Article

Teaching Students How to Study: A Workshop on Information Processing and Self-Testing Helps Students Learn

Kathrin F. Stanger-Hall,* Floyd W. Shockley,[†,‡] and Rachel E. Wilson[§]

*Department of Plant Biology, [†]Department of Entomology, and [§]Mathematics and Science Education, University of Georgia, Athens, GA 30602

We implemented a "how to study" workshop for small groups of students (6–12) for $N = 93$ consenting students, randomly assigned from a large introductory biology class. The goal of this workshop was to teach students self-regulating techniques with visualization-based exercises as a foundation for learning and critical thinking in two areas: information processing and self-testing. During the workshop, students worked individually or in groups and received immediate feedback on their progress. Here, we describe two individual workshop exercises, report their immediate results, describe students' reactions (based on the workshop instructors' experience and student feedback), and report student performance on workshop-related questions on the final exam. Students rated the workshop activities highly and performed significantly better on workshop-related final exam questions than the control groups. This was the case for both lower- and higher-order thinking questions. Student achievement (i.e., grade point average) was significantly correlated with overall final exam performance but not with workshop outcomes. This long-term (10 wk) retention of a self-testing effect across question levels and student achievement is a promising endorsement for future large-scale implementation and further evaluation of this "how to study" workshop as a study support for introductory biology (and other science) students.

INTRODUCTION

In recent years, many studies have documented a lack of proficient critical thinking skills in college students (Holschuh, 2000; Weimer, 2002; Lord and Baviskar, 2007; Crowe *et al.*, 2008; Lord, 2008; Stanger-Hall *et al.*, 2010). Critical thinking, defined here as higher-order thinking skills or levels 3–6 of Bloom's taxonomy of cognition (Bloom, 1956), is generally viewed as an essential part of college training (Boyer Commission, 1998; National Research Council, 2003). Unfortunately, many students fail to understand the importance of higher-order thinking skills (Stanger-Hall, unpublished data) and consequently struggle to develop these skills.

For instructors of large introductory college lecture classes, it is often difficult to assess whether students practice critical thinking when they study for class, and failure to practice these skills is recognized by the student and the instructor only when students fail to do well on exams that assess critical thinking (application, analysis, evaluation, and synthesis) skills. This is an issue especially during the first year of college when students discover that study techniques that have been successful in high school may not necessarily be successful in college (Matt *et al.*, 1991; Yip and Chung, 2005). More motivated students then find their way to the office of the instructor, where their study routine can be assessed and suggestions for modification of their study habits can be made and practiced. Teaching these skills to individual

students during office hours, however, is not a very efficient means by which to teach students new study skills, especially in large classes. As a result, we decided to develop (K.S.-H.) and implement (F.W.S.) a "how to study" workshop for students in a large introductory biology lecture class.

There is a growing body of literature on study techniques for college students (e.g. Paulk, 2000; Nist-Olejnik and Holschuh, 2008; van Blerkom, 2008), and online study advice is now available from many colleges and universities. Topics range from more mechanistic advice (note-taking, study scheduling) to processing skills (reading strategy, organizing information, creating exam questions, flowcharts, question design) to metacognitive skills (self-reflection on learning goals and adjustment of study time and techniques based on learning outcomes). What exactly constitutes "good" study behaviors and which behaviors may be detrimental because they use up study time at the expense of more effective study techniques (Gurung *et al.*, 2010) likely depend on both the specific learning goals for an individual class and the knowledge and skill levels taught and assessed by the instructor.

In the introductory biology class in this study, critical thinking was emphasized and Bloom's taxonomy of cognition (Bloom, 1956) was taught to students as a communication and study tool during the first week of class. At this time, the instructor emphasized to students that 25–30% of the questions on each exam would be asked at Bloom levels 3–5, assessing application, analysis, and evaluation skills. Because the exam format is limited to multiple choice, level 6 (creation/synthesis) is generally not assessed. As a result, students who desire to earn a grade of "C" or higher must master these critical thinking skills. It was apparent to the instructor during office hours, however, that students found it difficult to distinguish what they did know and what they did not know from the class material (Tobias and Everson, 2002), and generally most overestimated their knowledge and thinking skills (Isaacson and Fujita, 2006). In other words, struggling students tended to lack metacognitive skills and study strategies that would have allowed them to self-monitor their knowledge and to adjust their studying and learning outcomes to better achieve their learning goals for the class, a skill set that tends to be a trademark of higher-achieving students (Isaacson and Fujita, 2006). Furthermore, students tended to study facts in isolation rather than putting them in the context of their current knowledge: They were not using contextual thinking to improve their recall or as a basis for critical thinking (reasoning through connections to identify mistakes or misconceptions). As a result, we decided that a "how to study" workshop would demonstrate the benefits of these skills to our introductory biology students.

Visualization (in the form of mental images or as external representations) is important for all learning, but especially for learning in the sciences (Mathewson, 1999; Gilbert, 2005, 2008; Schönborn and Anderson, 2006). Furthermore, visualization relies on context and thus promotes contextual learning. Therefore, we specifically selected two different exercises as workshop activities demonstrating to students how visualization techniques can be used for contextual thinking and recall, for encouraging feedback on existing and missing knowledge and understanding, and as an opportunity for developing cognitively active learning and for practicing critical thinking skills.

EXERCISE 1. INFORMATION-PROCESSING EXERCISE: VISUAL VERSUS AUDITORY PROCESSING

Many college instructors have found themselves at some point teaching a large class and facing a sea of students looking down at their notes, trying to write down word-for-word *what* the instructor says, rather than paying attention to *why* it is being said (why it is important, what the context is, etc.). Most students taking notes during lecture tend to process information in an auditory manner by listening and simply recording what they hear. Students who process information visually (imagine what they hear), however, usually outperform students who process information in an auditory manner in terms of their recall ability (Revak and Porter, 2001). Visualization in the form of creating mental images is a fundamental cognitive process that has been shown to improve student learning (Pressley, 1976), particularly in the sciences (Wu and Shah, 2004). We decided to demonstrate this to our students through personal experience.

