$\chi^2=13.5$** $p=0.004$	$\chi^2=3.6$ $p=0.16$	First year
Ability to evaluate effect of assumptions	$\chi^2=3.4$ $p=0.34$	$\chi^2=31.7$*** $p<0.001$	Second year
Ability to evaluate result with second method	$\chi^2=16.3$*** $p<0.001$	$\chi^2=8.95$* $p=0.03$	First year mostly

outs to address the inadequate acquisition of each of the chosen four abilities may have impacted students' performance.

Ability to evaluate uncertainties. Table III shows the significant changes that we made for 2005 to emphasize the necessity to evaluate experimental uncertainties. As a result, in the year 2005 the number of students who did not attempt to evaluate uncertainties was virtually zero. In 2005 students showed a significant improvement in mastering the ability to evaluate uncertainties comparing to year 2004. Over four times as many students received perfect rubric scores (black bar in Fig. 2) in this ability in 2005 (35%) compared to 2004 (8%). During 2006, we implemented in labs 1, 3, 4, and 8 the set of homework exercises specially designed to facilitate the development of this ability. The homework showed students how to evaluate uncertainties and gave them the opportunity to practice before the lab where many other things need to be attended to. As a result, in 2006 students' performance on this ability was even better than in 2005. If we combine the students who received scores of "needs improvement" (i.e. most uncertainties are evaluated, see Appendix A) and "adequate" (all uncertainties are evaluated correctly), then over 90% of students were able to master this ability in 2006 (black and dark grey bar in Fig. 2).

Ability to identify assumptions. In 2005 we attempted to address students' difficulty in identifying assumptions. The modified handouts used in 2005 had explicit scaffolding that resulted in much higher scores for the ability to identify assumptions than the scores in 2004. However, our observations suggested that these scores might not represent real improvement because instead of scaffolding students' own work we were giving the answers in the handouts. In 2006, we paid special attention to this ability during the course, made new changes in the lab handouts, and removed explicit answers. As a result, in 2006, when having to figure out by themselves all the implicit assumptions, students received scores as high as in 2005. Over 90% of students in 2005 and all students in 2006 were able to identify most of the assumptions (score: "needs improvement" and "adequate" in Fig. 2).

Ability to evaluate the effects of assumptions and validate them. We found that the ability to determine how assumptions, implicit in the procedure, might affect the results was an ability difficult to master. In Fig. 2, note the combined percentage of students who received rubric scores corresponding to "needs improvement" (dark grey) and "ad-

equate" (black). This percentage refers to proportion of students who were at least able to evaluate the effects of assumptions (score: needs improvement, see Appendix A) and some who were also able to validate their assumptions (score: adequate). We shall refer to these combined scores as high scores. Only 15% of students received high scores in 2004. Our attempt to scaffold students in this ability was unsuccessful in the year 2004 or in 2005. Students' competence on this ability was slightly better in 2005 (25% received high scores) than in 2004, but the difference was insignificant. In 2006, we made several important changes in several of the handouts and developed targeted homework exercises to be completed in the fourth week of the lab. A large improvement was seen in 2006: 75% of students received high scores on this ability.

These results appear to indicate that by providing students with the "wrong" type of support, we may hinder their learning. In the year 2005 when we tried to help students by telling them specifically what some of the assumptions implicit in the mathematical procedure were; their scores improved slightly, however they were just repeating the answers given in the handout without any further elaboration. Interestingly, at the same time that year they had a great difficulty explaining the implications of these assumptions. We believe that by helping students *too much* we removed the agency from them. In this sense too much scaffolding or the wrong type has detrimental effects.

The ability to evaluate the result by second method. The revisions in 2005 were mostly focused on this ability (see Table III). This led to a significant decrease in the number of students who evaluated the result inadequately by second method. There is a significant difference in the general performance on this ability between 2004 and 2005. In 2006 we did not make any additional changes with relation to this ability. However, as this ability is closely connected to the ability to evaluate uncertainties and the effects of assumptions, other improvements may have affected this ability as well. As a result, students scored higher on this ability, although we did not take any specific actions to improve it.

