A multiyear approach to student-driven investigations in exercise physiology

Kathleen A. FitzPatrick and Jay Campisi
Department of Health Sciences, Merrimack College, North Andover, Massachusetts

Submitted 8 July 2009; accepted in final form 15 September 2009

FitzPatrick KA, Campisi J. A multiyear approach to student-driven investigations in exercise physiology. Adv Physiol Educ 33: 349–355, 2009; doi:10.1152/advan.00056.2009.—Many undergraduate institutions offer individual research opportunities for upper-level students in independent study courses and summer undergraduate research programs. These are necessarily limited to a small number of subjects. Greater numbers of subjects can benefit from incorporating student-directed investigative experiences into laboratories in standard courses. In human performance investigations, any single course may not offer sufficient numbers of subjects to adequately test hypotheses comparing population groups or to examine longitudinal trends. In this exercise physiology course, exercise testing was conducted in three areas: 1) techniques of body composition analysis, 2) field tests for the estimation of maximal oxygen consumption, and 3) maximal anaerobic and aerobic power. All students enrolled over a 10-yr period participated as subjects and as testers. Working in small research groups, students added their results to those from previous years, generated a variety of hypotheses (correlations between tests, subgroup differences, etc.), and tested them statistically using the complete data set of 217 subjects. They then engaged in collaborative writing and peer review to prepare formal papers on their results. The multiyear approach allowed students to situate their work within and contribute to the accumulation of a large database and to practice essential scientific skills of hypothesis formation, data collection and analysis, collaborative work, and scientific communication. In addition, due to the larger number of subjects available to analyze, students observed statistically significant differences between test groups in the multiyear database that they were unable to demonstrate when conducting analysis on a single course. Finally, the large number of subjects and statistical power offered by the use of the database provides distinct pedagogical advantages.

The advantages of active, investigative pedagogies in the science laboratory have been emphasized by many national organizations [National Science Foundation (28), National Research Council of the American Academies of Science (25–27), the American Association for the Advancement of Science (3), and the Association of American Colleges and Universities (5)]. An increasing body of evidence from classroom assessment has demonstrated the use and effectiveness of active, investigative pedagogies in the biology laboratory (1, 6, 8, 11, 19–21, 23, 24, 30, 31, 33, 37). We have previously described a sequence of two laboratory experiences within our programs in which students design and execute their own experiments, analyze data, and communicate their findings (12, 13). The course described here represents the third in this sequence.

As a discipline, exercise physiology encompasses both basic experimental physiological science and the application of research-based information to sport and health-related fitness and performance. The current increased focus on the public health and epidemiological aspects of physical activity in relation to chronic disease (15) makes it important for health profession students to understand and be introduced to fundamental research methods unique to the study of physical activity and human performance. The laboratory experience described here is designed to provide an investigative student-directed, classroom-based experience. Since many of the students in this course are preparing for certification exams in athletic training and the health and fitness fields, it is necessary that they develop an understanding of and acquire the skills to carry out standard physical fitness test protocols (equipment and interpretation for cardiorespiratory fitness, body composition, flexibility, strength, power, and endurance). It is also important that students appreciate the principles of experimental design, measurement, evaluation, and data interpretation unique to human performance research.

Additionally, the education literature challenges professors to be more creative and innovative in their active learning exercises (22) and to incorporate evidence-based practice in the classroom (7).

In some circumstances, student-designed experiments may not achieve the goal of ensuring that all class members acquire the necessary skills in all areas, and often the small numbers of subjects involved in such experiments make it unlikely that the results will achieve statistical significance. It is still important, however, that students are able to engage in hypothesis formation, accurate collection, statistical analysis, and interpretation of data and scientific communication. To this end, we have developed a process in which student teams practice and perform a variety of fitness assessments on class members, adding their data each year to an existing database. They formulate and test hypotheses, based on a large aggregate population sample, by examining the appropriateness of various field tests for the population and characterizing the fitness levels of a population over time and compared with norm populations. Students then communicate their results through collaborative scientific writing and peer-reviewed reports.