EXERCISE 2. SELF-TESTING EXERCISE: VISUALIZATION OF EXISTING AND MISSING KNOWLEDGE AND UNDERSTANDING

Many introductory biology students tend to limit their study activities to reviews of their lecture notes and textbook, aimed at memorization of facts and explanations (Karpicke *et al.*, 2009; Stanger-Hall, unpublished data). They are generally unaware of self-testing as a learning strategy other than using old exams or index cards as a memorization tool. Instead, students tend to believe that, once they can recall an item, they have learned it (Karpicke, 2009). Students tend to neglect practicing information retrieval (self-testing) when studying on their own (Karpicke, 2009), but research has shown that testing inserted into the learning phase enhances long-term retention (Agarwal *et al.*, 2008). To illustrate the importance of self-testing for student learning, we used a self-testing exercise as the second workshop activity. This self-testing exercise used drawing as a visualization tool (Gobert and Clement, 1999) to demonstrate to students what they remembered from class and what they didn't.

Previous studies have shown that visualization of scientific principles through the use of imagery and external representations (diagrams, models) improves student conceptual understanding and ultimately student performance on assessments. We predicted that we would see such effects in our workshop students as well.

GENERAL METHODS

The prescribed "how-to-study" workshop was implemented in Fall 2009 for 99 students, as part of a larger study on the effects of various study supports on the use of critical thinking skills in introductory biology students. This study was conducted in a large introductory biology class ($N = 300$

students). It was the goal of this larger study to test three different study supports as candidates for future large-scale implementation(s) in large introductory biology classes. The "how to study" workshops constituted one of these three study supports. Students were assigned to the three different study support groups based on their Exam 1 performance: exam scores were sorted in descending order and students were assigned to the three groups in rotating order: 1–2–3–1–2–3, etc. (top score: Group 1, second score: Group 2, third score: Group 3, fourth score: Group 1, etc.). Each of these three groups received a specific study support and group-specific assignments, which were part of the class grade (40 points of 1100). We generated the control groups for the three treatment groups by applying the same methodology to the same class from the year before (Fall 2008), which was taught by the same instructor (K.S.-H.) and received no treatments. This approach had the advantage of generating control groups of similar size to the treatment groups and of compensating for any possible bias generated by the 1–2–3 sequence of assigning students to groups.

We report here on the learning outcomes for the workshop group, whose study support consisted of two 90-min "how to study" workshops that each were offered during multiple (up to 20) time-slots over a period of 8 d. Workshop I was offered during the week following Exam 1, whereas Workshop II was offered during the week following Exam 2. Students signed up for their most desired time slot (maximum of 12 students) but were asked to choose another time if fewer than six students signed up. This study reports the outcomes of the first workshop in the series, which focused on the use of visualization for information processing and self-testing. There was no overlap between the workshop activities and the activities of another treatment group in Fall 2009; therefore, this treatment group served as an additional (same-semester) control group for this analysis.

The first workshop consisted of two visualization-based exercises, which demonstrated 1) different approaches to information processing (as it applies to class and reading) and their effects on remembering information, and 2) how to use self-testing as a routine study tool for remembering and critical thinking, defined as the upper four cognitive levels of Bloom's taxonomy: application, analysis, evaluation, and synthesis (Bloom, 1956). The desired benefits of these exercises were 1) to help students learn by reviewing the class material actively: This is expected to inform them on what they know and do not know as well as improve both their lower (remembering facts and explanations) and their higher (critical) thinking skills, and 2) to help students identify misconceptions by placing what they remember into the context of other knowledge from previous classes or lectures and using their thinking skills to detect possible discrepancies.

We used assessment of student learning on the cumulative final exam to answer the following questions: 1) Does self-testing lead to learning gains for the topic used for self-testing? If so: 2) Can learning gains be documented for both lower- and higher-order thinking skills? And 3) Do students across achievement levels (as measured by grade point average [GPA]) benefit? To qualify for future large-scale implementation as a study support in introductory biology (and other science) classes, affirmative responses to all questions were required.

Workshop Implementation

After the first exam in Fall 2009, 99 students were assigned to the workshop group; however, only 93 students consented to participate in this study. All 93 students attended the workshop and earned 10 points (of 1100) of class credit. As a result, the evaluation of workshop activities is based on a sample size of 93 students. Students could earn additional points by filling out an online feedback survey on their workshop experience, and 79 of the 93 students elected to do so (sample size for feedback survey: $N = 79$). Eleven of the 93 consenting workshop students withdrew from the class by the midpoint of the semester, leaving $N = 82$ students for analysis of their final exam performance (sample size for final exam analysis: $N = 82$). In comparison, the same-semester control group (control 2009) had 90 consenting members, and the previous-year control group (control 2008) had 87 consenting members at the end of the semester (control samples for the final exam analysis).

The workshop students met in small groups (a maximum of 12 students per workshop) outside of the regularly scheduled class time in a small conference room to better enhance instructor–student and student–student interactions. All workshop sessions were taught by the same workshop instructor (F.W.S.). Using a workshop format rather than incorporating these activities into existing lecture classes (where the total number of students can easily exceed 300 per class at many large public universities) is not only more practical but also more personal, and better promotes follow-up discussions on how implementation of the workshop activities can directly impact student learning.

