

## Scientific abilities and their assessment

Eugenia Etkina,\* Alan Van Heuvelen, Suzanne White-Brahmia, David T. Brookes, Michael Gentile, Sahana Murthy, David Rosengrant,\* and Aaron Warren

*Department of Physics and Astronomy, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854-8019, USA*

(Received 20 December 2005; published 1 August 2006)

The paper introduces a set of formative assessment tasks and rubrics that were developed for use in an introductory physics instruction to help students acquire and self-assess various scientific process abilities. We will describe the rubrics, tasks, and the student outcomes in courses where the tasks and rubrics were used.

DOI: [10.1103/PhysRevSTPER.2.020103](https://doi.org/10.1103/PhysRevSTPER.2.020103)

PACS number(s): 01.40.Fk, 01.50.Qb

### I. INTRODUCTION

Several hundred thousand science and engineering students take introductory physics courses each year. What are the goals of these courses? In most courses the goals are to help students acquire a conceptual and quantitative understanding of major physics principles and the ability to use this understanding for problem solving. In addition to these goals, we argue that our introductory courses (and, in fact, all courses in physics) should help students develop other abilities that will be useful in their future work. According to many studies our students after leaving academia will be asked to solve complex problems, design experiments, and work with other people.<sup>1-4</sup> Several documents guiding the K-16 program design and evaluation incorporate the development of these abilities as primary goals. National and international science tests at the 4-12th level have items assessing how students achieve these goals (reference to NAEP, TIMSS and the UK tests). Another example here is the new accreditation requirements for engineering colleges.<sup>5</sup> As opposed to the old checklist of courses taken, engineering colleges must now show that their students have acquired various abilities that are important in the practice of science and engineering. The National Science Standards<sup>6</sup> suggest that students need to learn to (i) identify questions and concepts that guide scientific investigation; (ii) design and conduct scientific investigations; (iii) use technology and mathematics to improve investigations and communications; (iv) formulate and revise scientific explanations and models using logic and evidence; (v) recognize and analyze alternative explanations and models; and (vi) communicate and defend a scientific argument.

Today even the most reformed introductory physics curricula do not focus explicitly on developing these abilities and more importantly on assessing them. The physics education research (PER) community uses summative assessment instruments that tell us whether students mastered the concepts of Newton's laws, thermodynamics, electricity and magnetism, and so on. Physics by Inquiry, Workshop Physics, Interactive Lecture Demonstrations, and the Washington tutorials<sup>7-12</sup> use a formative assessment of student learning in the process of learning, but their focus is also mostly on conceptual understanding. Some reformed curricula such as SCALE-UP (Ref. 13) have recognized some of these goals and implemented strategies to achieve them. However, in the PER community, there are no instruments that assess

whether students can design and conduct investigations, communicate, and defend scientific argument, etc. This does not mean that such instruments do not exist. More than 60 years ago, Kruglak and colleagues developed performance tests to evaluate student achievement in introductory college physics courses at the University of Minnesota. These tests involved the use of laboratory equipment to measure different aspects of experimentation such as control of variables, selection of an appropriate method to answer an experimental question, and analysis and interpretation of experimental data.<sup>14-17</sup> Later, items similar to Kruglak's questions but adapted to the paper-and-pencil environment appeared in national (NAEP) and international (TIMSS-R) science tests for middle school and high school students.<sup>18,19</sup> These tests show the importance of achieving science process goals. Although a valuable resource, the performance tests and paper-and-pencil items on the written tests are summative in nature; they assess the results of learning and do not help students in the process of learning.

In this paper we describe a set of the tasks and formative assessment instruments that can be used to help achieve "science-process" goals, as formulated by the National Research Council, the National Science Foundation, ABET, and others. We also describe the results of using these tasks and instruments in introductory physics courses whose curriculum was designed to specifically achieve these goals in addition to the traditional goals of a physics course.

### II. DEFINING SCIENTIFIC ABILITIES

We use the term "scientific abilities" to describe some of the most important procedures, processes, and methods that scientists use when constructing knowledge and when solving experimental problems. We use the term scientific abilities instead of science-process skills to underscore that these are not automatic skills, but are instead processes that students need to use reflectively and critically.<sup>20</sup> The list of scientific abilities that our physics education research group developed includes (A) the ability to represent physical processes in multiple ways; (B) the ability to devise and test a qualitative explanation or quantitative relationship; (C) the ability to modify a qualitative explanation or quantitative relationship; (D) the ability to design an experimental investigation; (E) the ability to collect and analyze data; (F) the ability to evaluate experimental predictions and outcomes,

conceptual claims, problem solutions, and models, and (G) the ability to communicate.

This list is based on the analysis of the history of the practice of physics,<sup>21–23</sup> the taxonomy of cognitive skills,<sup>24,25</sup> recommendations of science educators,<sup>26</sup> and an analysis of science-process test items.<sup>19</sup>

To help students develop these abilities, one needs to engage students in appropriate activities, and to find ways to assess students' performance on these tasks and to provide timely feedback. Activities that incorporate feedback to the students are called formative assessment activities. As defined by Black and Wiliam, formative assessment activities are “all those activities undertaken by teachers, and by their students in assessing themselves, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged.”<sup>27</sup> These authors reviewed 580 articles and found that learning gains produced by effective use of formative assessment are larger than those found for any other educational intervention (effect sizes of 0.4–0.7). Black and Wiliam also found that self-assessment during formative assessment is more powerful than instructor-provided feedback; meaning the individual, small-group, and large-group feedback system enhances learning more than instructor-guided feedback. Sadler<sup>28</sup> suggested three guiding principles, stated in the form of questions, that students and instructors need to address in order to make formative assessment successful: (1) Where are you trying to go? (Identify and communicate the learning and performance goals.) (2) Where are you now? (Assess, or help the student to self-assess, current levels of understanding.) (3) How can you get there? (Help the student with strategies and skills to reach the goal.)

As noted above, students need to understand the target concept or ability that they are expected to acquire and the criteria for good work relative to that concept or ability. They need to assess their own efforts in light of the criteria. Finally, they need to share responsibility for taking action in light of the feedback. The quality of the feedback rather than its existence or absence is a central point. The feedback should be descriptive and criterion-based as opposed to numerical scoring or letter grades without clear criteria. With all the constraints of modern teaching, including large-enrollment classes and untrained teaching assistants, how can one make formative assessment and self-assessment possible?

One way to implement formative assessment and self-assessment is to use assessment rubrics. An assessment rubric is one of the ways to help students see the learning and performance goals, self-assess their work, and modify it to achieve the goals (three guiding principles as defined by Sadler above). The rubrics contain descriptions of different levels of performance, including the target level. A student or a group of students can use the rubric to help self-assess her or their own work. An instructor can use the rubric to evaluate students' responses and to provide feedback.

### III. FINE-TUNING SCIENTIFIC ABILITIES AND DEVISING RUBRICS TO ASSESS THEM

After making the list of scientific abilities that we wanted our students to develop, we started devising assessment ru-

brics to guide their work. Rubrics are descriptive scoring schemes that are developed by teachers or other evaluators to guide students' efforts.<sup>29</sup> This activity led to a fine-tuning of the abilities, that is, to break each ability into smaller sub-abilities that could be assessed. For example, for the ability to collect and analyze data we identified the following sub-abilities: (i) the ability to identify sources of experimental uncertainty, (ii) the ability to evaluate how experimental uncertainties might affect the data, (iii) the ability to minimize experimental uncertainty, (iv) the ability to record and represent data in a meaningful way, and (v) the ability to analyze data appropriately.

Each item in the rubrics that we developed corresponded to one of the subabilities. We agreed on a scale of 0–3 in the scoring rubrics to describe student work (0, missing; 1, inadequate; 2, needs some improvement; and 3, adequate) and devised descriptions of student work that could merit a particular score. For example, for the subability “to record and represent data in a meaningful way” a score of 0 means that the data are either missing or incomprehensible, a score of 1 means that some important data are missing, a score of 2 means that all important data are present but recorded in a way that requires some effort to comprehend, and a score of 3 means that all important data are present, organized, and recorded clearly.

Simultaneously, while refining the list of abilities, we started devising activities that students could perform in recitations and laboratories. Defining subabilities and developing rubrics to assess them informed the writing of these activities. After we developed the rubrics, we started using them to score samples of student work. Each person in our nine-person group (eight coauthors and one member who left before the project was completed) assigned a score to a given sample using a particular rubric; we then assembled all the scores in a table and discussed the items in the rubrics where the discrepancy was large (See Table I for an example).