Methods

Context of the study. Our institution is a small (2,100 students), private, comprehensive college. In the 10 yr described here (1998–2008), 217 students (58% female and 42% male) completed the exercise physiology laboratory, which accompanies a semester-length lecture course. The course enrolls 10–16 students per laboratory section per year. All students were junior Sports Medicine majors. They had all completed courses in cell biology, biological investigations, and introductory genetics as well as two terms of general chemistry, human anatomy, human physiology, and two to three clinical courses and statistics; these students were preparing for careers and graduate study in athletic training, physical therapy, and exercise physiology.

Students attended 12 weekly 3-h faculty-facilitated laboratory sessions. Initially, the unique features of experimentation with human...
subjects were introduced; informed consent and confidentiality were emphasized, and each student reviewed and signed a consent form and was assigned a numerical identifier to be used in all subsequent experiments. The experimental protocol was approved by the Institutional Review Board of Merrimack College. The subjects were medically screened using guidelines from the American College of Sports Medicine for fitness assessment and exercise participation (4).

Course organization. In the years before 2003, students simply performed the assessments and reported on the data from that year. Beginning in 2003, after introductory remarks on the nature and theory of each assessment, students were introduced to the database accumulated over previous years from 1998 to the present (which was available to them as a downloadable Excel spreadsheet on the Blackboard course website). Before each test was performed, research groups formed a variety of hypotheses based on the data from previous years, previous courses, background reading, and experience. Hypotheses were discussed with the instructor and refined. The assessments were then performed, using identical protocols, on all class members unless medically contraindicated. Calculations and analyses were conducted as needed on raw data and reviewed by the instructor. The data collected from all subjects in each year’s class (beginning in 1998) were added to the master electronic database from previous years, which was posted on the Blackboard course website (beginning in 2003). The aggregate data were analyzed graphically and statistically relative to each hypothesis. Our collaborative writing and peer review process has been described previously (12). Briefly, a principal investigator in each group drafted a research report. Team members completed a structured review and peer editing of the draft. The principal investigator revised the draft based on peer comments and produced the final report for the team.

Investigators/subjects. The investigators/subjects in the aggregate database (1998–2008) included 90 males [mean age (SD): 21.0 (2.0) yr and mean body mass index (BMI) (SD): 26.5 (4.1) kg/m²] and 127 females [mean age (SD): 20.4 (1.43) yr and mean BMI: 23.9 (3.1) kg/m²]. Working in teams of four, students performed standard health fitness assessments on all team members.

Body composition estimations. Bioelectric impedance analysis (BIA-Tanita TBF 512 body composition analysis scale (BIA)) and skin folds (sum of 3 gender-specific site skin-fold thicknesses measured with standard calipers, Siri equation) were used to determine body density and body fat percentage (4). The body fat percentage of 85 students was also assessed with the BodPod Body Composition Tracking System (Life Measurement), which determines body density from air displacement plethysmography and body fat percentage from the Siri equation. Height was measured using a wall-mounted stadiometer, and weight in minimal workout clothes was recorded from a medical-grade digital scale. BMI was calculated as weight (in kg)/height (in m²). Percentile values of body fat percentage for a predominantly white college-educated population of 20–29 yr olds were used for comparison (4).

Cardiorespiratory fitness (maximal O₂ consumption) assessments. Maximal O₂ consumption (V̇O₂max) was determined using a treadmill with the Bruce protocol (4) and cycle ergometer with the YMCA cycle protocol (4) by monitoring O₂ consumption and CO₂ production in subjects performing exercise of increasing intensity to the point at which the subject was unable to continue. Field estimates for V̇O₂max included the progressive 20-m shuttle run (9) and the nonexercise activity questionnaire (18). Laboratory estimates included the YMCA and Astrand-Rhyming submaximal cycle ergometer tests (4). Percentile values of V̇O₂max, derived from student-generated maximal treadmill tests and for a predominantly white college-educated population of 20–29 yr olds were used for comparison (4).

Power assessments. The Sargent vertical jump (VJ) test estimated explosive power from VJ height and body weight. The anaerobic test estimated power from the maximum effort pace number of one-leg steps completed in 15, 30, and 60 s and step height and body weight (1). Each year, a subset of students completed the Wingate anaerobic power test on a computerized Monark Ergomedic 874E cycle (14). Results from the Wingate test and population norms for power were used for comparison (1).