VI. DISCUSSION AND IMPLICATIONS

In this paper we have shown that during the term of one semester, undergraduate students can develop competence in their exercise of scientific abilities (these are processes, pro-

cedures, and strategies that scientists apply when developing new understandings or solving problems). We believe that learning of these complex abilities was so fast and efficient because of a thoughtfully designed learning environment. If we assume that the populations of students that took the course were alike,[24] with similar aptitudes and comparable previous knowledge, the result implies that the modifications we made to the lab handouts, together with the homework targeted to the development of scientific abilities increased the course effectiveness significantly.

The research we have reported in this paper focused not only on the instructional practice but also on product design, as we sought both to establish a model for developing innovations and to produce an apt set of instructional strategies and materials. Therefore we intended to achieve two types of goals: first we wanted to devise a viable and effectual methodology to improve laboratory instruction and second, but not less important, we wanted to generate high quality materials and determine their relevant features. We posed six research questions at the beginning of the paper. Below we will discuss how we answered each of them.

Question 1: How does the "action-research" approach to the revision of the curriculum material work in practice?

According to our goals, this study can be classified as Design Research because it attempts to give responses to fundamental questions but it is inspired by and rooted in practice. Three characteristics of Design Research are applicable to our research: (i) it addresses complex problems in real contexts, (ii) it connects design principles with the affordances of the environment to generate solutions, and (iii) it engages in reflective inquiry to refine instruction and to define new design principles.[25,26]

According to the revision procedure that we followed, our work can be framed in the action-research tradition because it can be described as a fine tuning of instruction to the learners' needs though repeated cycles of planning, implementation, and assessment.[27] How does the "action-research" approach to the revision of the curriculum material work in practice? For the design and refinement of ISLE lab curriculum and instructional materials, we followed a sequence of steps very alike to Dick and Carey's model of systematic design of instruction.[8] First of all, we thoughtfully determined the instructional goals. The goals directed all the subsequent design decisions: the selection of the most appropriate experiment tasks and their progression, the choice of the laboratory equipment, the writing of handouts, and the selection of additional exercises. Our set of nuanced and detailed performance goals can be summarized in a single statement: students should demonstrate approaches and behaviors similar to those of scientists in the laboratory.

The design and refinement of ISLE laboratories is a complex undertaking and therefore it required multiple cycles of revisions and implementations. We integrated the assessment for learning paradigm into the action-research framework. In AFL, assessment should inform instruction.[11] Assessment functions as a diagnostic tool to find out what students know and are able to do, determining the gaps in their knowledge so that instructors can adjust their teaching to learners needs. In summary the two key features of AFL are: (i) students' progress is continuously evaluated, and (ii) instruction is modified and remodeled to fit students.

Our approach to curriculum design is not foreign to the PER community but is firmly rooted in the research tradition started by McDermott. She has argued that research must guide the development of the physics curriculum.[4] Researchers should first investigate students understanding and develop educational interventions tailored to learners' previous knowledge and then the instructional materials and strategies must be tested to find if they have a positive impact on students' learning and to determine under what circumstances a specific innovation is effective.[28] An example of the work by McDermott and the PER group at the University of Washington on the development of a curriculum for introductory electricity illustrates how to improve instruction by engaging in a three-phase cycle of design, test, and modification.[29] They concluded that by revising the curriculum, they gained valuable knowledge on how students understand physics (p. 1012). The difference between our work and that of McDermott and colleagues is that we applied this approach to the scientific abilities as opposed to the learning of physics content.

Question 2: How might instructors manage the introduction of new curricular materials? Should the changes be made and implemented simultaneously or progressively?

