Enhancing active learning in the student laboratory

Harold I. Modell, Joel A. Michael, Tom Adamson, and Barbara Horwitz

1Physiology Educational Research Consortium, Seattle, Washington 98115; 2Department of Molecular Biophysics and Physiology, Rush Medical College, Chicago, Illinois 60612; and 3Section on Neurobiology, Physiology and Behavior, University of California, Davis, California 95616

Submitted 26 January 2003; accepted in final form 22 April 2004

We previously examined how three approaches to directing students in a laboratory setting impacted their ability to repair a faulty mental model in respiratory physiology (Modell, HI, Michael JA, Adamson T, Goldberg J, Horwitz BA, Bruce DS, Hudson ML, Whitescarver SA, and Williams S. Adv Physiol Educ 23: 82–90, 2000). This study addresses issues raised by the results of that work. In one group, a written protocol directed students to predict what would happen to frequency and depth of breathing during exercise on a bicycle ergometer, run the experiment, and compare their results to their predictions (“predictor without verification”). In a “predictor with verification” group, students followed the same written protocol but were also required to show the instructor their predictions before running the experiment. Students in a third group reported their predictions verbally to an instructor immediately before exercise and reviewed their results with that instructor immediately after exercise (“instructor intervention group”). Results of this study were consistent with our earlier work. The predictor with verification and predictor without verification protocols yielded similar results. The instructor intervention protocol yielded higher success rates in repairing students’ mental models. We subsequently assessed the efficacy of a prediction period at the beginning of the lab session and a wrap-up period at the end to compare predictions and results. This predict and prediction period at the beginning of the lab session and a wrap-up period at the end of the experiment (predictor protocol). In the third approach, students were given the written predictor protocol, but they were also required to verbalize their predictions to an instructor before running the experiment (instructor intervention protocol). The instructor’s role in this process was only to elicit a verbal response. No comment was made regarding the students’ predictions (see study 1, Treatment 3).

Results indicate that participation in a laboratory experience, regardless of the protocol followed, helped some students recognize the error in their mental models and provided a basis for modifying them. However, the improvement shown by students using the cookbook and written predictor protocols was not different. Although this finding implies that predicting the outcome of the experiment (i.e., applying the student’s current mental model to the problem) was no more effective in helping students recognize the limitations of their models than following a series of instructions that required no prior thought of the mechanisms involved, the results of the instructor intervention protocol showed that this was not the case. Only about 35% of students using the cookbook or predictor protocol performed better on the posttest than on the pretest. However, when students were required to verbalize their predictions, 75% of the students who provided incorrect answers on the posttest correctly indicated on the posttest that tidal volume would increase when ventilation increased.

The large discrepancy between the results using the written predictor protocol and the instructor intervention suggests that those students who used the written predictor protocol may not have made predictions as directed. Another possible explanation of the discrepancy may be related to the presence of a guest instructor in the learning environment. The instructor who prompted the students’ predictions before running the experiment in the instructor intervention protocol was not a regular member of the course faculty. Hence, the discrepancy may have been related to the students’ response to interacting with a guest instructor.

The current set of experiments was designed to address these issues. If the results proved to be consistent with those of the previous study, confirming the necessity of interaction with the instructor for optimal results, success in the laboratory may involve changes in classroom management for the class period. In some settings, it may not be possible for lab instructors to interact with each lab group before each lab activity. Hence, in a second set of experiments, we examined an alternate approach to classroom management that would provide an opportunity for eliciting predictions from students.
STUDY 1

Methods

Students from five institutions participated in our original study (8). Courses ranged from the introductory anatomy and physiology level to the upper-division biology major level. Because the results of that study were consistent among participating institutions, we chose to conduct this study in an upper-division biology course for majors at one institution. We followed the same general experimental design as in our original study.

Basis for assessment. Assessment of the success of each experimental intervention was based on the success rate with which students repaired their faulty mental model of tidal volume changes that occur when minute ventilation is increased. Michael et al. (3) presented the following problem to 700 undergraduate students.