Data analyses. Students routinely conduct analyses using Microsoft Excel and, when appropriate, StatView or KaleidaGraph. Two-factor ANOVA was performed to analyze body composition (athlete and nonathlete) data. Three-factor ANOVA was performed to analyze body composition modality (BIA, BodPod, and skin folds) data. Two (athlete vs. nonathlete) × two (male vs. female) ANOVA was used to examine the effect of physical activity status and sex on body composition and power. Where appropriate, for multiple hypothesis testing, post hoc analyses were conducted using the false discovery rate (FDR) procedure (10); two-tailed α was set at P < 0.05, and the FDR was set at <5%. In all of the figures, the values are shown as group means (SD).

Assessment. At the end of this term, students were given instructions to complete the student assessment of their learning gains (Salgains), a web-based instrument that includes several items specifically related to various aspects of the laboratory (35, 36). A standard departmental teacher-course evaluation was administered in class. Data from 2004 are not included.

RESULTS

Student experiments and results. After discussion of the theory and nature of each test and before the performance of the fitness assessment, each team formed a working hypothesis concerning some aspect of the expected outcomes. These hypotheses were discussed with the instructor before measurements were performed. After students had been instructed on the proper technique for each test, the assessments were completed on all team members, and the necessary calculations were made and checked; the final value of the parameter being studied for each new subject was entered into the database containing values for all previous years for the same assessments performed using the same protocols. Each group then analyzed the aggregate data to determine if their hypothesis was supported.

The multiyear database makes it more likely that statistically significant results will be observed when analyzing data. For example, one recent group of students hypothesized that if all field tests of body composition are equally valid, the results for these subjects measured by each test should not be significantly different from each other. Figure 1A demonstrates the results of this investigation when a single year’s data were used. ANOVA indicated no significant difference between the three measures of percent body fat when BIA, BodPod, and skin fold data were used (n = 10 subjects). These results would lead one to believe that all three measures of body composition produce the same average value for the same population. However, analysis of the multiyear database (n = 85) revealed that a significant difference between the different test means existed (P < 0.0001; Fig. 1B and Table 1). The FDR analysis indicated that the skin-fold technique resulted in percent body fat estimates that differed significantly from both BIA and BodPod.

The multiyear database also allows for hypotheses to be tested that would not otherwise be feasible using data from a single year. For example, a group of students wanted to examine if percent body fat differed between athletes (defined as members of a varsity collegiate athletic team) and nonathletes and if there was a difference between male and females (measured via BodPod). Figure 2A shows the results of this investigation when a single year’s data were used. In this class...
of 10 students only 2 students were classified as athletes, and both were of the same sex, making significant differences between athletes and nonathletes unlikely ($P = 0.8465$) and not allowing for a comparison between sexes to be made. However, as shown in Fig. 2B, when all of the data were used, ANOVA indicated a significant difference ($P = 0.0495$) between athletes ($n = 20$) and nonathletes ($n = 65$). In addition, as shown in Fig. 3, not only did athletes have significantly less percent body fat, but males had significantly lower percent body fat than did females ($P < 0.0001$). This analysis also allows one to examine if a sex × athlete interaction occurred (e.g., is the difference between males and females different in athletes than nonathletes). ANOVA indicated that no such interaction existed ($P = 0.9150$; Table 2).

Similar advantages offered by the use of the multiyear database set were seen in the area of power and $\dot{V}O_{2\max}$. The multiyear database allows students to look for significant interactions between groups that would not be possible using a single laboratory section. For example, one hypothesis examined if explosive power (estimated from the Sargent VJ test) differed between male and female athletes versus male and female nonathletes. Using data from a single laboratory section resulted in a significant difference between males and females ($1,054$ (172) vs. $532$ (87), $P = 0.0002$) but no difference between athletes and nonathletes ($631$ (45) vs. $768$ (327), $P = 0.5876$) in a class of 10 students. The small number of students in each group did not allow for an analysis of interactions to occur (e.g., is the difference between males and females different in athletes than nonathletes). However, when all of the data were used, this comparison was possible, and, in fact, ANOVA indicated a significant sex × athletic status interaction ($P = 0.0478$). This analysis demonstrated that while female athletes tended to achieve greater power than female nonathletes ($606$ (98) vs. $518$ (97) W), male athletes tended to demonstrate lower power than male nonathletes ($957$ (119) vs. $1,045$ (236) W).