Each workshop session was scheduled for 75 min (the two workshop activities in the first workshop collectively took between 60 and 75 min to complete). We found it useful to break the workshop up into two discrete blocks, with each activity separated by short instructor-led group discussions of Bloom's taxonomy: reminding students why this was introduced to them in the first week of lecture, how it can be applied to learning in science, and how self-testing helps improve understanding of complex material and can lead to critical thinking skills.

At the end of the workshop, students answered three basic questions about each of the two exercises in an online survey on the class website:

- How useful did you find the exercise?
- How useful was it for you to actually do the exercise (during the workshop), rather than just hearing about it?
- How likely are you to implement the workshop activity into your own learning after the workshop?

Students were asked to respond to each question by rating their opinion on a Likert scale from 1 (least: not useful at all, not at all likely to implement) to 5 (most: extremely useful, highly likely to implement).

DATA COLLECTION AND ANALYSIS

The immediate effect of the workshop activities on student recall and understanding was assessed during the workshop, mainly to demonstrate those effects to the students, but also

to allow the instructor to immediately assess the impact of the workshop activities. The long-term effect (10 wk) of the workshop on student learning was assessed via workshop-related questions on the cumulative final exam. Final exams were not returned to the students so students in subsequent semesters could not benefit from memorizing old exams. If the workshop was effective, we predicted that workshop participants would perform better on final exam questions related to the workshop topic than students who did not participate in the workshop (F2009 and F2008 controls). We compared the distributions of correct–incorrect answers for the individual multiple-choice exam questions, the total correct answers, the total correct lower-level answers (Bloom levels 1 and 2), and the total correct higher-level answers (Bloom levels 3 and 4) between the workshop and control groups. To control for other potential influences, such as pre-existing student achievement, we compared self-reported student GPA (at the beginning of the semester), as well as overall final exam performance (all questions) between workshop and control groups. Under the Null hypothesis (that overall student achievement did not differ between control and treatment groups), we predicted no significant differences between the workshop and control groups in GPA and total final exam scores.

For each group, we tested all variables for normality (Goodness of Fit: Shapiro Wilkes Test) using JMP 8 software (SAS Institute, Cary, NC). We used SPSS 18.0 for Mac software (2010; SPSS, Chicago, IL) for quantitative statistical analyses. Only the final exam scores were normally distributed with homogeneous variances between groups. As a result, we report the results of nonparametric tests for all analyses. For the student performance data (e.g., overall performance on exam questions relating to the workshop, the total final exam scores, as well as start-of-semester GPA [pre-existing student achievement]), we used nonparametric Mann-Whitney U-tests for independent samples. This is a test for both location and shape to test for differences between distributions of ranked variables. To test whether the performance of workshop and control groups on individual exam questions was the same (null hypothesis) or different (alternative hypothesis: the workshop helped students learn), we used a Pearson χ^2 test. The data from the respective control groups were used to calculate the expected values for the workshop group. To correct for multiple comparisons (inflated Type I error) we applied a false discovery rate (FDR) correction (Benjamini and Hochberg, 1995) and report the adjusted P values. For paired samples (pre–post comparisons) of workshop exercises, we used the Wilcoxon-signed rank test, and for correlation tests (e.g., GPA and performance on exam questions) we used Spearman correlations. With the exception of the χ^2 tests (alternative hypothesis: workshop students perform better: one-tailed test), all reported results are based on two-tailed tests and significance levels of $P < 0.05$.

EXERCISE 1: INFORMATION-PROCESSING EXERCISE—VISUAL VERSUS AUDITORY PROCESSING

Visualization in the form of creating mental images is a fundamental cognitive process that helps student learning. We decided to demonstrate this to our students through a personal experience. For this purpose, we chose the Slippery

Snakes exercise, which was developed in 1993 by Don Irwin and Janet Simons from the Development Educational Learning Institute (Des Moines, IA) to illustrate the differences between visual- and auditory-encoded memory (Irwin and Simons, 1993; Bolt, 1996). We used this exercise to demonstrate different approaches to information processing while listening and the potential impacts that the different approaches can have on student recall ability. We hoped that, by learning how to engage their visual memory instead of their auditory memory alone when taking notes in class, students would not only improve their ability to remember facts, but also to contextualize information, a prerequisite for critical thinking.

For this exercise, we divided the students at the beginning of the workshop randomly and evenly into two groups. Each group was provided with a different set of written instructions on what they were supposed to do as each phrase was read aloud. One group (the visual-processing group) was given the task of trying to form a vivid mental picture or image of the action in each phrase, and rate each phrase (on a scale of 1–10) on how simple or difficult it was to visualize. The other group (the auditory-processing group) was given the task of listening to each phrase with an emphasis on pronunciation and to rate each phrase on how simple or difficult it would be to pronounce. The students were not aware that there was a difference in instructions. Each sheet of instructions had the same numbered blanks for the students to write down their ranks as each phrase was read.

The workshop instructor read the phrases aloud slowly and deliberately, one after the other. Once all the phrases had been read, the instructor gave the students a quick unannounced quiz on the content of the phrases they had just heard. The students wrote down their answers—the subject of the phrase and an associated adjective—on the back of their instruction sheets, as the questions were read aloud at the same pace as the original phrases. After the quiz, the students were asked to score their own answers as the instructor read the correct answers. The original exercise consists of 20 phrases (Irwin and Simons, 1993), as well as questions (and answers) about those phrases. For time management reasons, we only used 12 of these.

After reassuring students that memory does not equal intelligence, following the instructions by Irwin and Simons (1993), students reported their scores individually to the instructor, who recorded them on the board in separate columns for the two groups. The differences between the two groups were immediately obvious (Figure 1), with the students in the visual-processing group (mean $\pm$ SD $= 10.23 \pm 1.32$) scoring significantly higher than those in the auditory-processing group (mean $\pm$ SD $= 5.4 \pm 2.12$; Mann-Whitney U-test for independent samples: $U = 47, P < 0.001$). After looking at the reported scores on the board and averaging for each group, the instructor asked one student from each group to reveal their set of instructions by reading aloud their respective group's instructions. The students were generally surprised by the revelation of the difference in instructions and the improved performance of students using the visualization strategy for the processing of complex information.