A classmate is late for the exam and runs up five flights of stairs to the exam room. When she arrives at the top, 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 it was at rest
b) less than it was at rest
c) unchanged

Approximately 60% of the population surveyed predicted that, when minute ventilation increases, the tidal volume either decreases or remains unchanged. In reality, increased minute ventilation is a result of increased frequency and increased tidal volume. We have subsequently increased the size of our survey population to ~2,000 students. Approximately 55% of students in this larger sample incorrectly predicted that tidal volume would decrease or remain unchanged when minute ventilation increased.

Pretest. During the week before the laboratory session that included the exercise activity, students completed a pretest to determine the prevalence of the tidal volume-frequency-ventilation misconception. The pretest question was the same as that posed by Michael et al. (3). On the basis of 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. As in our earlier study, the misconception was not specifically addressed in the lecture portion of the course.

Laboratory. During the laboratory session dealing with the respiratory system, students rotated among three activity stations: spirometry, breathholding, and exercise. The exercise portion of the lab was the focus of our experiment. At this station, students measured the breathing pattern (frequency and tidal volume) of a subject before and during a period of exercise on a bicycle ergometer. Data were obtained online using a pneumotachograph, integrator, and related computer data acquisition equipment. All students in each lab section were provided with direction for this portion of the lab according to one of three treatments. In treatment 1 (predictor), students were provided with the same written predictor protocol as was used in the original experiment. In this protocol, students were asked to complete a table in which they predicted how breathing frequency and tidal volume measured during exercise would compare with what they measured during rest (increase, decrease, not change). The protocol also directed the students to compare their predictions with the experimental results that they obtained.

In treatment 2 (predictor with verification), students were given the same instructions as in treatment 1. However, to verify that predictions had, in fact, been made, students were required to show their completed prediction table to the instructor before performing the exercise portion of the lab.

Treatment 3 (instructor intervention) also involved making predictions before running the experiment. However, as in our original study, a verbal prediction was elicited from the students. The same written predictor instructions used in the other two treatments were given to the students in treatment 3. However, immediately before the students performed exercise, a guest instructor (H. Modell or a graduate student teaching assistant) 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 reasoning behind the prediction, and he did not provide any response to the prediction that would signal correctness. After exercise, the students and guest instructor reviewed the group’s data to determine whether their results were consistent with their predictions. The exact nature of the discussion depended on questions asked by the students. In some cases, the instructor only reviewed the data with the students. In others, specific issues raised by the students were discussed.

The guest instructor served as additional faculty for the lab session and did not replace regular lab instructors in the course. The guest instructor was introduced to the students in the lab section as being a guest from the Physiology Educational Research Consortium interested in physiology education and in class to conduct an experiment. The students were not told about the nature of the experiment.

Posttest. A posttest was administered during lab discussion session held the week after the lab. Students were not aware that a posttest would be given. The posttest question is shown below.

Over summer vacation, you and a friend drive to the top of Pike’s Peak, where the elevation is 10,000 ft. You notice that, while on the top of the mountain, your friend’s breathing frequency (no. of breaths/min) is increased. At the same time, his depth of breathing (the amount of air he takes in each breath) is

a) greater than it was at the bottom of the mountain.
b) less than it was at the bottom of the mountain.
c) the same as it was at the bottom of the mountain.

The posttest again allowed us to categorize each student as either holding the misconception (posttest wrong) or not holding the misconception (posttest correct) after the laboratory experience (see Discussion). Results of the posttest were matched to the results of the pretest for each student.

Study 1 Results

Two hundred nineteen students participated in this study. On the pretest, ~52% of these students incorrectly predicted what would happen to the depth of breathing (tidal volume) when ventilation increased. This is consistent with the prevalence of this misconception seen in our earlier studies (2, 3, 8).
The data were analyzed with a $\chi^2$ test of independence (1) to determine whether improvement from the pretest to posttest was related to the way in which the students had been directed in the lab (treatment group). To perform this analysis, the pretest-posttest results 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 after the laboratory activity (either wrong on both pretest and posttest or 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 lab activity (correct on both pretest and posttest). Because the influence of the lab 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 each of the three treatments is shown in Table 1. Data analysis revealed (Table 1) that there was no significant difference in success rate of students who were given only the written protocol (14.6%) and those whose written predictions were verified before their running the experiment (12.8%). In contrast, the success rate of students who had interacted with the guest instructor (78%) was significantly higher than that of students in either of the other treatments ($P < 0.001$). There was no difference detected between the performance of students who interacted with the investigator (H. Modell) and those who interacted with the graduate student teaching assistant.