**Assessment.** The Salgains results for 2003, 2005, 2006, and 2007, completed by 58 of 71 students (82%), indicated a positive response to various aspects of the laboratory experience, with mean ratings in the $3.59$–$4.12$ range (where $3 = $ moderate/somewhat and $4 = $ much help/a lot; Table 3). The responses did not seem to differ among the years and were not tested for significance, given the small numbers of respondents in early years. Table 2 also shows mode scores of $4$ and a high percentage of ratings in the $4$–$5$ range for the majority of responses. Representative narrative comments regarding the laboratory experience (as opposed to characteristics of the instructor) are shown in Table 4. The majority of comments identified positive features of the laboratory, including the use of the database, group work and interactions, and the active nature of the laboratory as well as the laboratory report writing. Negative comments generally related to heavy workloads involved with laboratory reports, although other comments noted the reports as helpful, and the difficulties of peer reviewing and group collaboration. Such comments are common among students participating in this peer review mode of report writing (12). Grades on laboratory written reports were consistently high, but it is interesting to note that the mean laboratory report grade increased from 86/100 during the final year of using

### Table 1. Percent body fat measurements by BIA, BodPod, and skin folds for a single year’s class and from the multiyear database

<table>
<thead>
<tr>
<th></th>
<th>Mean (SD)</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Single year’s class</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIA</td>
<td>21.6 (6.5)</td>
<td>16.9, 26.2</td>
</tr>
<tr>
<td>BodPod</td>
<td>22.8 (9.3)</td>
<td>16.2, 29.4</td>
</tr>
<tr>
<td>Skin folds</td>
<td>16.3 (7.7)</td>
<td>10.8, 20.9</td>
</tr>
<tr>
<td><strong>Multiyear database</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIA</td>
<td>25.5 (8.7)</td>
<td>23.7, 27.3</td>
</tr>
<tr>
<td>BodPod</td>
<td>25.5 (9.2)</td>
<td>23.6, 27.5</td>
</tr>
<tr>
<td>Skin folds</td>
<td>19.1 (8.6)</td>
<td>17.2, 20.9</td>
</tr>
</tbody>
</table>

$n = 10$ subjects in the single year’s class and $85$ subjects in the multiyear database. BIA, bioelectric impedance analysis.

![Fig. 1. Mean percent body fat values for college-age subjects. Values were estimated using bioelectric impedance analysis (BIA), the BodPod air displacement plethysmography instrument, and skin folds. Values are means; error bars indicate SD. A: results for one class year ($n = 10$ subjects). ANOVA indicated no significant differences between methods ($P = 0.1678$). B: results from 10 class years ($n = 85$ subjects). *False discovery rate analysis resulted in a significant difference between males and females [1,054 (172) vs. 532 (87), $P = 0.0002$] but no difference between athletes and nonathletes [631 (45) vs. 768 (327), $P = 0.5876$] in a class of 10 students. The small number of students in each group did not allow for an analysis of interactions to occur (e.g., is the difference between males and females different in athletes than nonathletes). However, when all of the data were used, this comparison was possible, and, in fact, ANOVA indicated a significant sex × athletic status interaction ($P = 0.0478$). This analysis demonstrated that while female athletes tended to achieve greater power than female nonathletes [606 (98) vs. 518 (97) W], male athletes tended to demonstrate lower power than male nonathletes [957 (119) vs. 1,045 (236) W].](http://advan.physiology.org/DownloadedFrom/10.220.33.6)
single class data to 89/100 during the years when the multiyear database was used, although the difference was not statistically significant and could depend on other variables across the years.

**DISCUSSION**

**General comments.** Student-driven investigative experiences can be incorporated into standard human performance laboratories by generating and adding to an ongoing database, which can then be used to formulate and test population-based hypotheses and characterize the fitness of this college age population. Student responses to the experience were positive, with students rating various aspects of the laboratory experience as of “much help” in improving their skills and understanding. The multiyear approach allows students to situate their work within a larger investigative context, to improve the statistical power of their analyses, to learn to manipulate a large database, and to practice essential scientific skills of hypothesis testing, data collection and analysis, collaborative work, and scientific communication.