Owing to these striking results, we decided to control for the possibility of a bias in the higher achieving visualization group by analyzing start-of-semester GPA (pre-existing student achievement) and student preferences for the
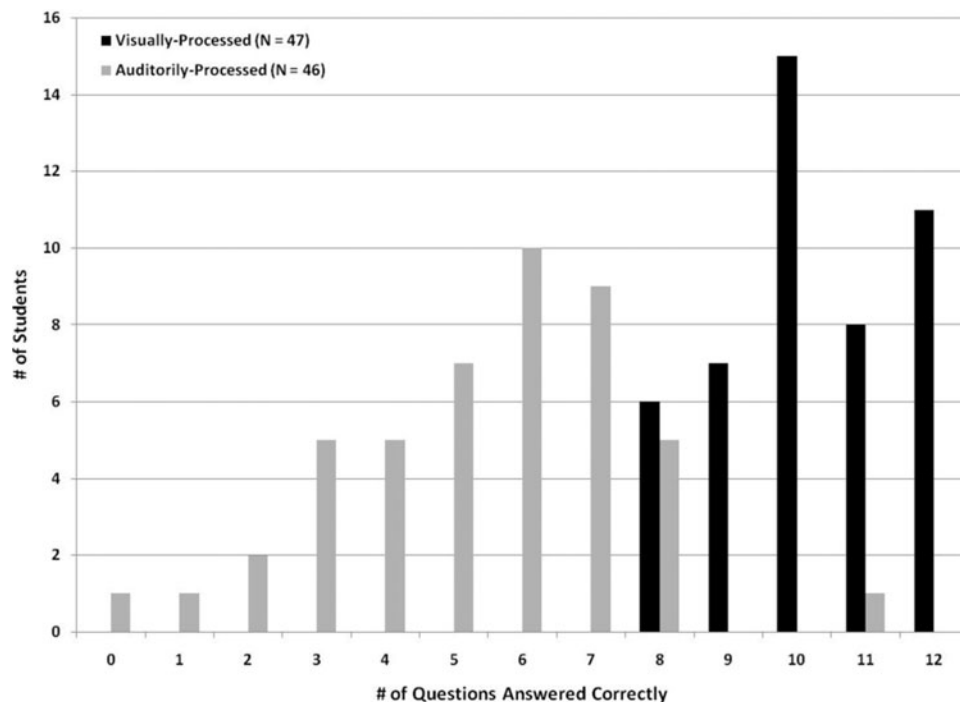
Figure 1. Information-processing exercise. Student performance by processing instructions (auditory vs. visual). Students who processed information visually scored significantly higher on the recall test ($N = 12$ items) than did students who processed information auditorily only (visual mean $\pm$ SD $= 10.23 \pm 1.32$; auditory mean $\pm$ SD $= 5.4 \pm 2.12$; Mann-Whitney U-test for independent samples: $U = 47$, $P < 0.001$).

information exchange medium (e.g. visual, acoustic). There was no difference in GPA between the visual- and auditory-processing groups (Mann-Whitney $U = 656.5$, $P = 0.494$). To quantify student preferences for information exchange media, we used the online VARK survey (version 7.0, 2006) with permission from Neil D. Fleming, Christchurch, New Zealand, and Charles C. Bonwell, Springfield, MO (Fleming and Mills, 1992). This survey consists of 16 questions asking students about their preferred information exchange medium in everyday situations (multiple answers possible). The output scores are visual (V), auditory (A), reading (R), and kinesthetic (K) scores (total instances that medium was chosen). Please note that the VARK survey is advertised as a "learning style" assessment (for a review of the vast amount of literature on "learning styles," see Coffield *et al.*, 2009); however, we used the VARK as a tool to determine a student's current preferred medium for information transfer (regardless of which influences may have contributed to this preference). After calculating the relative contribution of each medium to the preferred information processing of each student, we tested whether the students in the auditory- and the visual-processing groups of the listening exercise differed in their preferred information-processing medium. We found no significant differences between the two groups (Visual $U = 541.5$, $P = 0.659$; Auditory $U = 603.5$, $P = 0.749$; Reading $U = 512$, $P = 0.421$; Kinesthetic $U = 690$, $P = 0.167$). These findings emphasize that the significant differences in retention between the two groups were not due to differences in pre-existing student achievement (GPA) or prior information-processing preferences, but most likely due to the different processing instructions given to the two groups of students.

Although this exercise in itself was a valuable learning experience for our students, we found it useful to relate the value of this experience to student learning in class. For example, by visualizing the information they hear in class, students can incorporate context into the processing of this information and into their lecture notes. This strategy also works well for reading assigned textbook material. Whereas textbooks tend to emphasize isolated terms (in boldface type) over context, by visualizing the textbook material students can focus on context and higher-level processing. At the end of the workshop, most students rated this exercise as very useful (Table 1), especially doing the exercise rather than just hearing about it. Most students planned to implement what they had learned into their own information processing and note-taking during class (mean $\pm$ SD: 4.11 ± 0.9 on a scale from 1 to 5), and to a lesser extent, during their textbook reading (mean $\pm$ SD: 3.75 ± 1.1).