**STUDY 2**

**Rationale**

The results of study 1 suggest that verbalizing predictions before running the experiment is a critical element in determining the success of the lab activity as a vehicle to help the learner recognize the limitations of his or her mental models and provide a basis for revising them. This finding may raise a serious classroom management issue in some settings. Laboratory instructors may not be able to interact with lab groups as the guest instructors did in our study. For example, the lab instructors in the course in which our experiments were run were concerned with making sure that students had appropriately working equipment for each activity, monitoring (and, if necessary, directing) student progress during the lab, answering students’ questions, and managing a myriad of other unforeseen details that may have arisen during the lab period. Hence, in this course, the instructor could not always be available when it was time for individual lab groups to make their predictions and compare their results with their predictions. One approach to addressing this problem is to have the students make their predictions for all lab activities for the day during a “prediction period” held at the beginning of the laboratory period. This would be followed by a “wrap-up” session held at the end of the lab period in which students would compare their results with their predictions and discuss their results. To determine whether this design (i.e., prediction period at the beginning of the lab followed by a wrap-up period at the end of the lab) was as effective as the instructor intervention (treatment 3) protocol used in study 1, we ran another study during the next offering of the course.

**Methods**

The same general experimental design used in study 1 was used in this study. Three protocols were followed with respect to providing direction to the students. The predictor protocol without verification (study 1, treatment 1) served as the control. To address the classroom management question, the instructor intervention protocol (study 1, treatment 3) was compared with a protocol that we have called the predict & wrap-up protocol. In this new predict & wrap-up protocol, a prediction table was put on the white board in the lab. At the beginning of the lab period, each student was required to predict what would happen to measured parameters for each lab activity that constituted an experiment. For example, in addition to the exercise experiment, the lab included a breath-holding experiment. Students measured breath-hold times after various breathing maneuvers (e.g., hyperventilation, rebreathing, etc.). In the predict period, each student was asked to indicate by a show of hands whether he or she predicted that breath-hold time would increase, decrease, or not change after one of the breathing maneuvers. All students were required to make a prediction. The predictions were tallied, and the tally numbers were recorded in the predictions table. For the exercise experiment, students predicted what would happen to breathing frequency and depth of breathing (tidal volume) during exercise. The students then performed all of the scheduled lab activities for the day.

During the last 20 minutes of the lab period, the class reconvened for the wrap-up session. During this time, the results obtained by the various lab groups were compared with the predictions made during the predict period, and possible reasons for discrepancies were discussed. The guest instructor (H. Modell or a graduate teaching assistant) led the predict period and wrap-up discussions.

**Results**

Two hundred thirty-four students participated in this experiment. The prevalence of the frequency-tidal volume misconception uncovered by the pretest was consistent with previous experiments. The number of students falling into the success
and failure groups for the three protocols in this experiment are shown in Table 2. The performance of students in the predictor and instructor intervention groups was not significantly different from the corresponding groups in study 1. As in study 1, student performance was independent of the specific guest instructor with whom students interacted (instructor intervention and predict & wrap-up protocols).

Analysis of the data in Table 2 indicates that, although the predict & wrap-up protocol results in a greater success rate than the predictor protocol alone (21.2 vs. 44.2%; \( P < 0.01 \)), it is not as effective as the instructor intervention protocol (44.2% success rate vs. 84.4% success rate; \( P < 0.001 \)).

### DISCUSSION

Assessment of learning in our experiments was based on the success of students refining their mental models of the determinants of minute ventilation. Undergraduate texts and courses do not generally discuss the determinants of breathing pattern (tidal volume-frequency combination). Students are told only that minute ventilation depends on tidal volume and breathing frequency. This was the case for students participating in this experiment.

