HELPING UNDERGRADUATES REPAIR FAULTY MENTAL MODELS IN THE STUDENT LABORATORY

Harold I. Modell, 1 Joel A. Michael, 2 Tom Adamson, 3 Jack Goldberg, 3 Barbara A. Horwitz, 3 David S. Bruce, 4 Margaret L. Hudson, 5 Shirley A. Whitescarver, 6 and Stephen Williams 7

1 National Resource for Computers in Life Science Education, Seattle, Washington 98115; 2 Department of Molecular Biophysics and Physiology, Rush Medical College, Chicago, Illinois 60612; 3 Section on Neurobiology, Physiology, and Behavior, University of California, Davis, California 95616; 4 Department of Biology, Wheaton College, Wheaton, Illinois 60187; 5 Department of Biology, Seattle University, Seattle, Washington 98122, 6 Department of Biological Science and Nursing, Lexington Community College, Lexington, Kentucky 40506; and 7 Department of Biology, Glendale Community College, Glendale, Arizona 85302

Over half of the undergraduate students entering physiology hold a misconception concerning how breathing pattern changes when minute ventilation increases. Repair of this misconception was used as a measure to compare the impact of three student laboratory protocols on learning by 696 undergraduate students at 5 institutions. Students were tested for the presence of the misconception before and after performing a laboratory activity in which they measured the effect of exercise on tidal volume and breathing frequency. The first protocol followed a traditional written “observe and record” (“cookbook”) format. In the second treatment group, a written protocol asked students to complete a prediction table before running the experiment (“predictor” protocol). Students in the third treatment group were given the written “predictor” protocol but were also required to verbalize their predictions before running the experiment (“instructor intervention” protocol). In each of the three groups, the number of students whose performance improved on the posttest was greater than the number of students who performed less well on the posttest (P < 0.001). Thus the laboratory protocols helped students correct the misconception. However, the remediation rate for students in the “instructor intervention” group was more than twice that observed for the other treatment groups (P < 0.001). The results indicate that laboratory instruction is more effective when students verbalize predictions from their mental models than when they only “discover” the outcome of the experiment.


Key words: active learning; misconceptions; respiratory physiology

An active learning environment is one in which students are encouraged to engage in the process of building and testing their own mental models from information that they are acquiring (7). One traditional format for providing students with “active learning” experiences, especially in undergraduate science courses, is the student laboratory. This setting potentially provides an ideal environment in which students can test their mental models by making and testing predictions of how the system under study
functions. This process of applying mental models to new situations has been shown to help students repair faulty mental models in a variety of disciplines (1, 2, 8). However, there is evidence that student laboratories do not always achieve this goal (11, 12).

Most student laboratories share some common features regardless of the specific discipline for which they are intended. Instructors define the “experiment” of the day and ensure that appropriate supplies and resources are available for the chosen exercise. In anatomy and physiology, the experiment can range from examining anatomic structures to gathering data from live preparations to running simulations using a computer model. Instructors usually provide a set of written instructions—a laboratory protocol—that directs students through the component activities of the exercise. The design of the laboratory protocol may be a critical factor in determining the success of the laboratory with respect to student mental model testing and repair. The study described here was intended to serve as a first step in examining the impact of the laboratory protocol on the extent to which students repair faulty mental models during student laboratories.

A faulty mental model of a natural phenomenon, often termed a misconception or an alternative conception (10, 13), leads to erroneous predictions of system behaviors involving that phenomenon. For example, over half of undergraduate students entering physiology hold the misconception that, while breathing frequency increases when minute ventilation increases (e.g., as during exercise), tidal volume either decreases or remains unchanged (4, 5). The most prevalent explanation for this belief is that breathing faster does not allow time to breathe as deeply (4, 5). In fact, when minute ventilation increases, both breathing frequency and tidal volume increase.

This study was designed to address the following two questions. Can student laboratories in which the students observe the changes in breathing pattern that occur with exercise help them repair (correct) this misconception? Is the format of the protocol that directs the students through the laboratory an important factor in helping students repair this misconception?

**METHODS**

Experiments in this study were conducted at five member institutions of the Physiology Educational Research Consortium (PERC). Two were community colleges, one was a private liberal arts college, one was a private university, and one was a state research university. Participating courses were introductory courses that serviced preprofessional students (pre-med, pre-nursing, etc.), biology majors, and non-majors. Students in each course registered for one of several laboratory sections. The experiments were conducted as part of regularly scheduled lab periods within the course. The focus activity (measuring tidal volume and frequency before and after exercise) was a routine part of the laboratory and was one of several activities that students performed during that day's lab period. Six hundred and ninety-six undergraduate students participated in the study. Each student served as his or her own control.