EXERCISE 2: SELF-TESTING EXERCISE— VISUALIZING EXISTING AND MISSING KNOWLEDGE AND UNDERSTANDING

Many students in introductory biology classes are generally unaware of self-testing as a learning strategy. As a continuation of the visualization theme, we used a visual representation of the generalized plant life cycle to illustrate the importance of self-testing for student learning. Life cycles were taught in lecture the week before the workshop sessions were conducted. During class, the course instructor (K.S.-H.) emphasized that, during all sexual life cycles, specific

Table 1. Student feedback

Exercise	Learning "How useful did you find this part of the workshop?"		Doing "How useful was it to actually do the exercise during the workshop?"		Implementing "How likely are you to implement this in your own learning after the workshop?"	
	Mean	SD	Mean	SD	Mean	SD
Information processing	4.20	0.85	4.38	0.87	4.11	0.91
Self-testing	3.90	1.03	4.04	1.04	3.93	0.98

Student responses to three survey questions regarding the two workshop exercises, based on a Likert scale from 1 (least) to 5 (most). On average, students found the workshop, including its practical aspects, very useful and reported that they were very likely to implement what they had learned in their learning.

structures—as defined by two characteristics: their cellularity and their ploidy-–are transformed into one another via three basic processes: mitosis (cell division that maintains ploidy), meiosis (cell division that reduces ploidy), and fertilization (cell fusion that increases ploidy). After this introduction, all three generalized life cycles, including the generalized plant life cycle (Figure 2), were developed and drawn on the overhead camera in a collaborative effort between students and the instructor. Visual representations of complex information during class are extremely useful because instructors and students alike can use them to incorporate a large amount of information into a simple schematic diagram that provides context for conceptual understanding and recall. Another advantage of this approach is its utility for practicing reasoning skills, checking logical connections and relationships between different pieces of information, thereby helping students construct a more comprehensive understanding. To demonstrate how to self-test in a productive way during studying, we conducted the following exercise with our workshop students.

At the beginning of the self-testing exercise, the workshop instructor handed out one blank note card (4″ × 6″) to each

Figure 2. Generalized plant life cycle used for the self-testing exercise. This diagram includes five different structures, each defined by name, cellularity and ploidy, and five different processes (mitosis is involved in the transformation of structures in three distinct instances during the plant life cycle).

student. Students were asked to draw as much of the generalized plant life cycle as they could remember on one side of the note card. Students were given as much time as they needed and were encouraged to recall as many structures and processes as possible. Once all students finished the self-test, the workshop instructor had the students turn the note cards over so they could not see their initial drawings and led a group discussion of the generalized plant life cycle that served to remind students of what they knew. Students volunteered the information, which structures (names) are part of the generalized plant life cycle, which structural characteristics are used to define them, and which processes are involved in transforming one structure into the next. The instructor generated a table from student responses on the whiteboard (Table 2). Students were not allowed to take notes during this review, rather they had to remain engaged in the discussion. If needed, the instructor assisted only by revealing the number of items in each category, asking students to recall what they had learned in class until the list was completed. Please note that this list contained only the names of structures and the structural characteristics and processes to be considered. There was no discussion about the details of the plant life cycle, specifically, how the individual structures were defined or what the processes did to those structures. At this point, the instructor commended the students on their brainstorming and reminded them that they themselves had generated
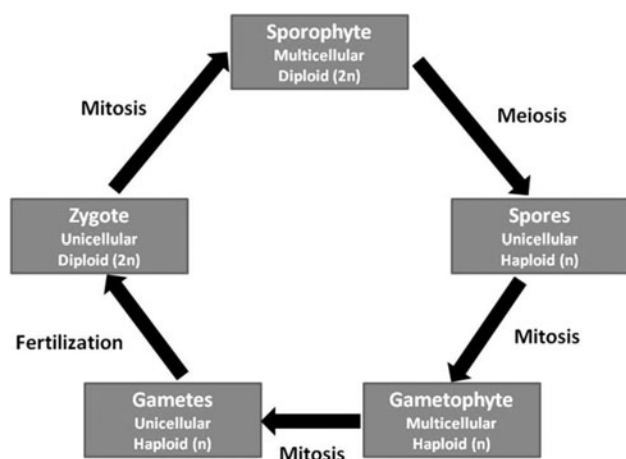
Table 2. Self-testing group review

Structure names	Structure characteristics	Processes
Sporophyte Gametophyte Spore Gamete Zygote	Cellularity (unicellular, multicellular) Ploidy (haploid, diploid)	Mitosis Meiosis Fertilization