We have found that there is no need to implement all the changes at once, but instructors can revise different aspects of the curriculum separately and introduce changes as they are made. This way makes it easier to determine what effects correspond to which modification. In addition, revising materials and strategies gradually is a better fit to the fine tuning necessary to respond to learners' needs and characteristics. For instance in the fall semester of 2005 our revision efforts concentrated on students' improvement in their treatment of uncertainties and the following year we focused on their understanding of assumptions.

Question 3: What happens to the quality of student work when the students are required to put more effort toward a particular aspect of learning?

Some may argue that when instructors direct students' attention to the development of a particular scientific ability and expect students to answer more questions and complete more exercises, learners will neglect other aspects of the scientific investigation that they may have previously mastered. We have noticed that, although the cognitive load of ISLE labs is higher than in traditional labs and students tend to resent it during the first weeks, they are able to succeed and indeed they thrive as the semester progresses. A period of adjustment is necessary and the accomplishments can be explained in the different forms of scaffolding embedded in the labs: rubrics, prompts, questions, instructors' feedback, targeted exercises, and peer support (as students always work in groups). Our data clearly show that an increase in scores on a particular scientific ability is not accompanied by any reduction in the other abilities (see Figs. 1 and 2). Our complete list of scientific abilities contains almost 50 different scientific abilities that we use to assess students work and we have never observed that gains in a certain ability was associated with loss in another. In fact, disregarding some normal fluctuations, student improvement is steady and firm as the scores for all the abilities increase (although at different rates) with time.[30]

Our results of students' improvement in scientific abilities might be affected by some other factors. Over three years, different course professors and different lab instructors have taught this course. One might ask the question whether the variability of instructors affected student outcomes. Variables are notoriously hard to isolate in educational research. Possible issues arising from the variability of instructors in our three-year study are in part mitigated by the facts that all course professors and almost all lab instructors have either been part of the PER group, or graduate students in the science education program, where they study many seminal ideas of PER. Lab instructor training was conducted by one of the researchers so that lab instructors were aware of the goals, techniques, and pedagogical content associated with every lab. In addition, great care was taken to integrate the different components of the course to make all the curriculum materials and activities consistent with the ISLE approach, which is inherently a student-centered activity. Another limitation is that we do not have firm evidence that student populations were alike to begin with in terms of their scientific abilities. However, the steadily increasing trend of improvement of scientific abilities over multiple years is a strong indication that it is the targeted instruction, materials, and student experience in labs that has affected the positive outcome. A third limitation is that some instructors initially find the scientific abilities rubrics daunting to work with. Our response is that one can choose which abilities the instructors want to focus on at different points in their course. Mastery of scientific abilities can be gradually built.

Question 4: What can be done to improve the quality of students' experimental investigations?

We have developed a curriculum to improve student learning in introductory physics labs in a large enrolment course. The ISLE labs are our answer to the question on what can be done to improve the quality of introductory instructional labs. In these reformed laboratories, students learn and are able to exhibit the same processes employed by scientists in their research, over the length of a single semester. We have found that there is no need to oversimplify laboratory assignments to ensure that students are able to complete the tasks. Although it is true that engaging in scientific inquiry requires a very complex set of abilities and it presents considerable demands for students, they can succeed in the challenge and demonstrate these abilities when provided with the necessary support and learning opportunities. On a separate note we want to clarify that students' learning of physics concepts does not suffer in ISLE where they do not receive the right answers or procedures. We are able to make this affirmation because in 2006 we conducted an experiment splitting the students in two groups, one attended ISLE design labs and the other one attended nondesign labs. The rest of the treatment was identical for the two groups. The students obtained similar grades in the common exams but the nondesign students did not acquire scientific abilities as well as the design students.[31]

Question 5: What are some of the steps that instructors may take to better support students' lab work?