Each experimental assessment and analysis challenged students to examine both the basic principles of the scientific method in human performance as well as features unique to each type of assessment. In their reports, drafted by the principle investigator and reviewed by team members and revised for final submission, there were many fruitful areas for literature research and discussion. The teams considered issues such as test purpose and practicality, intra- and intertester reliability, gender differences, role of motivation, subject compliance with pretest and test protocols, measurement and calculation error, and a host of other factors that can influence test results. Students noted that test validity may vary in relation to their work within a larger investigative context, to improve the statistical power of their analyses, to learn to manipulate a large database, and to practice essential scientific skills of hypothesis testing, data collection and analysis, collaborative work, and scientific communication.

Fig. 2. Mean percent body fat values for college-age subjects comparing athletes and nonathletes. Values were estimated using the BodPod air displacement plethysmography instrument. Values are means; error bars indicate SD. A: results for one class year ($n = 2$ athletes and 8 nonathletes). ANOVA indicated no significant differences between groups ($P = 0.8465$). B: results from 10 class years ($n = 25$ athletes and 60 nonathletes). *Significant difference between groups ($P = 0.0495$).

Table 2. Percent body fat measurements by BodPod by gender and athletic activity

<table>
<thead>
<tr>
<th></th>
<th>$n$</th>
<th>Mean (SD)</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Single year’s class</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Athletes</td>
<td>2</td>
<td>24.1 (2.9)</td>
<td>–2.0, 50.1</td>
</tr>
<tr>
<td>Nonathletes</td>
<td>8</td>
<td>22.5 (10.4)</td>
<td>13.8, 31.2</td>
</tr>
<tr>
<td><strong>Multyear database</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Athlete</td>
<td>25</td>
<td>22.5 (8.1)</td>
<td>19.2, 25.8</td>
</tr>
<tr>
<td>Nonathlete</td>
<td>60</td>
<td>26.8 (9.4)</td>
<td>24.4, 29.2</td>
</tr>
<tr>
<td><strong>Athletic activity by gender</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female athletes</td>
<td>16</td>
<td>27.1 (4.2)</td>
<td>24.9, 29.3</td>
</tr>
<tr>
<td>Female nonathletes</td>
<td>39</td>
<td>31.2 (7.0)</td>
<td>29.0, 33.4</td>
</tr>
<tr>
<td>Male athletes</td>
<td>9</td>
<td>14.3 (6.7)</td>
<td>9.2, 19.4</td>
</tr>
<tr>
<td>Male nonathletes</td>
<td>21</td>
<td>18.7 (8.1)</td>
<td>15.1, 22.4</td>
</tr>
</tbody>
</table>
Table 3. Student assessment of their learning gain results over time

<table>
<thead>
<tr>
<th>Year</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (SD)</td>
<td>Mode</td>
<td>Percent 4/5</td>
<td>Mean (SD)</td>
<td>Mode</td>
</tr>
<tr>
<td>Writing laboratory reports</td>
<td>3.7 (1.3)</td>
<td>4</td>
<td>75</td>
<td>4.3 (0.51)</td>
</tr>
<tr>
<td>Performing/interpreting exercise tests</td>
<td>4.1 (1.0)</td>
<td>4</td>
<td>87</td>
<td>4.1 (0.33)</td>
</tr>
<tr>
<td>Finding trends in data</td>
<td>3.9 (1.0)</td>
<td>4</td>
<td>75</td>
<td>4.1 (0.61)</td>
</tr>
<tr>
<td>Reviewing laboratory reports</td>
<td>3.7 (0.71)</td>
<td>4</td>
<td>62</td>
<td>4.1 (0.61)</td>
</tr>
<tr>
<td>Working effectively with others</td>
<td>3.5 (0.53)</td>
<td>3</td>
<td>50</td>
<td>3.7 (0.71)</td>
</tr>
</tbody>
</table>

n = 8 respondents in 2003, 9 respondents in 2004, 14 respondents in 2006, and 27 respondents in 2006. 3 = Somewhat, 4 = a lot, and 5 = a great deal.