Michael and colleagues (2, 3) examined the frequency-tidal volume misconception that we used for assessment in this study. These authors asked undergraduates to predict what would happen to the depth of breathing when subjects increased their ventilation in response to exercise or altitude. The responses were essentially the same. Fifty to sixty percent of students predicted that the depth of breathing would decrease regardless of the physiological stimulus for the increased ventilation. A prevalent explanation provided by students for the incorrect prediction in both situations was that breathing faster did not allow time to breathe deeper. Hence, although our pre- and posttest questions were not equivalent from a physiological mechanism standpoint, they were considered to be equivalent to assess the prevalence of the frequency-tidal volume misconception in our undergraduate student population. The validity of this assumption is further supported by the fact that the results from the predictor and instructor intervention protocols in these experiments are consistent with those of our earlier experiments (8) in which both the pretest and posttest questions dealt with a response to exercise.

Table 2. “Success” and “failure” rates for the three treatments in study 2

<table>
<thead>
<tr>
<th>Treatments</th>
<th>Success rate</th>
<th>( P )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predictor Protocol</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of successes</td>
<td>14 (21.2%)</td>
<td></td>
</tr>
<tr>
<td>No. of failures</td>
<td>52 (78.8%)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>66</td>
<td></td>
</tr>
<tr>
<td>Instructor Intervention</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of successes</td>
<td>65 (84.4%)</td>
<td></td>
</tr>
<tr>
<td>No. of failures</td>
<td>12 (15.6%)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>77</td>
<td></td>
</tr>
<tr>
<td>Predict &amp; Wrap-up Protocol</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of successes</td>
<td>34 (44%)</td>
<td></td>
</tr>
<tr>
<td>No. of failures</td>
<td>43 (55.8%)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>77</td>
<td></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>Predictor vs. instructor</td>
<td>21.2 vs. 84.4%</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Predictor vs. predict &amp; wrap-up</td>
<td>21.2 vs. 44%</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Instructor intervention vs.</td>
<td>84.4 vs. 44%</td>
<td>&lt;0.005</td>
</tr>
<tr>
<td>predict &amp; wrap-up</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Success = pretest wrong and posttest correct; failure = pretest wrong and posttest wrong, or pretest correct and posttest wrong.

Note: students who demonstrated an appropriate mental model both before and after the laboratory activity are not included.

The frequency-tidal volume misconception represents the result of applying an ill-formed mental model of the mechanism determining minute ventilation to the question of how minute ventilation increases during exercise. In the learning process, students must recognize that their current model(s) will not provide accurate predictions of real-world events before they are willing to refine their model(s). That is, students must test their mental models before they will refine them. The iterative process of building, testing, and refining mental models is the key element of an active-learning environment (7, 4).

The student laboratory provides an excellent opportunity for students to engage in the process of building, testing, and refining mental models. However, in many laboratory courses, students don’t behave in a manner consistent with this process; hence, learning is compromised. The factors that potentially contribute to this failure to engage in active learning are many and varied in their origin. They may be related to the design of the laboratory activities, the direction (or lack of direction) that students receive in carrying out these activities, the expectations of the student with regard to the purpose of the laboratory, or other issues yet to be defined.

Our earlier study (8) examined the influence of some of these factors. The validity of the build-test-refine mental model learning strategy was confirmed when students were required to make verbal predictions. However, similar success was not observed when students were told to make predictions on the written protocol. We hypothesized that the large difference in performance between the predictor protocol and the instructor intervention protocol could be due to one or more of three factors: 1) students with the predictor protocol did not, in fact, make predictions before running the experiment; 2) the conditions under which the students were asked to verbalize their predictions resulted in their giving more thought to their predictions than the students in the predictor protocol; and 3) 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. A fourth factor could have been related to the “novelty” of interacting with a guest instructor who may have been perceived by the students as an “expert” in the field.

The experiments conducted as study 1 in this investigation were designed to test these hypotheses. Results from students whose predictions on the written protocol were verified before running the experiment were not different from those of students whose predictions were not verified. The success rates on the posttest were comparable to those obtained in the cookbook and predictor protocols in our earlier study (8). Although we verified that students made predictions before running the experiment, we do not know when those predictions were made relative to running the experiment. They could have been made prior to the students coming to class, or they could have been made while the lab group was engaged in one of the other two activities that made up the remainder of the lab period.