Based on these discussions we revised the wording of the rubrics and tested them by scoring another sample of student work. This process was iterated until we achieved 80% or higher agreement among our scores.

In the sections below we list scientific abilities and corresponding subabilities that we identified, provide examples of scoring rubrics that we devised, and discuss where in the instructional process we use the rubrics. For each scientific ability, we provide examples of the tasks written for the students. Often the tasks target several abilities. In subsequent sections we will report on how we used the rubrics to study students' acquisition of some of the suggested abilities.

#### A. Ability to represent physical processes in multiple ways

While constructing and using knowledge, scientists often represent the knowledge in different ways, check for consistency of the representations, and use one representation to help construct another.<sup>30,31</sup> For example, in the 1950s Feynman diagrams helped quantum electrodynamics move forward somewhat more rapidly by providing a more visual and

TABLE I. Scoring student work using the rubrics. For each student write-up, nine people assigned scores based on the descriptors of a particular subability in a rubric. The rubric descriptors with large discrepancies (more than 1 score point) were discussed and revised. In this case the largest discrepancy was in the assessment scores for the subability to represent data in a meaningful way.

| A member of the group scoring student work | Score assigned for the demonstrated ability to identify sources of experimental uncertainty | Score assigned for the demonstrated ability to evaluate how experimental uncertainties might affect the data | Score assigned for the demonstrated ability to record and represent data in a meaningful way | Score assigned for the demonstrated ability to analyze data appropriately |
|--|---|--|--|---|
| Alan                                       | 2   | 1  | 3  | 2   |
| Sahana                                     | 2   | 1  | 2  | 2   |
| David                                      | 2   | 2  | 2  | 2   |
| Michael                                    | 3   | 2  | 1  | 3   |
| Dave                                       | 2   | 2  | 3  | 2   |
| Jen  | 2   | 2  | 2  | 2   |
| Aaron                                      | 2   | 2  | 2  | 2   |
| Suzanne                                    | 2   | 2  | 3  | 2   |
| Marina                                     | 2   | 2  | 1  | 2   |

understandable representation of a scattering process. Rules were also developed for converting these diagrams into complicated scattering cross section equations. Such qualitative representations, particularly diagrammatic or in some cases graphical representations, help physicists reason qualitatively about physical processes and to see patterns in data without engaging in difficult mathematical calculations.

In our introductory physics courses students are often given a verbal description of a physical process and a problem to solve relative to that process. They can start their analysis by constructing a sketch to represent the process and include in the sketch the known information provided in the problem statement. They construct more physical representations that are still relatively easy to understand—for example, motion diagrams, free-body diagrams, qualitative work-energy and impulse-momentum bar charts, ray diagrams, and so forth. Finally, they use these physical representations to help construct a mathematical representation of the process.

What subabilities help to make this multiple representation strategy productive for reasoning and problem solving? (i) The ability to correctly extract information from a representation; (ii) the ability to construct a new representation from another type of representation; (iii) the ability to evaluate the consistency of different representations and modify them when necessary.

In addition to such subabilities that students need to master while using multiple representations, there are specific subabilities needed for each type of representation. For example, to use free-body diagrams (FBDs) productively for problem solving, students must learn to: (i) Choose a system of interest before drawing the diagram; (ii) use force arrows to represent the interactions of the external world with the system object or objects; (iii) label the force arrows with two subscripts (for example, the force exerted by the Earth on the object is labeled as  $\vec{F}_{E \text{ on } O}$ ); (iv) try to make the relative lengths of force arrows consistent with the problem situation (the net force should be in the same direction as the system object’s acceleration); (v) include labeled axes on the diagram.

Such diagrams, if drawn correctly, can be used to help write Newton’s second law in a component form; to represent the situation mathematically. Based on these considerations we constructed a rubric to help students self-assess themselves while drawing FBDs (see Table II).

How can a student use this rubric for self-assessment? After she draws a free-body diagram she can use the rubric to ask herself whether she labeled all the forces with two subscripts to indicate interacting objects or for some of the forces she cannot find a pair of the objects? She can also check whether the number of forces on the diagram is equal to the number of objects that interact with the object of in-

TABLE II. A scoring rubric to assess a free-body diagram.

| Scientific ability                       | 0 (Missing)                          | 1 (Inadequate)   | 2 (Needs improvement)   | 3 (Adequate)   |
|--|--------------------------------------|--|---|--|
| Is able to construct a free-body diagram | No free-body diagram is constructed. | FBD is constructed but contains major errors such as incorrect force vectors; length of vectors; wrong direction; extra incorrect force vector, or missing vector. | FBD contains no errors in vectors but lacks a key feature such as labels of forces with two subscripts; vectors are not drawn from a single point; or axes are missing. | The diagram contains no errors and each force is labeled so that it is clear what each force represents. |

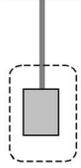
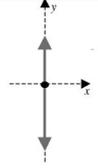
| Write a word description of the situation. | Sketch the situation and circle the system object.                                | Free-body diagram with perpendicular axes. Label the forces if needed.            | Draw a motion diagram Is it consistent with the FBD? | Write Newton's second law in component form. |
|--|---|---|--|--|
| An elevator is slowing down on its way up. |  |  |  |  |

FIG. 1. An example of a multiple representation task (cells 3–5 are not completed).

terest. Then she can check for less significant things such as coordinate axes. Obviously using the rubric for self-assessment requires the knowledge of physics; thus this process cannot proceed completely without an interaction with an instructor. The same is true for using other rubrics for self-assessment: they help students to ask relevant questions about their work but they do not provide answers. Thus a rubric guides a student in the process of self-assessment but does not provide feedback by itself.

We also made a list of several types of multiple representation activities (a task may consist of some combination of these activities). For example: (i) Provide students with one representation and have them create another; (ii) provide students with two or more representations and have them check for consistency between them; (iii) provide students with one representation and have them choose from a multiple-choice list a consistent different type of representation (for example, provide a mathematical description of a process and have students select from a list a consistent word description of the process); (iv) have students use a representation while solving a problem.

An example of a task that helps students develop the subabilities to construct new representations from other representations and to evaluate the consistency of different representations is given in Fig. 1. In addition, this example helps students learn to draw free-body diagrams.

### B. Ability to devise and test a qualitative explanation or quantitative relationship

One of the purposes of science is to explain observed phenomena.<sup>32,33</sup> Hypotheses that scientists generate to explain phenomena need to be testable—this means that they can be used to make predictions about the outcomes of new experiments.<sup>34</sup> If the outcome matches the prediction, it does not mean that the hypothesis under the test is always correct; it only means that the hypothesis was not ruled out by the testing experiment. Thus, it is more productive to try to design an experiment whose actual outcome may not match the prediction based on the hypothesis under the test. However, the outcome of the testing experiment depends not only on the correctness of the hypothesis but also on other auxiliary hypotheses used to make a prediction. These are usually simplifying assumptions about objects, interactions, systems, or

processes involved in the phenomenon.<sup>35</sup> Based on these considerations we identified the following subabilities that we want our students to develop: (i) ability to make a reasonable prediction based on the proposed hypothesis; (ii) ability to identify assumptions used in making the prediction; (iii) Ability to determine specifically the way the assumptions might affect the prediction; (iv) Ability to revise the hypothesis based on new evidence.

For each of the subabilities, we developed an item in the rubrics to guide students. A sample set of rubrics is shown in Appendix A. In item 7 in the set of rubrics in Appendix A, one can see an example of the descriptors used to assess student's ability to make a prediction based on a relationship or explanation. In addition to the assessment goal this item helps students see the importance of using the explanation to make a prediction and asks them to pay attention to the assumptions. Although it does not tell them whether the assumptions they described are correct, it reminds the students to describe the assumptions.

To engage students in testing hypotheses, we provide them with alternative hypotheses that they need to test. This is usually done when students are constructing a new understanding and have to reconcile new ideas with their prior knowledge. We emphasize that they need to try to design an experiment to rule out the hypothesis, not to support it. For example, (i) design an experiment to test the following proposed hypothesis: an object always moves in the direction of the unbalanced force exerted on it by other objects; (ii) design an experiment to test the following proposed hypothesis: in an electric circuit the current is used up by different elements.