From the results of the study we draw the following implications for the features of the instructional materials. Lab handouts have to include prompts and questions to direct students' attention to important aspects of the experiments investigation that otherwise students will ignore. Scaffolding should be mainly heuristic, that is, scaffolding directed to the development of approaches and procedures instead of concepts.[18] For those scientific abilities that present the biggest challenges, additional exercises tailored to address the difficulties may be given. Instructors might present the students with accounts of actual research done by scientists stressing the reasoning processes that underline the investigations. Finally, rubrics are an indispensable tool to facilitate the development of scientific abilities as they break down complex procedures into elemental steps and provide concrete descriptors of different levels of competence for each of them. In this paper we have shown that the approach is remarkably successful.

Question 6: When does instructor help become inefficient or counterproductive?

After establishing that scaffolding is indispensable for completing demanding tasks such as in ISLE design labs, we may ask if there is any type of scaffolding that is inefficient or counterproductive for supporting students in their experimental work and in their learning. We have found that overscaffolding or overfacilitating student work had negative consequences. It is crucial that instructors do not remove the agency from students.[32,33] Students need to create their own solutions to be able to amend them later because, as we know, knowledge can not be transmitted but must be recreated in the minds of learners.[34] Instructors are often tempted to "help students too much." As a result they deactivate instruction assigning simple one-step exercises in place of meaningful challenges. We think that scaffolding should be heuristic and as general as possible designed to support student work but not to do the work for them. Results that point in this same direction were obtained by Davis when she studied the effects prompts for reflection with middle school student population.[35] The researcher compared generic prompts (unspecific calls to think as they worked on their science project) and directed prompts (that offer students hints about in which direction to think). A priori directed prompts may seem to be a better option because generic prompts appear insufficient. However, Davis found that, when working on complex science projects, the students who were given generic prompts were able to develop more articulate insights than the students who received directed prompts.

In the future most of the students will not use most of the concepts that they learned in introductory physics courses; many of them will not conduct scientific research (this does not mean we think that scientific abilities are irrelevant for our every day lives). Therefore the most important aim of our introductory courses should be the creation of an essential cognitive and behavioral residual: the ability to learn by their own means. In our program we seek that students become progressively acculturated in the scientific way of producing knowledge. For this reason, scaffolding is such a focal point of our work.

APPENDIX A. RUBRICS FOR THE ABILITIES USED IN THE STUDY

The following shows the Rubrics for the abilities used in the study:

Scientific ability	Missing (0)	Inadequate (1)	Needs some improvement (2)	Adequate (3)
Is able to evaluate specifically how experimental uncertainties may affect the data.	No attempt is made to evaluate experimental uncertainties.	An attempt is made to evaluate experimental uncertainties, but most are missing, described vaguely, or incorrect.	Most experimental uncertainties are evaluated correctly, though a few contain minor errors, inconsistencies, or omissions.	All experimental uncertainties are correctly evaluated and the final result is written with the percent uncertainty.
Is able to identify the assumptions made in using the mathematical procedure.	No attempt is made to identify any assumptions.	An attempt is made to identify assumptions, but most irrelevant, described vaguely, or incorrect.	Most relevant assumptions are identified.	All relevant assumptions are identified.
Is able to determine specifically the way in which assumptions might affect the results.	No attempt is made to determine the effects of assumptions.	An attempt is made to determine the effects of some assumptions, but most are missing, described vaguely, or incorrect.	The effects of relevant assumptions are determined correctly.	The effects of relevant assumptions are correctly determined and the assumptions are validated.
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.

APPENDIX B. LAB HANDOUT VERSIONS FOR THE CALORIMETRY LAB FOR THREE YEARS. CHANGES ARE SHOWN IN BOLD.