The testing modality, the estimation equation used, and the population studied under the conditions of testing.

For example, students predicted that skin-fold values would compare most closely with BodPod values, whereas others predicted better agreement with BIA estimates. In fact, there were no significant differences between mean BodPod and BIA values when the complete database was analyzed, whereas skin-fold values were significantly lower (Fig. 1A). These results, in contrast to the results generated with only a single year of data, are consistent with the literature (16, 17). Thus, different and potentially more accurate conclusions can be reached when using a multiyear database to test hypotheses. It is also apparent that subgroup comparisons (male/female and athlete/nonathlete) are only possible with the multiyear database (Table 2). Students also found it difficult to form testable hypotheses and find statistical significance when using $V\dot{O}_{2\text{max}}$ and power assessment data from single laboratory sessions. Similarly, these problems are largely overcome by using power and $V\dot{O}_{2\text{max}}$ data from the multiyear database.

Students noted the loss of statistical power in their comparisons, and the different conclusions made, of body composition values from their class alone (Fig. 1A) compared with comparisons of larger numbers of subjects that they had done when analyzing the entire database (Fig. 1B and Table 2). These observations allowed students to truly appreciate that statistical power, although affected by the chosen α-level, inherent variability, and the value of the tested parameter under the alternate hypothesis, is significantly influenced by sample size. Moreover, these examples emphasized the importance, when designing experiments in the research setting, of performing power estimations to determine the number of subjects required to detect significant differences.

Advantages. This laboratory approach offers several general advantages. All class members participate in the research experience and the research questions and hypotheses are student driven. The research questions address real-world applications relevant to student interests and future careers. Therefore, the larger aggregate database provides improved statistical significance and allows a cross-sectional and longitudinal characterization of the student subpopulation for a comparison with population norms. This can provide information for targeted interventions with this population when health/fitness norms are not met. Students clearly observed that when using their own class data of 10–12 subjects, differences predicted by their hypotheses rarely achieved significance, and they then appreciated the improvement in statistical power provided by a large database. In addition, each individual student receives a complete fitness evaluation while mastering the skills for the performance of evaluations on clients. Students gain experience with group collaboration and functioning as members of a research team. They also gain experience with scientific communication, including writing, peer review, and revision as well as the use of technology, including a graphing package, spreadsheets, statistics, and Blackboard. It is especially advantageous to have complete updated data always available to all students on the Blackboard site to be downloaded, shared, manipulated, and analyzed at any time and place with internet access.

Limitations. There are, however, some limitations to this laboratory approach. With multiple investigators initially inexperienced with the techniques, inter- and intratester reliability may not be high. While time and the laboratory schedule did not permit the collection of replicate measures, reliability experiments with this group in a previous class for right calf skin-fold thickness, students showed a Pearson correlation of $r = 0.723 (0.256)$ [mean (SD), $n = 62$] for test-retest reliability and a mean correlation of $r = 0.570 (0.293)$ [mean (SD), $n = 38$] for intertester comparison. This design would be unsuited for a laboratory research study, but it closely resembles the practical situation seen in health/fitness/sport facilities, where multiple testers evaluate clients over extended time periods. It would be instructive to devote an additional laboratory week to repeating the body composition measures to reinforce the importance of reliability control. Due to the lack of time and equipment, field estimates of $V\dot{O}_{2\text{max}}$, power, and body composition cannot be validated directly against their corresponding criterion tests for all students (open circuit spirometry,
Wingate anaerobic power test, and hydrodensitometry); however, numerous validity studies of these methods have been reported in the literature (16, 17, 29). Additionally, the student population in this class may not resemble the norm population, since this sample is clustered at the lowest point of the age range (male mean age = 21.0 yr and female mean age = 20.4 yr) of the comparison group (age: 20–29 yr) and contains intercollegiate athletes as well as average students. It is important for students to understand that field tests are validated relative to criterion tests in specific populations and that norms are generally highly specific to the test population. As the database grows and individual class sizes are larger, it will offer the potential to examine longitudinal trends in fitness over several years.