**Pretest.** To determine the prevalence of the misconception, each student was asked to answer the following question either at the beginning of the course or at the beginning of the respiration section of the course:

A classmate is late for the exam and runs up 5 flights of stairs to the exam room. When she arrives, you notice that her breathing frequency is increased. At the same time you notice that her depth of breathing (how much air she takes in with each breath) is

- a. greater than normal
- b. less than normal
- c. unchanged

The students also answered a follow-up question intended to provide information related to the reasoning behind their answers (4, 5). Answers to the follow-up question were solicited for a related study but did not play a role in data analysis in this study. Students were not told if their pretest predictions were correct.

Based on the answer provided for the pretest question, each student was categorized as either holding the misconception (pretest wrong) or not holding the misconception (pretest correct) at the outset of the experiment. The misconception was not specifically addressed in the lecture portion of the courses.
Laboratory exercise. During the course, students participated in a laboratory session in which they examined responses to exercise. The form of exercise differed at different institutions. At one institution, students exercised on a bicycle ergometer. At the other institutions, students ran in place, ran up several flights of stairs, performed step exercise, performed jumping jacks, or ran around the outside of a building. In all cases, students measured tidal volume and breathing frequency in a test subject before and immediately after performing the exercise. In the institution utilizing the ergometer, tidal volume and frequency data were obtained on-line during the exercise period using a pneumotachograph, integrator, and related computer data acquisition equipment. At the other institutions, data were obtained using a spirometer and stopwatch.

The key to achieving an active learning environment is to have students test the mental models that they are developing. Unfortunately, the directions that students receive in undergraduate laboratories are seldom designed to encourage this process. An examination of “traditional” student laboratory protocols from physiology, biology, chemistry, and physics reveals an interesting pattern (6). Procedures generally tell students to “do this,” “record this,” “tell me what you observed,” and, in some cases, “interpret what you observed,” or they ask, “what are the implications of what you observed?” Although students following these “cookbook” protocols are “active,” they are not necessarily involved in active learning because they may not explicitly be testing their mental models against the phenomena that they are observing.

An alternative approach is to design the laboratory protocol so that it directs the student to predict the results of each experiment before the experiment is run. Figures 1 and 2 show examples of a cookbook protocol and a “predictor” protocol for the same...
Predictor Protocol Excerpt

Carbon dioxide production is directly related to oxygen consumption. For this part of the exercise, connect the one-way breathing valve to the spirometer valve, and position the spirometer valve to the spirometer bell. This arrangement will allow you to collect expired gas in the spirometer. One person in the group will be the timer.

A) Determine the subject’s minute ventilation (tidal volume and breathing rate) when he/she is sitting quietly and breathing normally.

B) Consider what will happen to ventilation when the subject exercises. Predict how breathing frequency and tidal volume during exercise will compare to the resting values (more than, less than, same as) on the prediction table below

<table>
<thead>
<tr>
<th>Exercise</th>
<th>Frequency compared to rest</th>
<th>Tidal Volume compared to rest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Now have the subject exercise for several minutes by running in place, doing jumping jacks, or performing similar exercise. Immediately after the exercise is completed, measure the subject’s respiratory rate and tidal volume as you did in part A. Did ventilation change as you predicted it would?

<table>
<thead>
<tr>
<th>Rest</th>
<th>Exercise</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Did the subject increase frequency and tidal volume at the same time? What advantage might this strategy have over increasing just frequency or tidal volume? (Hint: Consider the work of breathing as ventilation increases using different strategies.)

FIG. 2. Excerpt from the “predictor” protocol directing students through the laboratory exercise activity. In this protocol, students are asked to predict the results of the experiment before they run it.

laboratory activity run by students at one of the participating institutions. The contrast is clear. Unlike the cookbook protocol, the “predictor” protocol instructs the students to make a prediction based on their current knowledge, run the experiment, and compare the actual results with the predicted results.

In this study, students in each of the participating courses were divided into three “treatment” groups. Each treatment group included one or more lab sections, with all of the students in those sections using the same protocol. Students in treatment 1 (cookbook) were directed through the laboratory with a traditional, written cookbook protocol that instructed them to make measurements of tidal volume and frequency at rest, record these results, perform exercise, repeat the measurements, record the new data, and discuss their results among themselves. This format was followed for the exercise component of the laboratory as well as components of the laboratory that were not related to this study.