The primary setting where students can strengthen their ability to develop and test hypotheses is the laboratory. In place of a laboratory write-up that tells students what to do, we structure the write-up around the rubric abilities that we think it should develop. The lab write-ups provide guidance as shown below.

#### 1. Testing experiment

Your friend says that as current flows through a circuit, it is used up by the elements of the circuit. Design an experiment to test your friend's idea.

*Equipment.* Voltage source, resistors, light bulbs of different ratings, ammeter, voltmeter.

*Write the following in your lab report,*

- State the hypothesis you will test.
- Devise a procedure to test the hypothesis. Write an outline of your design;
- Draw a circuit diagram. Use different resistors or light bulbs as your circuit elements.
- Make a prediction of the outcome of your experiment, based on the hypothesis that you are testing. State your assumptions.
- Perform the experiment. Record your data in an appropriate format.
- What is the outcome of the experiment? Did the outcome match the prediction?
- Based on your prediction and the outcome of the experiment, what is your judgment about the hypothesis you were testing?

TABLE III. A scoring rubric to assess a student's revision of an explanation

| Scientific ability  | 0 (Missing)   | 1 (Inadequate)  | 2 (Needs improvement)  | 3 (Adequate)  |
|---|---|---|--|---|
| Is able to revise the explanation of a prediction, based on the results of an experiment. | No attempt is made to explain the outcome of the experiment, to revise the previous explanation or assumptions. The difference between the prediction and the outcome of the experiment is not addressed. | An attempt is made to explain the outcome and revise the previous explanation or assumptions, but is (a) mostly incomplete and/or (b) based on incorrect reasoning. | The revision of the previous explanation or assumptions is partially complete and correct, yet still lacking in some relevant details. | The revision of the explanation or assumptions is explained completely and correctly. |

(h) How can you modify the hypothesis to account for the outcomes of the experiment?

In Appendix A we provide examples of student work in a laboratory where they design experiments to test proposed hypotheses and rubrics that help students self assess their work. The rubric scores assigned by instructors can be used for research purposes. (Examples are given in Sec. IV.)

### C. Ability to modify a qualitative explanation or quantitative relationship

Another important ability that scientists use in their work is the ability to account for anomalous or unexpected data. Often when a scientist performs an experiment, she obtains some information that seems to contradict her expectations. After performing the experiment she needs to modify the explanation or revisit the simplifying assumptions. We devised “surprising data tasks” that engage students in similar activities. They are used at the stage of learning when students have constructed some scientific understanding (explanation) of relevant phenomena and are ready to refine them. Students are asked to predict what will happen as a result of a particular new experiment. Students need to write the prediction and a justification of the prediction. After making their prediction and writing a justification, students observe the experiment directly. Most likely the outcome of the experiment will not match their prediction—they will have anomalous data. Then the students have to revise their prediction by revising the explanation on which their prediction was based, or the simplifying assumptions that they used to make it.

There are several reasons why surprising data tasks might be helpful in forming students' scientific abilities:<sup>36</sup> students learn to analyze data and revise explanations; students receive almost instant feedback about their prediction when they observe the experiment. Students learn to differentiate between a description and an explanation,<sup>37</sup> which is often confusing for the students.<sup>38</sup> These tasks also help develop an evaluation ability, as students have an opportunity to reconsider their reasoning after observing an experiment.

One can use them during instruction to build more sophisticated models of phenomena and, at the end of learning a particular topic, to assess students' understanding of assumptions that they often make unconsciously. An example of a surprising data task that engages students in revising a model of constant electric resistance is provided below. It can be

used in a laboratory. Students perform the activity after they learn how to build simple circuits, understand what current and voltage are, how to measure them, and that the voltage-versus-current ratio for commercial resistors is constant.

#### 1. Testing experiment

Voltage-versus-current ratio for a light bulb.

*Equipment.* Constant voltage source, a light bulb, a commercial resistor, ammeter, voltmeter, connecting wires, and a switch.

Connect a light bulb in series with an ammeter (to read the current through the bulb). Connect a voltmeter across the bulb (to read the voltage across the bulb). Do not close the switch yet.

*Write the following in your lab report.*

- (1) Draw a circuit diagram.
- (2) Close the switch, record the current through the bulb and the voltage across it. Determine the voltage to current ratio for the bulb.
- (3) Next, connect the bulb in series to a resistor. Draw a circuit diagram. Do not close the switch yet. Predict the voltage- to current-ratio for the bulb after you close the switch. Explain your prediction.
- (4) Close the switch and record the values of voltage and current. Determine the ratio. Did it match the prediction?
- (5) How can you explain the discrepancy between the predicted value and the experimental value?
- (6) Devise an experiment you could perform to test your explanation (you do not have to actually perform the experiment).

In addition to the rubrics in Appendix A, the rubric in Table III helps students self-assess their ability to revise an explanation based on the results of an experiment. More tasks of this type can be found at <http://paer.rutgers.edu/pt3>.

### D. Ability to design an experimental investigation

To devise and test relationships and explanations students need to develop experimental abilities. For pedagogical purposes we have classified experimental investigations that students perform in introductory courses into three broad categories:<sup>39</sup> observational experiments, testing experiments, and application experiments.

When conducting an observational experiment, a student focuses on investigating a physical phenomenon without

TABLE IV. Subabilities involved in designing three different types of experimental investigation.

| Subabilities involved in the ability to design and conduct an observational experiment         | Subabilities involved in the ability to design and conduct a testing experiment         | Subabilities involved in the ability to design and conduct an application experiment |
|--|---|--|
| Identifying the phenomenon to be investigated  | Identifying the relationship or explanation to be tested                                | Identifying the problem to be solved   |
| Designing a reliable experiment that investigates the phenomenon                               | Designing a reliable experiment that tests the prediction using the available equipment | Designing an experiment that solves the problem                                      |
| Deciding what is to be measured, and identifying independent and dependent variables           | Deciding what is to be measured, and identifying independent and dependent variables    | Deciding what is to be measured  |
| Using available equipment to make measurements   | Using available equipment to make measurements  | Using available equipment to make measurements                                       |
| Describing what is observed, both in words and by means of a picture of the experimental setup | Deciding whether to confirm the prediction based on the results of the experiment       | Making a judgment about the results of the experiment                                |
| Describing a pattern or devising an explanation  | Making a reasonable judgment about the relationship or explanation                      | Evaluating the results by means of an independent method                             |
| Identifying shortcomings in an experimental design and suggest specific improvements           | Identifying shortcomings in an experimental design and suggest specific improvements    | Identifying shortcomings in an experimental design and suggest specific improvements |

having expectations of its outcomes. When conducting a testing experiment, a student has an expectation of its outcome based on concepts constructed from prior experiences. In an application experiment, a student uses established concepts or relationships to address practical problems. In the process of scientific research the same experiment can fall into more than one of these categories.

What abilities do students need when designing these investigations? We have identified the following steps that students need to take to design, execute, and make sense out of a particular experimental investigation. We assigned a subability for each step and wrote corresponding descriptors in the rubrics. The results of these discussions are presented in Table IV.

For each of the identified subabilities, we devised a rubric item that describes different levels of proficiency. For example, for the subability “using available equipment to make measurements” a 0 level of proficiency is described as “at least one of the chosen measurements cannot be made with the listed equipment;” the level 1 is described as “all of the chosen measurements can be made but there are no details given of how it is done;” the level 2 is described as “all chosen measurements can be made but the details of how it is done are vague or incomplete;” and the level 3 “all measurements can be made and all details of how it is done are

provided.” The rubrics describe the levels of proficiency for each of the subabilities identified in Table IV.

Students use these rubrics in the laboratories. Ideally we want them to continuously refer to the rubrics while designing and performing the experiment. The rubrics guide them as to what experimental aspects they should specifically pay attention to. After they perform the experiment, they write a lab report (in the lab). During the process of writing, they use the descriptors in the rubrics to improve their report.