As in our earlier investigation, the instructor intervention protocol was very effective in helping students test and refine their mental models. It is doubtful that this success was related to the specific guest instructor. The success rate on the posttest achieved by students who interacted with one of the investigators, a respiratory physiologist by training (72%), was com-
parable to that achieved by students who interacted with the graduate student teaching assistant (79%).

The results of our earlier investigation and study 1 in this investigation indicate that students are most likely to engage in the build-test-refine mental model learning strategy if they are required to make predictions publicly followed by an examination and discussion of the experimental results in relation to those predictions. It is important to recognize that, in these experiments, explanation or justification of predictions was not solicited. Any discussion of the reasoning behind predictions was conducted as part of the follow-up period, in which experimental results were compared with predictions.

What do these results imply in terms of classroom practice? They suggest that each lab group should be polled for their predictions about the outcome of the experiment before each experiment is performed, and their results should be reviewed with an instructor immediately after the experiment is completed. However, this may not be practical in some laboratory settings. In our experiments, the guest instructor performed this duty. The instructor (or teaching assistant) who was in charge of the lab section was not involved in this process. Instead, he or she made sure that the lab ran smoothly. This involved making sure that students had the appropriate supplies for the other lab activities, that the equipment for the exercise activity as well as the other four lab stations were working properly for all groups, and that students’ questions were answered. We hypothesized that holding a prediction session for the whole class at the beginning of the period and a wrap-up at the end of the period might solve this classroom management problem. The results indicated that, although this approach yielded better results in terms of learning than the control, it was not nearly as effective as the instructor intervention protocol. One explanation for this result is that the students may not remember what they predicted or they may not remember the reasoning behind their predictions when considerable time has elapsed between prediction and comparison of results. Hence, when experimental results are compared with their predictions, the impact of the build-test-refine mental model learning strategy is diminished.

This is consistent with anecdotal evidence observed during a study on high school physics students examining force to explain motion (5). Minstrell observed that, “in having students work through worksheets of lab activity, they merrily went along answering the questions on the lab sheet. If there were several other questions between the prediction and observation, they did not make the connection between prediction and observation. In our instructional design, we now try to have the specific prediction that is going to be tested done just before we go to the ‘let’s find out’ part. That usually results in a closer tie between the predictions and the surprise or confirmation of the observation” (J. Minstrell, personal communication).

There may be another factor that may have a significant impact on this process. When instructors design any classroom activity, they assume that students will approach the activity in a way that is consistent with the design. This, however, is not necessarily the case. The reasons that students take the course are often varied. Some enroll because they are interested in learning as much as they can about the subject. Others enroll because they must take a laboratory course to meet curriculum requirements. Thus students’ expectations for the goals of the laboratory experience often do not match the instructor’s goals. In our experiment, for example, we assumed that the students would reflect on the predictions when it came time to run each of the laboratory activities. This was part of the reason that they were on display on the white board. However, many students view the laboratory as another setting in which they obtain information (i.e., getting the “right” answer) rather than as a setting to practice problem-solving skills. As most lab instructors would confirm, many students come to the laboratory unprepared for the day’s activities. Although they may be prepared in terms of factual information related to the lab (e.g., the students in our test course had a quiz before beginning the lab procedure), they are not prepared in the sense of having given much thought to the physiological mechanisms involved or the implications of the activities in terms of those mechanisms. For these students, the goal of the lab is to record the prescribed data as quickly as possible and leave rather than taking the time to apply the build-test-refine mental learning strategy and reflect on what they are doing. Because prediction is not immediately followed by the experiment and comparison components, these people may not view the prediction and wrap-up periods as linked activities that are important components of the lab. In the case of our experiment, the prediction and wrap-up periods had not been part of the laboratory experience in previous lab sessions. Hence, to some students, the relevance of these activities to the overall lab experience may not have been obvious. If, however, the prediction and wrap-up periods were regular features of the lab sessions, and if the importance of testing their current mental models to the learning process was emphasized, the degree of success achieved by this approach might improve as the students gain more practice with the process.

ACKNOWLEDGMENTS

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

GRANTS

This research was supported by National Science Foundation (NSF) Grant REC-9909411. 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.

REFERENCES