Students in treatment 2 (predictor) were given a written predictor protocol that instructed them to
make measurements of tidal volume and frequency at rest, 2) record these results, 3) complete a prediction table of expected results (increase, decrease, no change in frequency and tidal volume) immediately following exercise, 4) perform exercise, 5) repeat measurements, 6) record the new data, 7) compare actual results with expected results, and 8) discuss the results. Whenever possible in this group, the predictor protocol format was also followed for the laboratory components that were not part of this study.

Treatment 3 (instructor intervention) also involved making predictions prior to running the experiment, but in this treatment, a verbal prediction was elicited from the students by one of the investigators (H. I. Modell or J. A. Michael) serving as a guest instructor. These studies were not conducted at the guest instructors’ home institutions. The guest instructor served as additional faculty for the laboratory session and did not replace regular lab instructors in the course. The guest instructors were introduced to the students in the lab section as colleagues from PERC who were interested in physiology education and were on site to conduct an experiment. The students were not told about the nature of the experiment.

The written predictor instructions used in treatment 2 were given to the students in treatment 3. However, immediately before the students performed exercise, the guest instructor asked each student in the lab group to predict what would happen to the rate and depth of breathing during the exercise period. The guest instructor did not seek any additional information about the prediction, and he did not provide any response to the prediction that would signal correctness. Following exercise and measurement of tidal volume and breathing frequency, the students in this treatment recorded their data and discussed with the guest instructor how their results compared with their predictions. In some cases, this follow-up discussion took the form of a whole class discussion aimed at comparing predictions with results. In other cases, the guest instructor discussed results with each lab group as they completed the activity.

**Posttest.** A posttest that asked students to predict how the breathing pattern would change during exercise was administered several days after the laboratory experience. The students were not told that a posttest would be given. The rules of the course, however, made it clear that they were responsible for material covered in all laboratory activities.

The question in the posttest was equivalent in form and content to that presented in the pretest. Students were presented with a situation in which a subject performed exercise, and they were instructed to predict if the subject’s depth of breathing would be greater than normal, less than normal, or unchanged.

The posttest again allowed us to categorize each student as either holding the misconception (posttest wrong) or not holding the misconception (posttest correct) following the laboratory experience. At one participating institution, the posttest question was administered at the beginning of the scheduled class conference period during the week following the lab. At the other institutions, the posttest was administered at the beginning of the next week’s lab period. Results of the posttest were matched to the results of the pretest for each student.

**RESULTS**

Results were consistent among participating institutions. Hence, the individual institutional data were pooled for analysis.

Approximately 62% of our student test population exhibited the tidal volume-frequency misconception on the pretest. This prevalence is consistent with our earlier studies with undergraduate populations (4, 5).

Student performance on the pre- and posttest can be represented in several ways. Table 1 shows the distribution of the 696 participating students according to their pre- and posttest performance and the treatment group into which they fell. To answer the question, “Can the student laboratory experience per se help students repair the misconception?” we analyzed these data using the McNemar change test, a variation of the chi-square analysis (9). This analysis is concerned only with the changes observed between performance on the pretest and performance on the posttest (from wrong to correct or from correct to wrong). The null hypothesis in this analysis is that the number of changes in each direction is equally likely.
The data in Table 1 show that this is not the case. In fact, the analysis indicated that, in each of the three treatment groups, the number of students whose performance improved on the posttest (cookbook protocol: 68/98; predictor protocol: 76/103; instructor intervention: 69/73) was significantly greater than the number of students who performed less well on the posttest (P < 0.001 for each treatment). Thus participating in the laboratory activity in which they determined the change in breathing pattern that accompanies exercise helped students correct their misconception.

The next question to be answered was, "Is the format of the protocol that directs the student through the laboratory an important factor in helping students repair this misconception?" In this case, the pretest-posttest data were placed into one of three groups.

The "success" group consisted of students who demonstrated correction of the misconception (wrong on the pretest; correct on the posttest). The "failure" group consisted of students who did not demonstrate an appropriate mental model following the laboratory activity (students who were wrong on both pretest and posttest and students who were correct on the pretest but wrong on the posttest). The third group consisted of students who demonstrated an appropriate mental model both before and after the laboratory activity (correct on both the pretest and posttest). Because the influence of the laboratory could not be assessed in this group, it was not included in the analysis. The number of students falling into the success and failure groups for the three treatments is shown in Table 2.