**E. Ability to collect and analyze data**

The abilities to collect and analyze data are independent of the type of experiment that is being performed and hence, have been placed in a different category. We identified subabilities that students need for successful data collection and analysis and devised rubrics for each subability. (The simplified list below is appropriate for students—scientists do this at a much more sophisticated level.) (i) Ability to identify sources of experimental uncertainty. (ii) Ability to evaluate how experimental uncertainties might affect data. (iii) Ability to minimize experimental uncertainty. (iv) Ability to record and represent data in a meaningful way. (v) Ability to analyze data appropriately. The rubric for each subability has descriptors indicating what needs to be done for a satisfac-

TABLE V. A scoring rubric to assess a student’s presentation of the data.

| Subability   | 0 (Missing)                                 | 1 (Inadequate)                                      | 2 (Needs improvement)  | 3 (Adequate)  |
|--|---|---|--|---|
| Is able to record and represent data in a meaningful way | Data are either missing or incomprehensible | Some important data are missing or incomprehensible | All important data are present but require some effort to comprehend | All important data are present, organized, and recorded clearly |

tory achievement. An example of a rubric for the subability to record and represent data in a meaningful way is shown in Table V.

Students develop these subabilities in labs. As we discussed above, the lab write-ups we provide guide them through the process by focusing their attention on the subabilities outlined in the rubrics. Below, we show an example of an application experiment for use in labs. This lab activity helps students develop the abilities to design an application experiment to solve a practical problem, and to collect and analyze data.

### ***1. Application experiment: Coefficient of friction between a shoe and a floor tile***

A floor tile company needs to decide whether their new floor tiles meet minimum safety standards. They don't want people to slip on their tiles and sue the company! Design two independent experiments to determine the maximum coefficient of static friction between your shoe and the sample of floor tile provided. Equipment: Spring scale, ruler, protractor, floor tile, tape, string, clips. Include in your report the following for each independent experiment.

- (a) Draw a sketch of your experimental design.
- (b) Write a brief outline of the procedure you will use.
- (c) Decide what assumptions about the objects, interactions, and processes you need to make to solve the problem. How might these assumptions affect the result? Make sure you only consider relevant assumptions.
- (d) Draw a free-body diagram for the shoe for the situation. (Recall your assumptions.) Include an appropriate set of coordinate axes. Use the free-body diagram to devise the mathematical procedure to solve the problem.
- (e) What are possible sources of experimental uncertainty? Which instrument gives you the highest uncertainty? How would it affect the data? How could you minimize it?
- (f) Perform the experiment and record your observations in an appropriate format. Make sure you take steps to minimize uncertainties. What is the outcome of the experiment?
- (g) When finished with both experiments, compare the two values you obtained for the coefficient of static friction. Decide, using assumptions and uncertainties, if these values are different or not. If they are different, what are possible reasons?
- (8) List shortcomings, if any, in the experiments. Suggest specific improvements.

### **F. Ability to evaluate experimental predictions and outcomes, conceptual claims, problem solutions, and models**

We define an evaluation as making judgments about information based on specific standards and criteria.<sup>40</sup> More specifically, a given particular is judged by determining whether it satisfies a criterion well enough to pass a certain standard. Scientists constantly use evaluations to assess their own work and the work of others when conducting their own research, serving as referees for peer-reviewed journals, or serving on grant-review committees.

The evaluation is a crucial ability for our students. During a physics course, students are expected to identify, correct, and learn from their mistakes with the help of an instructor. This aid may come in many forms, such as when an instructor provides problem solutions to a class, or tutoring to an individual student. However, in each case the student relies upon an instructor (or sometimes a textbook) in order to determine whether, and how, their work is mistaken. Since the students are not given any other means with which to evaluate their work, the students come to see an evaluation by external authorities as the only way for them to identify and learn from their mistakes.

There are several sets of criteria and strategies that are commonly used by practicing physicists. Each of these strategies relies upon hypothetico-deductive reasoning,<sup>34</sup> whereby the information is used to create a hypothesis which is then tested. The logical sequence for this testing can be characterized as: *If (general hypothesis) and (auxiliary assumptions) then (expected result) and/but (compare actual result to expected result), therefore (conclusion)*. For example, when a student derives an equation and needs to evaluate it with dimensional analysis, the logical sequence is as follows:

*If* the equation is physically self-consistent,

*And* I correctly remember the units for each quantity in the equation,

*Then* I expect the units for each term in the equation to be identical,

*And/But* the units for each term are/are not identical,

*Therefore* the equation is/is not physically self-consistent.

The types of subabilities that students need to develop to be successful in evaluations are numerous. Some of them are (i) ability to conduct a unit analysis to test the self-consistency of an equation; (ii) ability to analyze a relevant limiting and/or special case for a given model, equation, claim; (iii) ability to identify the assumptions a model, equation, or claim relies upon; (iv) ability to make a judgment about the validity of assumptions; (v) ability to use a unit analysis to correct an equation which is not self-consistent; (vi) ability to use a special-case analysis to correct a model, equation, or claim; (vii) ability to judge whether an experimental result fails to match a prediction; (viii) ability to evaluate the results of an experiment by means of an independent method.

Evaluation subabilities are integral components of multiple representation abilities, design abilities, etc. Examples of evaluation subabilities rubrics are given in Table VI.

To help students learn evaluation strategies, we have developed two categories of tasks. One category consists of *supervisory evaluation tasks*, wherein students act like a supervisor by evaluating (and, if necessary, correcting) someone else's work (usually the work of an imaginary friend). The other category consists of *integrated evaluation tasks*, which ask the students to evaluate, and if necessary to correct, their own work. For both categories of task, the evaluated work may be a problem solution, experiment design, experiment report, conceptual claim, or a proposed model. Supervisory evaluation tasks are meant to help the students learn the goals, criteria, and method of use for each evaluation strategy, while integrated evaluation tasks encourage

TABLE VI. Scoring rubrics to assess a student's evaluation abilities

| Scientific ability  | 0 (Missing)   | 1 (Inadequate)  | 2 (Needs improvement)  | 3 (Adequate)   |
|---|---|---|--|--|
| Is able to evaluate the results by the means of an independent method             | No attempt is made to evaluate the result by means of 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 of both methods.              | A second independent method is used to evaluate the results. Some discussion about the differences in the results is present but there is little or no discussion of the possible reasons for the difference | A second independent method is used to evaluate the results. The discrepancy between the results in the two methods and possible reasons are discussed. A percentage difference is calculated. |
| Is able to analyze a relevant special case for a given model, equation, or claim. | No meaningful attempt is made to analyze a relevant special case.           | An attempt is made to analyze a special case, but the identified special case is not relevant. OR major steps are missing from the analysis (e.g., no conclusion is made) | An attempt is made to analyze a relevant special case, but the student's analysis is flawed. OR the student's judgment is inconsistent with their analysis.  | A relevant special case is correctly analyzed and a proper judgment is made.   |

students to incorporate evaluation into their learning behavior. During a semester we tend to use mostly supervisory tasks for the first few weeks so that the students can get acquainted with each strategy, and then transition to integrated tasks so that they gain experience at using the strategies to evaluate and correct their own work.

Below are two example tasks. Each of these tasks features the same physical scenario and question. The first task exemplifies the format of a supervisory task, and is structured to help the students work through each step in the evaluation. The second task is in the format of an integrated task. In general, any of our tasks may be framed either as supervisory or integrated tasks.

*Supervisory task:* You have been given a problem, which says: "As a certain type of green bean ripens, it builds up gas inside until the bean pod explodes from the pressure and shoots its seeds outwards. Let us assume one particular seed starts off at ground level, and is shot out at an angle of  $30^\circ$  above the ground at a speed of 12.0 m/s. What is the maximum *height* the seed reaches above the ground?" Your friend Scooter comes up with the following solution.

First, he solves for  $v_{0y}$  (the  $y$  component of the initial velocity),

$$v_{0y} = v_0 \cos(\theta) = (12.0 \text{ m/s})\cos(30^\circ) = 10.4 \text{ m/s}.$$

Then he uses this to solve for the maximum height:

$$v_y^2 = v_{0y}^2 + 2a(\Delta y),$$

$$(0 \text{ m/s})^2 = (10.4 \text{ m/s})^2 + 2(-9.8 \text{ m/s}^2)(\Delta y),$$

$$\Delta y = 5.5 \text{ m}.$$

Do a special-case analysis of Scooter's solution:

(a) Choose a special-case for a situation where you know what the answer should be, conceptually.

(b) For this special-case, state what you think the answer should be, and explain your reasoning.

(c) Next, recalculate your friend's answer for your special case.

(d) Compare your conceptual expectation with the result from part (3).

(e) Make a conclusion about the validity of your friend's solution based on this comparison.