In all workshop groups, students generated this overall list of structure names, structure characteristics, and processes at work in the general plant life cycle. This list was compiled on the white board by the workshop instructor as students generated this information. Only the terms, not their relationships, were listed during the review (students had to identify the specific characteristics of each structure and the processes that transformed one structure into another on their own during their postreview self-test).
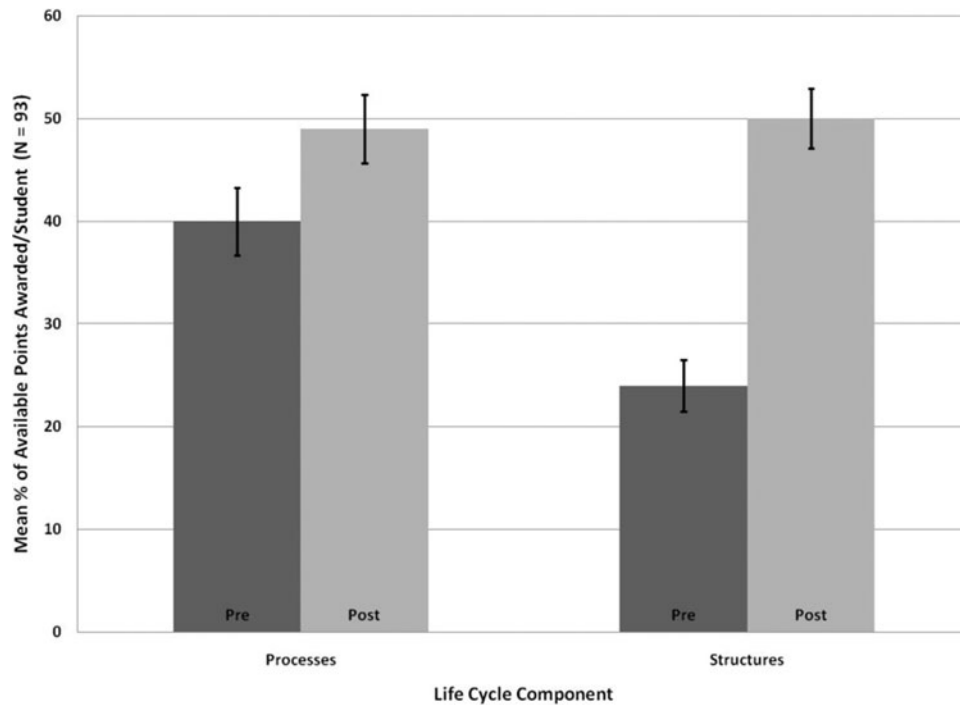
Figure 3. Pre- and postreview self-testing scores. Student performance ($N = 93$) is shown as mean (%) possible score (possible process score: 5 points; possible structure score: 15 points). Students performed significantly better in the postreview self-test than in the prereview self-test for both processes and structures (Wilcoxon signed rank test for paired samples (combined): $Z < -2.58$; $P < 0.01$; for processes only: $Z = -2.585$; $P = 0.01$; for structures only: $Z = -7.152$, $P < 0.001$).

this table, so as a group they clearly knew more than they initially thought they did, and that they should practice this approach while studying.

For the final step in this exercise, the list was removed from the whiteboard, and the students were asked to use the remaining blank side of the note card to diagram the plant life cycle again, using as many of the structures and processes as they could. At the end of the workshop, students were asked to identify which was their first (preview) and second (postreview) attempt and to hand in their note cards.

For the purpose of this study, we scored how well students did in their two attempts (this did not affect student grades). Their note cards were scored by assigning three points for each structure (one point each for the name and the two characteristics that define that structure) and one point for each correctly placed process for a total of 20 possible points (structures = 15 points; processes = 5 points). Both the preview and the postreview attempts were graded in the same manner for each student. To allow direct comparison of student performance on structures and processes, we transformed the data to percent of total points possible for each category (structures and processes).

Student Performance on Workshop Exercise

The life cycles produced during the first (preview) self-test were largely incomplete, and students were better able to re-call processes (mean $\pm$ SD $= 1.978 \pm 1.56$ of 5) than structures and their characteristics (3.66 ± 3.63 of 15). In contrast, during their second (postreview) attempt, students were able to produce a much more complete and accurate diagram of the

plant life cycle (2.43 ± 1.57 of 5 processes and 7.47 ± 4.21 of 15 characteristics; Wilcoxon signed rank test for paired samples: $Z < -2.58$; $P < 0.01$). Between the first and second attempt, the correct placement of structure names and their characteristics improved by 56% and 47%, respectively, whereas correct placement of processes only improved by 17% (Figure 3). This result is partly due to the fact that students struggled more to name the correct structures and their characteristics (of 15, 24% correct) on their first attempt, than they did in naming the processes (of 5, 40% correct). In their second attempt, students scored on average $\sim$50% correct for both structures (associating names with the correct cellularity and ploidy state) and processes (making sure that the process matched the change in structural characteristics). This represents a significant improvement for both categories (Figure 3) without looking up the complete life cycle.

Student Behavior during Self-Test

The workshop instructor observed students exhibiting "helpless" behavior during the first self-test: Students tended to give up at the first point in the cycle where they encountered a structure or process they could not recall, and did not attempt to start over from a different point and work the problem from there.

During their second self-test, the students who "gave up" during their first attempt were able to complete more of the life cycle, despite the fact that the placement of structures and processes and their logical connections were not practiced (or revealed) during the group review. When asked about the outcome of their second attempt after the exercise, the
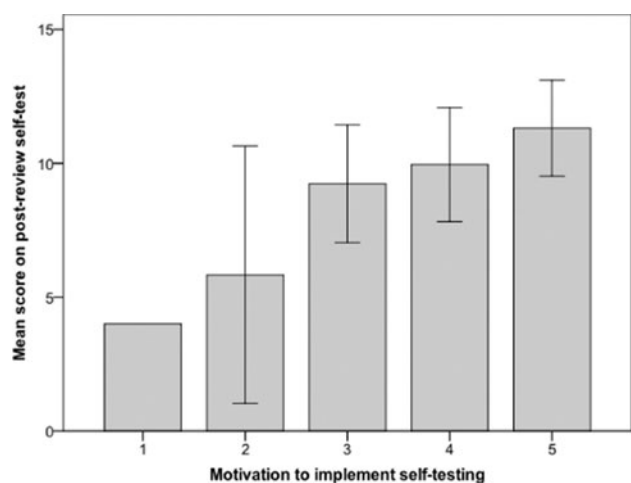
Figure 4. Student motivation to implement self-testing as a study tool. Students who reported a higher motivation to implement self-testing as a study tool after the exercise had performed better (points out of 20) during the postreview self-test than students who reported a lower motivation. For example, students with the highest level of motivation to implement self-testing (5: highly likely to implement, $N = 29$) scored 11 points (55%) in their postreview self-test; the student with the lowest level (not at all likely to implement, $N = 1$) scored 4 points (20%).

students generally attributed their higher success to the "help" by the instructor, even though they themselves generated the list of terms during the group review and applied the list to their second attempt, underestimating their own role in the process. We have found that this response is fairly typical for students not used to self-testing, including students who go through this exercise individually (Stanger-Hall, unpublished data).