These data were analyzed with the chi-square test of independence (3) to determine if the three treatments were equally effective in helping students repair the misconception. No difference was found between the cookbook treatment (success rate 35.9%) and the predictor treatment (success rate 34.4%). However, the success rate for the instructor intervention treat-

---

**TABLE 1**

Pretest and posttest performance for each of the treatment groups

<table>
<thead>
<tr>
<th>Treatment 1: Cookbook Protocol (Analysis of pretest to posttest change: P &lt; 0.001)</th>
<th>Posttest Wrong, no. of students</th>
<th>Posttest Correct, no. of students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest Wrong, no. of students</td>
<td>91</td>
<td>68</td>
</tr>
<tr>
<td>Pretest Correct, no. of students</td>
<td>30</td>
<td>61</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Treatment 2: Predictor Protocol (Analysis of pretest to posttest change: P &lt; 0.001)</th>
<th>Posttest Wrong, no. of students</th>
<th>Posttest Correct, no. of students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest Wrong, no. of students</td>
<td>108</td>
<td>76</td>
</tr>
<tr>
<td>Pretest Correct, no. of students</td>
<td>37</td>
<td>87</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Treatment 3: Instructor Intervention (Analysis of pretest to posttest change: P &lt; 0.001)</th>
<th>Posttest Wrong, no. of students</th>
<th>Posttest Correct, no. of students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest Wrong, no. of students</td>
<td>19</td>
<td>69</td>
</tr>
<tr>
<td>Pretest Correct, no. of students</td>
<td>4</td>
<td>46</td>
</tr>
</tbody>
</table>

---

**TABLE 2**

"Success" and "failure" remediation rates for the three treatments

<table>
<thead>
<tr>
<th>No. of Successes</th>
<th>Treatment 1: Cookbook Protocol</th>
<th>Treatment 2: Predictor Protocol</th>
<th>Treatment 3: Instructor Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Successes</td>
<td>68 (35.9%)</td>
<td>76 (34.4%)</td>
<td>69 (65%)</td>
</tr>
<tr>
<td>No. of Failures</td>
<td>121 (64%)</td>
<td>145 (65.5%)</td>
<td>23 (35%)</td>
</tr>
<tr>
<td>Total</td>
<td>189</td>
<td>221</td>
<td>92</td>
</tr>
</tbody>
</table>

Comparison of Treatments

<table>
<thead>
<tr>
<th>Treatments</th>
<th>Success Rate</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cookbook vs. Predictor</td>
<td>35.9% vs. 34.4%</td>
<td>NS</td>
</tr>
<tr>
<td>Cookbook vs. Instructor Intervention</td>
<td>35.9% vs. 75%</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Predictor vs. Instructor Intervention</td>
<td>34.4% vs. 75%</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Note: Students who demonstrated an appropriate mental model both before and after the laboratory activity are not included.
ment (75%) was significantly higher than that for either of the other two treatments ($P < 0.001$).

**DISCUSSION**

The results of this study confirm that student laboratories can help students repair misconceptions regarding physiological phenomena (Table 1) and that the manner in which students are directed through a laboratory exercise can influence their learning (Table 2).

Perhaps the most surprising result of the study was the apparent impact on learning of even a brief, face-to-face faculty-student interaction. The written predictor protocol (see Fig. 2) asked the student to make the same kind of commitment to his or her mental model (i.e., make a prediction about what will happen in the experiment) as did the instructor. Yet, the written predictor protocol alone yielded a much lower success rate for repairing the misconception than did the instructor intervention.

The power of an active learning environment in which students test their mental models (make predictions and seek confirmation of those predictions) is demonstrated by the success of the students in the instructor intervention treatment. Social custom dictates that when someone asks you a question, it is impolite not to answer. Thus when the instructor polled a lab group, the group members, unlike their colleagues in the predictor treatment, were bound by social custom to make a prediction and verbalize it. It is important to recognize that the instructor did not seek any additional information other than a prediction, nor did the instructor provide any responses to the predictions that would signal correctness. Hence, the query was made in a low-stress situation. When everyone in the group had verbalized a prediction, the group was told to do the experiment and see if their predictions were correct. This was the same instruction that was in the written predictor protocol.