*Integrated task.* As a certain type of green bean ripens, it builds up gas inside until the bean pod explodes from the pressure and shoots its seeds outwards. Let us assume one particular seed starts off at ground level, and is shot out at an angle of  $30^\circ$  above the ground at a speed of 12.0 m/s.

(a) What is the maximum *height* the seed reaches above the ground?

(b) Do a special-case analysis of your work in part (a).

(c) If your work does not pass the special-case analysis, describe how you should change your solution in order to pass the analysis. If your work did pass the special-case analysis, describe how you benefited from doing the analysis anyway.

Evaluation tasks are usually employed as part of recitations and homework. During recitations, students have a chance to try using each strategy while getting real-time feedback from their peers and instructor. Moreover, recitations provide a setting where the students can learn how to associate topic-specific knowledge with the general evaluation strategies. The homework then provides an opportunity for students to gain further practice at using each strategy.

### G. Ability to communicate

An important ability in the work of scientists is their oral and written communication, an ability that can be fostered in a physics course. For example, the quality of a lab report can be judged for its completeness and clarity. A communication

TABLE VII. A scoring rubric to assess a student's communication ability.

| Scientific ability   | 0 (Missing)   | 1 (Inadequate)   | 2 (Needs improvement)   | 3 (Adequate)   |
|--|---|--|---|--|
| Is able to communicate the details of an experimental procedure clearly and completely | Diagrams are missing and/or experimental procedure is missing or extremely vague. | Diagrams are present but unclear and/or the experimental procedure is present but important details are missing. | Diagrams and/or experimental procedure are present but with minor omissions or vague details. | Diagrams and/or experimental procedure are clear and complete. |

ability rubric can help students know what is expected in communications in the scientific world. An example of a communication ability rubric is provided below in Table VII.

#### IV. DO STUDENTS DEVELOP SCIENTIFIC ABILITIES?

In this section we will describe briefly four research projects whose goals were to investigate whether students who learn physics in courses which focus on the development of specific abilities actually acquire them and use them while solving problems, designing experiments, evaluating their work, etc. We also investigated students' attitudes towards the development of some of these abilities.

##### A. The study of multiple-representation abilities

This two-year descriptive study was conducted in a large-enrollment (about 500 students) algebra-based physics course for science majors. One of the course goals was to help the students use multiple representations for the analysis of phenomena and for problem solving. In lectures the instructor discussed with the students how to represent the same process in multiple ways, and how to use one type of representation to help construct another, and how to apply the representations for problem solving. Students worked individually and in small groups on activities that required them to represent problem situations in different ways without actually solving for a particular quantity. In recitations students worked on similar activities; they were also part of homework assignments and appeared as multiple-choice problems on the exams. When students solved traditional problems, they were encouraged to use a problem-solving strategy in which physical representations (such as free-body diagrams, energy bar charts, ray diagrams, etc.) were used to construct mathematical equations. Homework solutions for traditional problems used a multiple-representation strategy.

For the study we used a free-body diagram as an example of a physical representation. Students learned to construct free-body diagrams according to the advice provided in the rubric described earlier. Students learned to convert a diagram into Newton's second law in component forms. Occasionally, they were given Newton's second law in a component form as applied to an unknown situation and asked to construct a consistent free-body diagram and to describe in words a consistent physical situation.

The goals of the study were to (a) see if students who were explicitly taught how to draw FBDs and how to use them to solve problems actually used them to help solve multiple-choice problems when no credit was given for

drawing the diagrams, and (b) see if the use of a FBD correlated with success in solving Newton's second law problems. We collected data from several multiple-choice exams (the quantitative study) and interviewed some students (the qualitative study). In the first year, we chose five problems from four exams; in the second year, seven problems from four exams. The problems were chosen if they involved forces and were difficult enough to merit the construction of a diagram. In our first year, we followed 125 randomly chosen students and 120 in the second year.

We examined FBDs that students drew on the exam sheets near the problem statement. First we counted the number of students who drew diagrams on the problem sheets. The exams were multiple choice and students were only given credit for a correct answer. Thus, if a diagram appeared on the problem sheet next to some kind of mathematical solution, we considered that a student drew it to help solve the problem. We found that on average about 58% of our students drew a FBD for each of the chosen exam problems even when they knew they would not receive any credit for doing so. For traditionally taught students, this number is around 20%.<sup>41</sup>

To see if the FBD construction correlated with success in problem solving, we scored student diagrams using the rubric. Our coding scheme followed the rubric descriptors (missing, inadequate, needs improvement, and adequate) shown earlier. We found that for over 12 problems on 4 exams 85% of students who had a correct free-body diagram ("adequate," according to the rubric—Table II) had a correct answer. 71% of those who had an incomplete diagram ("needs improvement") had a correct answer. 38% of those who had an incorrect diagram (inadequate) had a correct answer and 49% of those who did not draw a diagram (missing) had a correct answer. The average percentage of students who solved these problems correctly was 60%. A two-sample *t* test for independent samples indicated that those who had an adequate diagram significantly outperformed the class average, and those who had inadequate diagrams were significantly below the class average. There were some small variations between the categories per individual question. However, there was never a case when the students who did not draw a diagram were more successful than those who drew a correct one.

We followed up with interviews (six students) during which students had to solve several problems using a think-aloud protocol. All students started solving a problem by constructing a sketch of the situation described in the problem statement and five out of six drew a free-body diagram after the sketch. Two most successful students then went

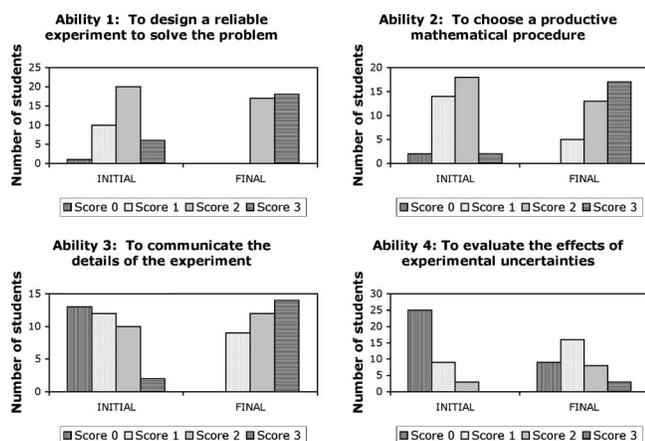


FIG. 2. Comparison abilities during the third and the tenth week of the semester.

back and forth between the sketch and their free-body diagram to understand the problem better. They used the diagrams to help create a Newton's second law equation in a component form, which they then used to solve the problem. These students determined the direction of the acceleration and constructed a diagram that was consistent in terms of force magnitudes with the acceleration direction. After completing their numerical solution, they went back to the free-body diagrams and used them to evaluate the final result. None of the students who were unsuccessful in problem solving during the interview used the free body diagram to construct the mathematical representation and none used it to evaluate the result. More details of this study can be found in Ref. 42.

### B. The study of experimentation abilities

The study was conducted in a large-enrollment introductory laboratory course (500 students) for science majors (pre-med, prevet, biology, exercise science, and environmental science). Although the laboratory course was separate from a lecture and/or recitations course, most of the students were enrolled in both. The development of scientific abilities was considered an important goal in both courses. In the lab course there were 20 lab sections, each with about 25 students and 9 teaching assistants. Students performed one three-hour lab per week for 10 weeks. In each lab, at least one experiment was a design task (similar to the example in Sec. III E). The experiments had guidelines that focused on different scientific abilities that we had identified.

To measure the development of students' acquisition of scientific abilities, we scored their lab reports each week based on the scientific abilities rubrics. We focused on the ability to design a reliable experiment to solve a problem, to choose a productive mathematical procedure, to communicate details of the experiment, and to evaluate the effects of experimental uncertainty. Our sample consisted of 35 randomly chosen students who were distributed among 4 lab sections. In Fig. 2, we show a histogram of the scores that students' reports received on four abilities during the third and the tenth week of the semester.

We found that the students' abilities to design an experiment, to devise a mathematical procedure to solve an experimental problem, and to communicate the details of the procedure performance improved.<sup>43</sup> The changes in the above abilities were statistically significant. However, the changes in students' ability to evaluate experimental uncertainties were not significant. We later found that another ability on which students do not improve is the evaluation of the effects of theoretical assumptions.<sup>44</sup> It could be that guidelines for the students in the lab, asking them to evaluate the effects of assumptions and uncertainties, are not sufficient for them to actually do it. Students probably need additional exercises helping them master these abilities.<sup>45,46</sup> We are currently working on the development and pilot testing of these exercises.