The following shows the lab handout versions for the calorimetry lab for three years:

2004	2005	2006
Task 1. You have a Styrofoam cup, half-filled with hot water, some ice at 0 °C, a weighing scale, a thermometer, and paper towels. Your task is to predict the amount of ice (in grams) that you should add to the cup of hot water in order that the ice and water reach a final temperature that is half the initial temperature of the hot water. Write the following in your notebook: (a) What idea or relationship will you use to make a prediction? (b) Write a brief outline of your procedure, with a labeled sketch. (c) Construct the mathematical procedure needed to make your prediction. (d) List the assumptions you are making in your procedure. How could these assumptions affect the experimental results? (e) What are the possible sources of experimental uncertainty? Use the weakest link rule to estimate the uncertainty in your prediction. (f) Then try the experiment. Did your experimental outcome match your prediction within experimental uncertainty? If not, can you explain the difference using the assumptions you made? (g) What can you conclude about the applicability of the idea/relationship in a) to this particular situation? Note: The ice might not be at 0 °C, so pour some water on it and let it sit for a few minutes. The water-ice solution will reach temperature 0 °C. Then pour off the water, and quickly pat the ice dry with a paper towel and weigh the desired amount. Task 2. Design two independent experiments to determine the specific heat of the given object. It is not known what material the object is made of.	**Specific heat of unknown object (Application experiment)** Design two independent experiments to determine the specific heat of the given object. **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… choose the best two designs For each method, write the following in your lab-report: **(Note: Try to reduce the amount of writing you have to do by referring to earlier points. For example: If some of your assumptions are the same for both methods, just write "see method 1 assumptions.")** (a) Write a verbal description and a draw labeled sketch of the design you chose… (b) List all assumptions you have made in your design. **For example, one of the assumptions in your procedure may be that no energy is lost from the calorimeter to the air. Then you have to determine if this assumption is valid in your experiment, and if not, evaluate whether the specific heat you determined from your experiment is greater than or smaller than the actual value**. (c) 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**. (d) Perform the experiment. **Make sure you take steps to** *minimize experimental uncertainties. Record your measurements in an appropriate format*	**Homework to lab 4: Uncertainty To decide whether two experiments give you the same result, you need to pay careful attention to theoretical assumptions and experimental uncertainties**. **(1). You were measuring how heavy an object is using an electronic platform scale with the smallest increment of** 0.1 g **and got the value of** 250 g. **Estimate the relative uncertainty of your measurement**. **Assumptions After you recorded the reading of the scale, you noticed that the table on which the scale was sitting was tilted a little bit. You measure the angle of the tilt and find it to be about 100. Can you assume that the table is not tilted**? I. Application experiment: Specific heat of unknown object… **Self-assessment** (the letters below indicate the rubrics that students are supposed to use for self-assessment) **D2, D7, D8, D9, D4 Warning: Be prepared to repeat the experiment if necessary**! *Equipment:* Water, ice, container for water, hot plate, calorimeter, weighing balance, and a digital (!) thermometer. First, recall… why it is important to design *two* experiments to determine a quantity. Devise as many designs as possible… Working with your lab partners, choose the best two designs. Indicate the criteria that you used to decide which designs were the "best." **Show it to your lab instructor**. For each method: … (a) Write a verbal description and draw a labeled sketch… (b) Construct the mathematical procedure you will use. (c) List all assumptions you make in your design.