The Salgains assessment and traditional course evaluation indicate a positive perception of the value of this laboratory approach. It does, however, only address student perceptions of their learning gains. A growing body of evidence supports the usefulness, validity, and reliability of the Salgains evaluation in the improvement of teaching and learning, and it has been adopted by national organizations such as Science Education for New Civic Engagements and Responsibilities in Science, Technology, Engineering, and Mathematics course reform efforts (34–36). Laboratory grades based on short and long Technology, Engineering, and Mathematics course reform efforts (34–36). Laboratory grades based on short and long-term fitness testing is emphasized to reinforce the concept that exercise physiology is an applied scientifically based discipline.

A number of reports have described the advantages of student-designed experiments in physiology laboratories. (8, 11, 12, 20, 31). In large introductory biology laboratories with Physiology majors, modular experiments designed by student research teams were performed using collaborative writing (21). Analysis indicates a higher level of positive student opinion about the experience and better performance on the assessment of content knowledge among students in investigative laboratories compared with those in traditional ones. In an organismal physiology course, students’ abilities to correctly answer increasingly complex questions concerning a mock experimental scenario were assessed (24). Students who participated in the laboratory centered on student-designed experiments performed better than those who took lecture only.

Another approach, similar to that described here, uses hypothesis formation and prediction before more standard experiments (23). After introductory remarks and the examination of a written protocol by students, predictions of the effect of cycle ergometer exercise on respiratory variables were elicited from class members. The predicted results were compared with the actual outcome of the experiments. Assessment indicated improvement in the realignment of students’ faulty mental models of respiratory processes (as assessed by pre-/posttesting), particularly when instructor verification and the postlaboratory wrap-up were included (23). Another study described prior hypothesis formation, focusing on the control of confounding variables and subsequent experimental testing in the biomechanics of human walking. The value of results that contradict the original hypothesis in forcing students to confront incorrect assumptions underlying their original predictions was noted. Student assessment indicated that the exercise was helpful in increasing their understanding of various aspects of the scientific method (6).

In an introductory population biology course, data were pooled among multiple sections of large laboratory courses for both major and nonmajor students, as a means of improving statistical power in experiments involving student-generated hypotheses in the areas of heritability and plant competition. The majority of student survey responses found such an approach helpful, consistent with the findings reported here (19).

In a study in exercise physiology similar to our report, students performed laboratory-based measures of cardiorespiratory endurance (i.e., $V_{O2max}$) and compared them with measures of human performance (20-m shuttle runs) in student subjects (31). Data collected with similar methodology, equipment, and staff over a 3-yr period were pooled for analysis. The goal was to enhance learning of fundamental exercise physiology concepts by having students collect and analyze their own experimental data and thus develop their analytic, problem-solving, and evaluative skills. The authors concluded that the learning of fundamental concepts was enhanced by the use of student-driven investigation using their own data. It is clear in this study that a larger pooled sample allowed students to demonstrate statistically significant relationships within the data.

Another study describes a project in which two groups of students, in a physiology class and in a basic statistics class, collaborated in the analysis of physiological data (heart rate, blood pressure, etc.) collected on student subjects by student investigators (30). The authors performed assessments of both attitudes and cognitive mastery of content knowledge before and after the course. They conclude that this student-centered, inquiry-based collaboration aided student recognition of the value of statistics in data analysis, helped students to use the appropriate methods and interpret statistics, and improved student attitudes concerning the value of mathematical analysis. The authors particularly identify the use of meaningful real-world open-ended data with no known answer as a major contributor to the positive outcomes of the project goals.

In summary, the use of a cumulative database, with information from assessments performed repeatedly over several years, provides opportunities for students to have an additional
experience exploring investigative questions and issues specific to human performance. Such experiences can reinforce and extend the investigative skills gained in previous courses in a variety of experimental systems and offer an active, interesting, and challenging experience for students. These experiences address directly the goals articulated by the Association of American Colleges and Universities (5): that education “provide multiple opportunities for students to engage in ‘inquiry-based learning,’ both independently and in collaborative teams…learn how to find and evaluate evidence, how to consider and assess competing interpretations, how to solve problems, and how to communicate persuasively.”

ACKNOWLEDGMENTS
The authors thank the students and faculty members of the Department of Health Sciences of Merrimack College for the effort and support in on-going curriculum revision and reform.

REFERENCES