### C. The study of transfer of scientific abilities

The goal of this study was to investigate whether students can transfer some scientific abilities to a new context. We use the term transfer here to mean an ability to apply something that one learns in one context in a different context or with a different content.

This study was conducted in a 190-student algebra-based physics course with two 55-minute lectures, one 80-minute recitation, and one three-hour lab. Each lab usually contained two design tasks in which students had to either test a proposed hypothesis (similar to the Testing experiment in Sec. III B) or experimentally solve a practical problem (similar to the Application experiment in Sec. III E). The lab handouts provided to the students resemble the example shown after the next paragraph.

In the first lab of the course, students were to design an experiment to test a proposed hypothesis (the full experiment write-up and sample student responses are given in Appendix A and Tables X and XI). They were encouraged to design an experiment to reject the hypothesis, as a supportive outcome meant only that the hypothesis was not ruled out—it did not prove the hypothesis. Students had guidelines provided in the lab handout and rubrics to help them design an experiment. They were to use hypothetico-deductive reasoning to test the hypothesis. The logic of hypothetico-deductive reasoning that we communicated to the students was as follows:

*If* the hypothesis being tested is correct.

*And* I perform the following testing experiment.

*Then* I predict the outcome of the experiment based on the hypothesis.

*And/But* the outcome did/did not match the prediction.

*Therefore* the hypothesis is not/is disproved.

A question on the final exam was similar to the above experiment: "Describe an experiment that you could design to test the proposed hypothesis that an object always moves in the direction of the unbalanced force exerted on it by other objects." The final exam was three months after the above experiment was performed in the lab. However, on the exam, students had no guidelines like parts (a)–(i) in Appendix A, and no rubrics. Also, students worked individually on the exam. According to a scheme by Barnett and Ceci,<sup>47</sup> the transfer we examined can be classified as near in terms of the

TABLE VIII. Students responses on the final exam ( $N=181$ ).

| Students' performance on final exam   | Transfer rate |
|---|---------------|
| Drew pictures   | 94%           |
| Drew physical representations (motion diagrams, free-body diagrams)         | 45%           |
| Designed experiment to try to reject the proposed rule                      | 58%           |
| Designed experiment that supports rule                                      | 41%           |
| Prediction was based on rule to be tested AND effect of assumptions         | 16%           |
| Prediction was based on rule to be tested but not on effects of assumptions | 64%           |
| Prediction was not based on rule but on previous knowledge of Newton's laws | 32%           |
| Considered assumptions  | 47%           |

knowledge domain, but far in terms of physical context (exam hall instead of a physics lab), functional context (writing an answer to a question versus designing and performing an experiment), social context (individual versus group), and modality (exam versus a lab).

The individual student's exam question was scored using the rubrics. Table VIII indicates the percentage of students whose exam answer received a score of 3 on the relevant items in the rubrics. We call this percentage the rate of transfer. The rate of transfer means that a certain percentage of the students are successful in a certain ability. Typically the rate of transfer is measured by the percentage of subjects who can solve a problem similar to a problem that they were taught to solve. These results are encouraging as this rate of transfer for some abilities was much higher than the typical 20% transfer rate reported in other studies (this is transfer of a skill without a hint).<sup>48</sup>

We believe that it is unlikely that students remembered the details of the laboratory work on this question performed three months earlier. However, they have been using hypothetico-deductive reasoning during the semester, and this practice could have contributed to the positive result. Was this problem too easy and hence a false indicator of transfer? To address this issue, we compared students' performance on this exam question and their overall performance on the exam. The average score on this special question was 60% (standard deviation of 21%) as opposed to the average score of 69% (standard deviation 15.5%) on the exam as a whole—it seems that the special question was not too easy. Thus we think that students did indeed transfer some experimentation-related scientific abilities that they acquired in the labs using the rubrics.

#### D. The study of evaluation abilities

A comparison-group study was conducted in two large-enrollment courses for science majors. Both courses are year-long courses but the experimental course has generally lower achieving students and is taught on a different university campus. During the year when the experiment was con-

ducted the two courses were run in parallel, with nearly identical lectures, recitations, homework, and labs. The only significant difference was that in the experimental course several evaluation tasks were included in recitation assignments and homework while the comparison course had no evaluation tasks in their coursework.

The study had two goals. One goal was to find out whether the use of our evaluation tasks did indeed help students to acquire evaluation abilities. The other goal was to find whether an improvement in students' evaluation abilities benefited their understanding of the subject matter.

We included an evaluation task on each of the six exams throughout the year for both courses. We used the rubrics to assign scores of 0–3 (0, missing; 1, inadequate; 2, needs improvement; 3, adequate) to student work related to each strategy. We found that in the experimental group students mastered the unit analysis strategy very quickly, scoring an average of 2.5 with a standard deviation of 1.05 on the first exam, and maintaining that average throughout the year. In contrast, the much more complicated strategy of special-case analysis took most of a semester to be learned. On the first exam, the average score was 1.17 with a standard deviation of 1.04. By the end of the first semester, though, the student average reached 2.30, with a standard deviation of 0.62, and remained roughly there for the entire second semester.

To address our second goal, we analyzed exam data from the two courses. The exams for the experimental and control groups shared several of the same multiple-choice problems. Of these shared problems, some were on topics on which only the experimental group students had evaluation tasks (such as Newton's Second Law), whereas the comparison group had additional standard problems on those topics instead of the evaluation tasks in their homework and recitations. The shared multiple-choice exam problems on such topics will be called *E* problems. (Though called *E* problems, these problems did not have an evaluation component, they were traditional physics problems. The letter *E* relates to the fact that students in the experimental course had received evaluation tasks related to the content of those problems instead of traditional problems.) The remainder of the shared exam problems were on topics on which no one had evaluation tasks (such as momentum). We will call these *NE* problems.

By comparing the relative performance of each class on *E* and *NE* problems, we could test whether the use of our evaluation tasks benefited the students' problem solving performance. What we found was that the experimental group students significantly outperformed control students on *E* problems, while the control group students significantly outperformed the experimental group students on *NE* problems.

We did, in fact, expect the control group students to do better on *NE* problems since these students have stronger math and science backgrounds. In light of this population difference, our result that the experimental group students did better on *E* problems is especially remarkable. It indicates that the use of evaluation tasks significantly benefited students' problem solving for those topics. In particular, we found that a mastery of special-case analysis was the primary cause behind this boost in relative performance. This is reasonable since a special-case analysis is one effective way for

students to test and refine their conceptual understanding, and to coherently organize different models into a hierarchical system.

## V. SUMMARY

This paper described the development of tasks and assessment rubrics that help students acquire some of the abilities useful in science and engineering. Development of these abilities is considered important in science courses at K–16 levels.<sup>14–19</sup> There are summative assessment questions developed for national and international evaluations of student learning of some scientific abilities but we are unaware of any systematic efforts to build a coherent and systematic library of formative assessment tasks and rubrics to help students develop these abilities.

We have developed an approach to learning introductory college physics in which the acquisition of various scientific abilities is one of the main goals. We list below the general process abilities that we have chosen to include in the learning system with the goal of helping students develop these abilities while learning physics. (i) Learn to represent physical processes in multiple ways. (ii) Learn to devise and test a qualitative explanation or quantitative relationship. (iii) Learn to modify a qualitative explanation or quantitative relationship. (iv) Learn to design an experimental investigation. (v) Learn to record, represent, and analyze data. (vi) Learn to evaluate experimental predictions and outcomes, conceptual claims, problem solutions, and models. (viii) Learn to communicate. Each ability includes subabilities that are needed for proficiency in that ability.

In order to help students acquire these abilities, we have developed a large number of activities used formatively during instruction. In addition, we have developed rubrics, which indicate what is needed for proficiency relative to the different subabilities. The rubrics can also be used by instructors to provide formative feedback to the students or for summative evaluation of the students and the learning system. The rubrics can also be used by students for self-evaluation as they perform the activities. Students need to be able to revise their work after they evaluated it using the rubrics. Here the instructor's help is essential as the rubrics themselves do not provide content-related feedback. The rubrics also can be used for research purposes to monitor students progress and to compare students from different courses. They can add to our library of assessment instruments that allow the PER community to evaluate learning. So far most of the PER-developed instruments assess conceptual understanding and graphing skills.