2004	2005	2006
You have access to the following equipment: water, Styrofoam container, weighing balance, thermometer and timer. For each method, write the following in your lab report: (a) First, come up with as many designs as possible to determine the specific heat. Write a brief outline of each procedure you come up with. Then choose the best design. Indicate the criteria that you used to decide which design was "best." (b) Include a verbal description and a labeled sketch of the design you chose. (c) Construct the mathematical procedure you will use. (d) What physical quantities will you measure? (e) List the assumptions you make in your procedure. How could they affect the result? (f) What are the sources of experimental uncertainty? How would you minimize them? (g) Perform the experiment and record your measurements. Make a table for your measurements, if necessary. (h) Calculate the specific heat, based on your procedure and measurements. (i) After you have done both experiments, compare the two outcomes. Discuss if they are close to each other within your experimental uncertainty. That is, if they results are different, can they be explained by the assumptions you made in your procedure? (j) List any shortcomings in the experiment design and how you would address them. Decide why this activity was included in the lab. Think of real life situations [and briefly describe them] in which you need to figure out things similar to this experiment	(e) Calculate the specific heat… **Include experimental uncertainty in each value of specific heat that you determine**. (f) After you have done both experiments, compare the two outcomes. Discuss if they are close to each other within your experimental uncertainty. **If not, specifically explain what might have gone wrong—perhaps one of your assumptions was not valid. If your experimental results are not close to each other within experimental uncertainty, perform the experiment again taking steps to improve your design. For example, you could take all measurements quickly so that hot objects do not cool off, or you could improve the thermal insulation of your calorimeter**.	**Decide which of the assumptions affects your results most**. Explain how the outcome of the experiment depends on this assumption, i.e. if the assumption increases or decreases your result (d) **Design an additional experiment to determine whether the main assumption is valid in your experiment. Quantitatively estimate the effect of this assumption on the value of your measurement and compare it with the instrumental uncertainty…** (e) List sources of experimental uncertainty… estimate the uncertainty in your result… Perform the experiment. Make sure you take steps to minimize experimental uncertainties and the effect of the assumptions… (f) After you have done both experiments, compare the two outcomes… If your experiments are not close to each other within experimental uncertainty explain what might have gone wrong and perform the experiment again taking steps to improve your design…

[1] E. Etkina and A. Van Heuvelen, in *Research Based Reform of University Physics, (AAPT)*, edited by E. F. Redish and P. Cooney (AAPT, Compadre, 2007), http://per-central.org/per_reviews/media/volume/ISLE-2007.pdf

[2] Scientific abilities are approaches and procedures that scientists use when engaging in the construction of knowledge or in solving complex problems. Examples of the abilities are designing experimental investigations, collecting and analyzing data, reasoning from the data, constructing explanatory models and testing them, evaluating reasoning and design, solving complex problems, working with other people, and communicate. 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).

[3] J. McNiff, *Teaching as Learning: An Action Research Approach* (Routledge, London, 1993).

[4] L. C. McDermott, Oersted Medal Lecture 2001: Physics Education Research—The key to student learning, Am. J. Phys. **69**, 1127 (2001).

[5] E. Etkina, A. Van Heuvelen, D. T. Brookes, and D. Mills, Role of experiments in physics instruction: A process approach, Phys. Teach. **40**, 351 (2002).

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

[7] A. Karelina and E. Etkina, Acting like a physicist: Student approach study to experimental design, Phys. Rev. ST Phys. Educ. Res. **3**, 020106 (2007).

[8] W. Dick and L. Carey, *The Systematic Design of Instruction* 3rd ed. (Scott Foresman, Glenview, IL, 1990).

[9] D. Kolb, *Experiential Learning. Experience as the Source of Learning and Development* (Englewood Cliffs, Prentice-Hall, NJ, 1984).

[10] P. Black and D. Wiliam, Assessment and classroom learning, Assessment in Education **5**, 7 (1998).

[11] P. Black, C. Harrison, C. Lee, B. Marshall, and D. William, *Assessment for Learning* (Open University Press, Philadelphia, PA, 2003).

[12] C. A. Chinn and B. A. Malhotra, Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks, Sci. Educ. **86**, 175 (2002).

[13] A. Collins, J. S. Brown, and S. E. Newman, Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics, in *Knowing, Learning, and Instruction: Essays in Honor of Robert Glaser*, edited by L. B. Resnick (LEA, Hillsdale, NJ, 1989), pp. 453–494.

[14] B. J. Reiser, Scaffolding complex learning: The mechanisms of structuring and problematizing student work, J. Learn. Sci. **13**, 273 (2004).

[15] D. Wood, J. S. Bruner, and G. Ross, The role of tutoring in problem solving, J. Child Psychol. Psychiatry **17**, 89 (1976).

[16] L. Vygotsky, Mind in Society (Trans. M. Cole) (Cambridge, MA, Harvard University Press, 1978).