Three hypotheses that could explain the observed differences between the predictor treatment and the instructor intervention treatment come to mind. We will explore these possibilities in a follow-up study. The first is that those students with the predictor protocol did not, in fact, make predictions prior to running the experiment. They were not required to show their prediction to the lab instructor prior to running the experiment, and they may have just ignored that part of the written protocol. Because our experimental design assumed that students with the written protocols would follow the instructions, this is an important issue. In support of this hypothesis is the observation made subsequent to this study by PERC colleagues that many undergraduate students who have been given predictor-format lab protocols skipped the part of the protocol that asked them to write down predictions (D. U. Silverthorn, personal communication).

A second hypothesis is that the conditions under which the students in the instructor intervention treatment were asked to verbalize their predictions resulted in their giving more thought to their predictions than the students in the predictor treatment. It could be argued that students perceive that a response to an instructor involves more personal risk than a private response, and, as a result, they may reflect more on their mental model before making a commitment. However, students were not asked to explain their predictions, nor was any indication given as to the correctness of their predictions. The “instructor” initiating the query was someone whom the students had not seen before, nor were they likely to see this instructor in the future. Furthermore, the guest instructor had no input into the students’ grades for the course. Finally, the query was made as a simple question to group members just before the experiment was run, and responses were made by individual group members in short order. It is doubtful, under these conditions, that the stress associated with verbal predictions was significantly different from that resulting from having to record predictions on the instruction sheet.

The third hypothesis is that the efficacy of the instructor intervention arose from the discussion of the results that followed the experiment rather than the verbalization of predictions by the students. This is unlikely because the extent and nature of the discussions led by the two guest instructors were quite different, and yet, the effect of their interventions were similar. Furthermore, students in all of the treatment groups were either required to turn in laboratory reports prior to the time that the posttest was administered or expected to engage in a discus-
sion of the lab in the subsequent discussion class that began with administration of the posttest. It is reasonable to assume that they, therefore, reviewed their data and noted the breathing pattern resulting from exercise prior to the posttest. By doing so, they had the information necessary to provide a correct answer to the posttest question.

Finally, the “novelty” of interacting with a guest instructor must be considered as a confounding factor. That is, because students were not accustomed to interacting with the guest instructor, the results of the lab exercise made a greater impact than on the groups that did not engage in this “novel” behavior. It is doubtful that the novelty of the situation influenced learning. The nature and duration of the encounter varied with lab groups and with guest instructor. In addition, if the encounter, per se, made a significant impact, one would expect that any interaction with the guest instructor during the lab would lead to significant learning. This was not the case. For example, one of the guest instructors (H. I. Modell) interacted with lab groups that were engaged in other aspects of the student lab. The nature of the interaction was to ask students to make predictions and then discuss results after the activity had been completed. In one laboratory, students performed a number of breath-holding trials following a variety of maneuvers (e.g., after a normal breath, following hyperventilation, following a maximal inspiration, following a maximal expiration). As part of this exercise, expired gas was collected and analyzed for carbon dioxide content. Following a discussion with the guest instructor that focused on factors limiting breath-holding time and the role of carbon dioxide in regulating ventilation, the group was queried about what would happen to breath-holding time following hyperventilation. Students correctly predicted that breath-holding time would increase. However, when the guest instructor asked them to explain the rationale behind their prediction, most responded that more oxygen would be available in their lungs. If the guest instructor had made an impact, one would expect that the discussion from several minutes earlier would have been remembered and that students would recognize that the primary factor would be the decreased carbon dioxide level in the lung rather than the increased oxygen level in the lung.

Perhaps the most important message that the results of this study convey relates to the factors that contribute to success of the student laboratory as a learning environment. In this experiment, all treatment groups at a given institution made the same measurements before and immediately following exercise. They certainly learned something from participating in the activity. However, the participation, per se, did not determine the efficacy of the learning environment. Instead, the way in which students participated appeared to be the critical factor in promoting learning. The process of verbalizing predictions and testing whether those predictions were correct resulted in approximately twice the “learning” (as measured by the rate of successfully repairing the misconception) that occurred when this process was not followed. To optimize the learning environment of the student laboratory, we must recognize that the important component is not “active,” but active learning.

We thank colleagues in the Physiology Educational Research Consortium for input during the planning phases of this study and for comments during the preparation of this manuscript.

This research was supported by National Science Foundation (NSF) Grant DUE-9652782. Any opinions, findings and conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the NSF.

Address for reprint requests and other correspondence: H. I. Modell, NRCLSE, PO Box 51187, Seattle, WA 98115 (E-mail: modell@ocke.hs.washington.edu).

Received 23 August 1999; accepted in final form 11 January 2000.

References