Our summative use of the rubrics to study if students in introductory physics courses (primarily the introductory physics course for biology majors) are getting better in using the abilities is positive for the most part. In the multiple-representation study we found that a relatively large number of the students (about 60% compared to 20% in traditionally taught courses) used free-body diagrams in their problem solving. Correct use of the diagrams was significantly correlated with success in the problem solution. In the second

study concerning student-experimental abilities, we found that their abilities to design an experiment, to devise a mathematical procedure to solve an experimental problem, and to communicate the details of the procedure performance improved significantly. However, their ability to evaluate experimental uncertainties did not change significantly. In the third study, we found that students could transfer some abilities learned in one context to new contexts. In the fourth study, we found that there was a significant increase in student problem solving performance in conceptual areas in which they had done limit case evaluation activities. It seems that this type of evaluation enhances student understanding and problem solving. Finally, we have evaluated student traditional problems solving performance in this learning system that emphasizes the development of science process abilities. These students do significantly better on multiple choice problems on final tests than their peers taught via traditional methods.<sup>49</sup>

In summary, it seems that it is possible to help students in introductory physics courses start acquiring some of the science process abilities that are needed for work in the 21st century workplace. The learning system also enhances student performance in terms of traditional measures. Developed rubrics can be used for research purposes to collect data about student progress and for comparison of student learning in different courses.

## ACKNOWLEDGMENTS

The authors wish to thank Michael Lawrence, Marina Milner-Bolotin, Hector Lopez, and Juliana Timofeeva for their help in the development of rubrics and formative assessment tasks. We wish to thank Richard Fiorillo and Gabriel Alba for their help in implementing the laboratories. We also thank the National Science Foundation (Grant No. DUE-0241078) for providing us with funding for the project.

## APPENDIX A: EXAMPLES OF STUDENT WORK AND ITS ASSESSMENT WITH THE RUBRICS

Design an experiment to test the following proposed hypothesis: An object always moves in the direction of the unbalanced force exerted on it by other objects.

- (a) State what hypothesis you will test in your experiment.
- (b) Brainstorm the task. Make a list of possible experiments. Decide what experiments are best.
- (c) Draw a labeled sketch of your chosen experiment.
- (d) Write a brief description of your procedure.
- (e) Construct a free-body diagram for the object.
- (f) List assumptions you make. How could they affect the prediction?
- (g) Make a prediction about the outcome of the experiment based on the hypothesis you are testing.
- (h) Perform the experiment. Record the outcome.
- (i) Make a judgment about the hypothesis based on your prediction and the experimental outcome.

We also discuss student responses and provide rubric scores for the subabilities shown in Table IX–XI.

TABLE IX. Sub-abilities rubrics for designing a testing experiment and making a prediction of the outcome of the experiment.

|                    |   | Ability to design and conduct a testing experiment   |  |  |   |
|--------------------|---|--|--|--|---|
| Scientific ability | 0 (Missing)   | 1 (Inadequate)   | 2 (Needs some improvement)   | 3 (Adequate)   |   |
| 1                  | Is able to identify the relationship or explanation to be tested.                                 | No mention is made of a relationship or explanation.   | An attempt is made to identify the relationship or explanation to be tested but is described in a confusing manner.  | The relationship or explanation to be tested is described but there are minor omissions or vague details.  | The relationship or explanation is clearly stated.  |
| 2                  | Is able to design a reliable experiment that tests the relationship or explanation.               | The experiment does not test the relationship or explanation.                                    | The experiment tests the relationship or explanation, but due to the nature of the design it is likely the data will lead to an incorrect judgment.  | The experiment tests the relationship or explanation, but due to the nature of the design there is a moderate chance the data will lead to an inconclusive judgment. | The experiment tests the relationship or explanation and had a high likelihood of producing data that will lead to a conclusive judgment. |
| 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 chosen measurements can be made, but the details about how it is done are vague or incomplete.   | All chosen measurements can be made and all details about how it is done are clearly provided.  |
| 4                  | Is able to decide whether or not to confirm the prediction based on the results of the experiment | No decision is made to confirm or disconfirm the prediction.                                     | A decision is made but it is not strongly based on the results of the experiment.  | A decision is made based on the results of the experiment, but the reasoning is flawed.  | A correct decision is made and is based on the results of the experiment.   |
| 5                  | Is able to make a reasonable judgment about the relationship or explanation.                      | No judgment is made about the relationship or explanation, or is not based on the results.       | A judgment is made but it is based only on the degree of agreement between the results and the prediction.   | A judgment is made based on the reliability of the experiment and the degree of agreement between the results and the prediction, but the reasoning is flawed.       | A reasonable judgement is made based on the reliability of the experiment and the degree of agreement between the results and prediction. |
|                    |   | Ability to construct, modify, and apply relationships or explanations                            |  |  |   |
| 7                  | Is able to make a reasonable prediction based on a relationship or explanation.                   | No attempt to make a prediction is made. The experiment is not treated as at testing experiment. | A prediction is made but it does not follow from the relationship or explanation being tested, or it ignores or contradicts some of the assumptions inherent in the relationship or explanation. | A prediction is made that follows from the relationship or explanation, but it does not incorporate the assumptions.   | A prediction is made that follows from the relationship or explanation and incorporates the assumptions.                                  |
| 8                  | Is able to identify the assumptions made in making the prediction.                                | No attempt is made to identify any assumptions.  | An attempt is made to identify assumptions, but most are missing, described vaguely, or incorrect.   | Most assumptions are correctly identified.   | All assumptions are correctly identified.   |

TABLE X. Student work: Student A.

| Transcript of student’s lab report  | Researcher’s comments based on scientific abilities rubrics  | Rubric item | Score |
|---|--|-------------|-------|
| In my experiment, I wish to test the hypothesis that an object always moves in the direction of the net force exerted on it by other objects.   | States the hypothesis to be tested experimentally.   | 1           | 3     |
| I will get the bowling ball moving to the left in a straight line. I will then hit it towards the right with a mallet gently.   | Description of the experiment: The proposed experiment tests the hypothesis and has a high likelihood of producing data that will lead to a conclusive judgment. | 2           | 3     |
| Student draws figures and free-body diagrams here.  | Multiple representations.  |             |       |
| I assume there is no friction between the ball and the floor. Then the only force on the ball in the $x$ direction is the force of the mallet on the ball to the right. Net force is to the right.  | Discusses an assumption in the procedure. However, the main assumption—that the floor is not tilted is not addressed. The friction assumption is irrelevant.     | 8           | 1     |
| Prediction: According to the hypothesis the ball should immediately start moving to the right when the mallet hits it no matter how hard we hit it.   | Prediction is based on the hypothesis.   | 7           | 3     |
| Outcome: When I tapped the ball gently the ball first slowed down, and only later moved to the right. When I tapped hard, the ball still slowed down a little but then moved to the right. It did not instantly reverse directions as the hypothesis predicted. | Note that the student has designed and performed a series of experiments where the outcome is different from the prediction, i.e., she tries to reject the rule. | 4           | 3     |
| Based on my prediction and the experimental outcome, the hypothesis is not supported.   |  | 5           | 3     |

TABLE XI. Student work: Student B.

| Transcript of student’s lab report  | Researcher’s comments based on scientific abilities rubrics   | Rubric item | Score |
|---|---|-------------|-------|
| Hypothesis we will test: An object always moves in the direction of the net force exerted on it.  | States the hypothesis to be tested experimentally.  | 1           | 3     |
| Possible experiments:<br>1. Hit a bowling ball with a mallet.<br>2. Push a dynamics cart along a dynamics track.<br>3. Push (gently) your lab partner forward.<br>The first choice would be the best because it would most clearly exhibit the hypothesis we are trying to prove. | Brainstorming different experiments. Note that the student is trying to support hypothesis.   |             |       |
| Chosen experiment: We will hit the bowling ball (which is sitting at rest) in a forward direction, using the mallet.  | This experiment will confirm the rule instead of trying to reject it. Due to the nature of the design it is likely the data will lead to an incorrect judgment. | 2           | 1     |
| Prediction: The ball will move in the forward direction, the same direction as the net force exerted on it by the mallet, and in no other direction besides that one.   | Prediction is based on the rule under test, but assumptions are not considered.   | 7           | 2     |
|   | Assumptions are not described.  | 8           | 0     |
| Outcome: The ball moved forward in the exact direction of the net force on it (force of the mallet). Yes—the prediction was confirmed.  | Note that the student has designed and performed an experiment where the outcome is supported by the rule.  | 4           | 3     |
| Based on our prediction and successful outcome, we can say that hypothesis is supported by experimental evidence.   | The judgment is based only on the degree of agreement between the results and the prediction.   | 5           | 1     |