Student Feedback

Many students found the self-testing exercise very useful (Table 1). Most students expressed intentions to implement what they had learned about self-testing into their regular study schedule. The motivation of students to implement self-testing as a study strategy was positively correlated with how well they did in their first (Spearman $\rho = 0.261$, $P = 0.02$) and second attempt (Spearman $\rho = 0.278$, $P = 0.013$) (Figure 4) to draw and label the plant life cycle.

Student Performance on the Final Exam

There were seven life cycle-related questions on the final exam in 2009 (14 points of 270), and four of these questions were also represented in 2008 (8 points of 270). Two of the questions (Questions 2 and 3) were higher-level application questions. We assessed overall performance on life cycle questions, performance on lower-level (Bloom levels 1 and 2) and higher-level (Bloom levels 3 and 4) life cycle questions, as well as on the individual life cycle questions separately. For comparison, we also assessed overall student performance on the final exam (all questions). To address the question of how the workshop affected students of different ability levels, we used two approaches: 1) We assessed whether GPA was correlated with performance on workshop-related questions on the final exam and how this correlation (if any) compared with the correlation between GPA and final exam performance overall; in addition, 2) we determined the median GPA for the workshop group (3.4) and divided the workshop students into two groups (low GPA = GPA < 3.4, and high GPA = GPA > 3.4; students with GPA = 3.4 [$N = 12$] were excluded from this analysis to create two distinct groups). We used the ratio of points earned for life cycle questions and total points earned for the other questions on the final exam to assess whether the potential workshop benefits (as assessed by the life cycle questions) differed between low- and high-GPA students.

If student achievement (GPA) did not influence workshop benefits (Null hypothesis), we expected to find 1) no significant correlation between GPA and exam performance and 2) no difference between the low-GPA and the high-GPA groups in how well the students did on the life cycle questions relative to all the other exam questions on the final exam.

Control 2009

The advantage of using this control group is that students were exposed to the exact same lectures (style, delivery, and examples used) as the workshop group. The disadvantage of this control group is that cross-talk between students (in the same class) may have occurred, that is, students may have learned from each other during study sessions. Despite this possibility, the workshop group scored significantly higher than the control group ($U = 4665.5$, $P = 0.001$) on the life cycle questions of the final exam (Figure 5A). Compared to the control group, the workshop students performed significantly better (all P values are reported after FDR correction for multiple comparisons) on Question 1 (lower-level: Pearson $\chi^2 = 3.739$, borderline significant at $P(1) < 0.05$), Question 2 (higher-level: Pearson $\chi^2 = 7.464$, $P(1) < 0.05$), and Question 7 (lower-level: Pearson $\chi^2 = 12.158$, $P(1) < 0.01$) of the life cycle question series on the final exam. In fact, workshop students tended to perform better than the control group on all life cycle questions (Figure 5B), but the differences for the other questions were not significant. The overall performance on the final exam was not significantly different ($U = 3723$, $P = 0.809$) between groups, and there was no significant difference in students' pre-existing achievement, as measured by self-reported GPA (workshop GPA = 3.353 ± 0.377, Fall 2009 control GPA = 3.283 ± 0.5; Mann-Whitney $U = 3041.5$, $P = 0.438$).

Control 2008

The advantage of using this control group is the absence of cross-talk between students, but the disadvantage is that the lectures (not content, but style and delivery) may have varied between the 2 yr, even though the classes were taught by the same instructor (K.S.-H.). Overall, workshop students scored significantly better on the life cycle questions of the final exam ($N = 4$ questions: $U = 4448$, $P = 0.004$) than the control group (Figure 5A). The workshop group performed significantly better on three of the four individual questions in the life cycle question series on the final exam (P values are reported after FDR correction for multiple comparisons): Question 1 (lower-level: Pearson $\chi^2 = 11.91$, $P(1) < 0.01$), Question 2 (higher-level: Pearson $\chi^2 = 16.089$, $P(1) < 0.01$), and Question 4 (lower-level: Pearson $\chi^2 = 5.149$, $P(1) < 0.05$).
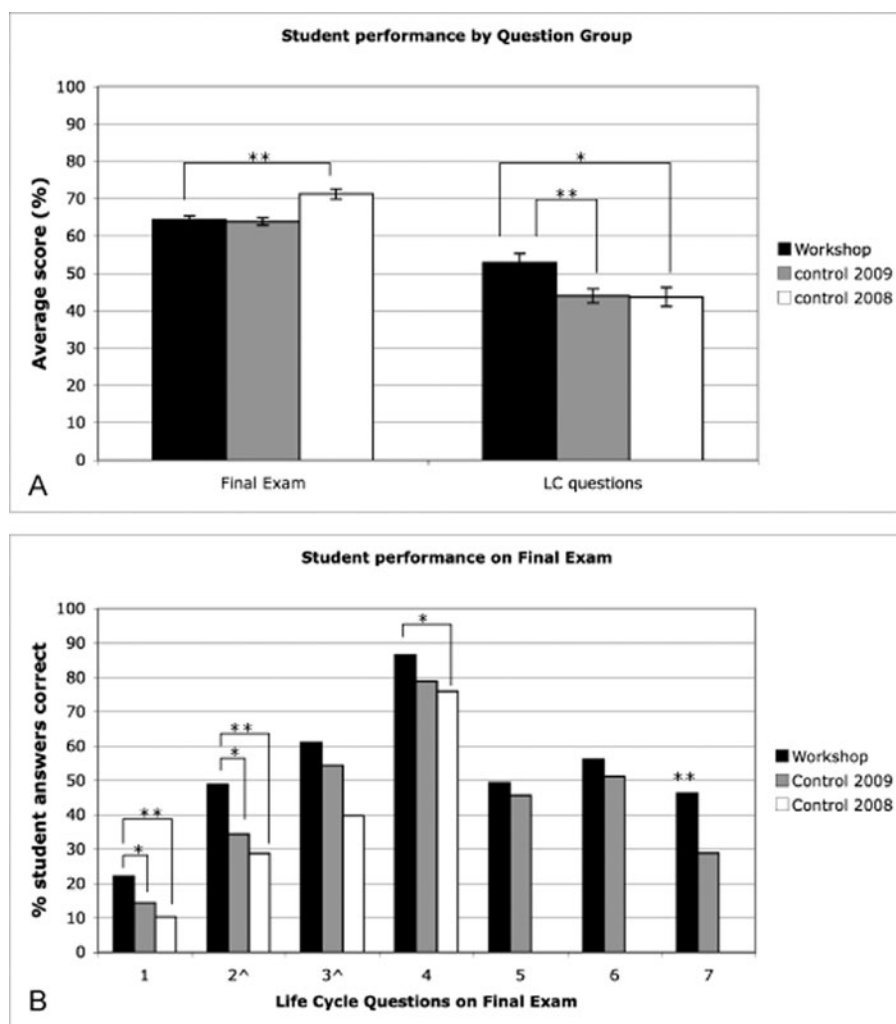
Figure 5. Student performance on life cycle questions on the final exam. Workshop students tended to perform better than students in the control groups. Significant differences between groups are noted as *, $P < 0.05$ or **, $P < 0.01$ (after FDR correction for multiple comparisons). (A) Overall student performance (mean ± standard error) on the final exam and on life cycle questions. (B) Individual life cycle questions on the final exam ($N = 7$ in 2009 and $N = 4$ in 2008): % students who answered correctly, ˆ = higher-level (application) questions.