[17] C. E. Hmelo-Silver, Design principles for scaffolding technology-based inquiry, in *Collaborative Learning, Reasoning, and Technology*, edited by A. M. O'Donnell, C. E. Hmelo-Silver, and G. Erkens (Erlbaum, Mahwah, NJ, 2006), pp. 147–170.

[18] D. Holton and D. Clarke, Scaffolding and metacognition, Int. J. Math. Educ. Sci. Technol. **37**, 127 (2006).

[19] A. H. Schoenfeld, Learning to think mathematically: problem solving, metacognition, and sense making in mathematics, in *Handbook of Research on Mathematics Teaching and Learning*, edited by D. A. Grouws (Macmillan Publishing Company, New York, NY, 1992), pp. 334–370.

[20] J. Wilson and D. Clarke, Modeling mathematical metacognition, *Annual Meeting of the American Educational Research Association*, New Orleans, 2002.

[21] E. Etkina and S. Murthy, in *2005 Physics Education Research Conference*, edited by P. Heron, L. McCullough, and J. Marx, AIP Conf. Proc. No. 818 (AIP, Salt Lake City, UT, 2005), pp. 97–100.

[22] C. Bereiter, Aspects of an educational learning theory, Rev. Educ. Res. **4**, 603 (1990).

[23] A. Karelina, S. Murthy, M. R. Ruibal-Villasenor, and E. Etkina, *Tricky Calorimetry: Making Sense of the Real World*, edited by P. Heron, L. McCullough, and J. Marx, AIP Conf. Proc. No. 818 (AIP, Salt Lake City, UT, 2005).

[24] Which is a reasonable assumption as the course requirements and advising did not change over the course of 3 years and Rutgers student population in general improved only by less than 1% on SAT scores over the last 10 years.

[25] A. L. Brown, Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings, J. Learn. Sci. **2**, 141 (1992).

[26] A. Collins, D. Joseph, and K. Bielaczyc, Design Research: Theoretical and Methodological Issues, J. Learn. Sci. **13**, 15 (2004).

[27] P. Riding, S. Fowell, and P. Levy, An action research approach to curriculum development, Information Systems Research **1**, 1 (1995).

[28] L. C. McDermott, Bridging the gap between teaching and learning: *The role of research, Proceedings of the International Conference on Undergraduate Physics Education (ICUPE)*, edited by E. F. Redish and J. Rigden (AIP, Woodbury, NY, 1997), pp. 139–165.

[29] P. S. Shaffer and L. C. McDermott, Research as a guide for curriculum development: An example from introductory electricity. Part II: Design of instructional strategies, Am. J. Phys. **60**, 1003 (1992).

[30] E. Etkina, A. Karelina, and M. R. Ruibal-Villasenor, How long does it take? A study of student acquisition of scientific abilities, Phys. Rev. ST Phys. Educ. Res. **4**, 020108 (2008).

[31] E. Etkina, A. Van Heuvelen, A. Karelina, M. R. Ruibal-Villasenor, D. Rosengrant, C. Hmelo-Silver, and R. Jordan, Transfer of Scientific Abilities, in *2007 Physics Education Research Conference*, edited by L. Hsu, C. Henderson, and L. McCullough, AIP Conf. Proc. No. 951 (AIP, Greensboro, NC, 2007), pp. 92–95.

[32] D. L. Schwartz and T. Martin, Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction, Cogn. Instruct. **22**, 129 (2004).

[33] D. L. Schwartz and J. D. Bransford, A time for telling, Cogn. Instruct. **16**, 475 (1998).

[34] C. H. Liu and R. Matthews, Vygotsky's philosophy: Constructivism and its criticisms examined, Int. J. Educ. Dev. **6**, 386 (2005).

[35] E. A. Davis, Prompting middle school science students for productive reflection: generic and directed prompts, J. Learn. Sci. **12**, 91 (2003).