- \*Also at The Rutgers Graduate School of Education, 10 Seminary Place, New Jersey 08901, USA.
- <sup>1</sup>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, New York, 1997), pp. 213–224.
  - <sup>2</sup>*What Work Requires of Schools: A SCANS Report for America 2000* (U. S. Department of Labor, Secretary's Commission on Achieving Necessary Skills, Washington D.C., 1991).
  - <sup>3</sup>J. D. Bransford, A. L. Brown, and R. R. Cocking, *How People Learn: Brain, Mind, Experience, and School* (National Academy Press, Washington, D.C., 1999).
  - <sup>4</sup>A. Carneval, L. Gainer, and A. Meltzer, *Workplace Basics: The Essential Skills Employers Want* (Jossey-Bass, San Francisco, 1990).
  - <sup>5</sup>See the ABET home page at URL <http://www.abet.org/forms.shtml>
  - <sup>6</sup>National Research Council, *National Science Education Standards* (National Academy Press, Washington, D.C. 1996).
  - <sup>7</sup>L. C. McDermott & the Physics Education Group at the University of Washington, *Physics by Inquiry* (Wiley, New York, 1996).
  - <sup>8</sup>E. F. Redish, *Teaching Physics with the Physics Suite* (Wiley, New York, 2003).
  - <sup>9</sup>D. R. Sokoloff and R. K. Thornton, *Interactive Lecture Demonstrations* (Wiley, New York, 2004).
  - <sup>10</sup>D. R. Sokoloff, R. K. Thornton, and P. W. Laws, *RealTime Physics* (Wiley, New York, 2004).
  - <sup>11</sup>P. W. Laws, *Workshop Physics Activity Guide*, 2nd ed. (Wiley, New York, 2004).
  - <sup>12</sup>L. C. McDermott and P. S. Shaffer, *Tutorials in Introductory Physics* (Prentice Hall, New York, 1998).
  - <sup>13</sup>See the SCALE-UP project webpage at URL <http://www.ncsu.edu/per/scaleup.html>
  - <sup>14</sup>H. Kruglak, Am. J. Phys. **20**, 136 (1952).
  - <sup>15</sup>H. Kruglak, Am. J. Phys. **22**, 442 (1954).
  - <sup>16</sup>H. Kruglak, Am. J. Phys. **22**, 452 (1954).
  - <sup>17</sup>H. Kruglak, Am. J. Phys. **23**, 82 (1955).
  - <sup>18</sup>National Assessment Governing Board, *Science Framework for the 2005 National Assessment of Education Progress* (US Department of Education, Washington, DC, 2004), URL [http://www.nagb.org/pubs/s\\_framework\\_05/761907-ScienceFramework%rk.pdf](http://www.nagb.org/pubs/s_framework_05/761907-ScienceFramework%rk.pdf)
  - <sup>19</sup>P. Gonzales, C. Calsyn, L. Jocelyn, K. Mak, D. Kastberg, S. Arafeh, T. Williams, and W. Tsen, Statistical Analysis Report No. NCES 2001-028 2001 (unpublished).
  - <sup>20</sup>G. Salomon and D. N. Perkins, Educ. Psychol. Meas. **24**, 113 (1989).
  - <sup>21</sup>S. Holton and S. Brush, *Physics, The Human Adventure* (Rutgers University Press, New Brunswick, New Jersey, 2001).
  - <sup>22</sup>A. E. Lawson, Am. Bio. Teach. **62**, 482 (2000).
  - <sup>23</sup>A. E. Lawson, Int. J. Sci. Educ. **25**, 1387 (2003).
  - <sup>24</sup>B. S. Bloom, *Taxonomy of Educational Objectives: The Classification of Educational Goals, Handbook I Cognitive Domain* (David McKay Co, Inc., New York, 1956).
  - <sup>25</sup>P. Dressel and L. Mayer, *General Education: Exploration in Evaluation*. Final Report of the Cooperative Study of Evaluation in General Education (Council on Education, Washington, DC, 1954).
  - <sup>26</sup>C. D. Schunn and J. Anderson, in *Designing for Science: Implications from Everyday, Classroom and Professional Settings*, edited by K. Crowley, C. D. Schunn, and T. Okada (Erlbaum, Mahwah, NJ, 2001), pp. 351–392.
  - <sup>27</sup>P. Black and D. Wiliam, *Inside the Black Box: Raising Standards through Classroom Assessment* (nferNelson, London, 1998).
  - <sup>28</sup>D. R. Sadler, Instr. Sci. **18**, 119 (1989).
  - <sup>29</sup>S. M. Brookhart, *The Art and Science of Classroom Assessment: The Missing Part of Pedagogy, ASHE-ERIC Higher Education Report* (The George Washington University, Graduate School of Education and Human Development, Washington, DC, 1999) Vol. 27, No.1.
  - <sup>30</sup>R. D. Tweney, in *Designing for Science: Implications from Everyday, Classroom and Professional Settings*, edited by K. Crowley, C. D. Schunn, and T. Okada (Erlbaum, Mahwah, NJ, 2001), pp. 141–173.
  - <sup>31</sup>D. Hestenes, Am. J. Phys. **55**, 440 (1987).
  - <sup>32</sup>J. A. Kournay, *Scientific Knowledge: Basic Issues in the Philosophy of Science* (Wadsworth, Belmont, CA, 1987).
  - <sup>33</sup>E. Nagel, *The Structure of Science* (Harcourt, Brace, & World, Inc., New York, 1961).
  - <sup>34</sup>A. E. Lawson, Sci. Educ. **9**, 577 (2000).
  - <sup>35</sup>E. Etkina, A. A. Warren, and M. Gentile, Phys. Teach. **44**, 34 (2006).
  - <sup>36</sup>C. A. Chinn and W. F. Brewer, Rev. Educ. Res. **63**, 1 (1993).
  - <sup>37</sup>M. Lawrence, Ph.D. thesis, Rutgers, the State University of New Jersey, 1997.
  - <sup>38</sup>R. H. Horwood, Sci. Educ. **72**, 41 (1988).
  - <sup>39</sup>E. Etkina, A. Van Heuvelen, D. T. Brookes, and D. Mills, Phys. Teach. **40**, 351 (2002).
  - <sup>40</sup>L. W. Anderson and D. R. Krathwohl, *A Taxonomy for Learning, Teaching, & Assessing: A Revision of Bloom's Taxonomy of Educational Objectives* (Longman, New York, 2001).
  - <sup>41</sup>A. Van Heuvelen, Am. J. Phys. **59**, 891 (1991).
  - <sup>42</sup>D. Rosengrant, A. Van Heuvelen, and E. Etkina, in *Proceedings of the 2005 Physics Education Research Conference*, edited by P. Heron, J. Marx, and L. McCullough, AIP Conference Proc. No. 818 (AIP, New York, 2006), pp. 49–52.
  - <sup>43</sup>S. Murthy and E. Etkina, in *Proceedings of the 2004 Physics Education Research Conference*, AIP Conf. Proc. No. 720 (AIP, NY, 2005), pp. 133–137.
  - <sup>44</sup>S. Murthy and E. Etkina, Using rubrics as self-assessment instruments for students in introductory laboratories, AAPT National Meeting, Albuquerque, NM, January, 2005.
  - <sup>45</sup>A. Buffler, S. Allie, F. Lubben, and B. Campbell, Int. J. Sci. Educ. **23**, 1137 (2001).
  - <sup>46</sup>R. Lippmann, Ph.D. thesis, University of Maryland, 2003.
  - <sup>47</sup>S. M. Barnett and S. J. Ceci, Sankhya, Ser. A **128**, 612 (2002).
  - <sup>48</sup>M. L. Gick and K. J. Holyoak, Cogn. Sci. **12**, 306 (1980).
  - <sup>49</sup>Study reported in A. Van Heuvelen, Implementing a science process approach in large introductory physics courses, talk given at Rutgers University, November, 2004.