Workshop students tended to perform better than the control group on all life cycle questions (Figure 5B), but the difference for Question 3 was not significant. The overall performance on the final exam was significantly different ($U = 2223, P < 0.001$) between the workshop and the control group, but the Fall 2008 control group performed significantly better than the Fall 2009 workshop group. This makes the significantly better performance of the workshop students on the life cycle questions even more relevant. There was no significant difference in students' pre-existing achievement (workshop GPA = 3.353 ± 0.37, Fall 2008 control GPA = 3.415 ± 0.34; Mann-Whitney $U = 3041.5, P = 0.438$).

Achievement Effects within the Workshop Group

There was no significant correlation between GPA and overall performance on the life cycle questions on the final exam within the workshop group (Spearman's $\rho = 0.219, P = 0.055$), but GPA was positively correlated with overall final exam performance (Spearman's $\rho = 0.552, P < 0.001$). In addition, low-GPA students did not gain significantly more or less from the self-testing exercise during the workshop than high-GPA students (relative to their performance on the other final exam questions; Mann-Whitney $U = 577, P = 0.383$). As a result, we can conclude that the self-testing exercise benefited students across achievement levels.

SAMPLES OF STUDENT FEEDBACK

Both workshop exercises were designed to help students learn, specifically to illustrate how visualization techniques can be used to improve information processing during class and reading, as well as for self-testing as a means to review and study. When students were asked, "What was the best or most useful part of the workshop?" the majority (52%) of respondents listed the information-processing exercise, 10% listed the self-testing exercise, and 36% had no preference (most of these liked both). The following are sample student responses to this question, demonstrating the wide variability in responses:

> "The self-testing part. I had never thought about how I would go about making up my own questions but now I know how."

> "Learning to visualize the concepts that we talk about in class, not just listen to them."

"The self test really made me see how much I did, and didn't know."

"It showed me the difference between the effectiveness of visual and auditory learning styles. Also, it showed me how to focus my learning on ways that will help me remember it in the long-run."

"I though [sic] participating in the demonstration was very helpful and allowed us to actually see the effectiveness of the process."

"Realizing that if I better visualize my notes as I am taking them/studying them, I can better retain and understand the information."

DISCUSSION

There is growing recognition that visualization is an essential thinking skill in science and science education (Mathewson, 1999; Gilbert, 2005, 2008; Schönborn and Anderson, 2006), and we have shown in this study that visualization in the form of mental images (during information processing) or external representations of complex information (during self-testing) can help college students learn. The next logical step is the large-scale implementation and further evaluation of this workshop for large introductory biology (or other science) classes, including development of more learning activities that allow students to practice their visualization skills during class and at home.

In the sciences, instructors and students alike can use visual representations to incorporate complex information into a simple schematic diagram that provides context for conceptual understanding and recall. Another advantage of creating visual representations is their utility for practicing reasoning skills, for example by checking logical connections and relationships between different pieces of information. This helps students construct a more holistic and comprehensive understanding, as compared with simply describing the relationships in writing (Gobert and Clement, 1999). Agarwal *et al.* (2008) showed that additional testing enhances long-term retention, and the self-testing approach described here would serve that purpose.

In general, during any self-testing exercise, students should write down all the information on a given topic that they can remember (notes and text closed) and organize it on paper (e.g., in the form of a diagram). After identifying possible gaps and missing pieces, students should then brainstorm what else they might know about this or closely related topics, making a list as they go. This process is the first step in identifying possible candidates for the missing pieces to fill gaps in their diagram, the goal of this exercise. The next step is to define everything on that list, and possibly compare–contrast similar terms, structures, or processes. This process further helps the students learn the reviewed material, recognizing possible connections to their diagram, and deducing some of the missing pieces and relationships—a strategy many students already use when solving Sudoku puzzles in the school newspaper before class but fail to employ during studying. When students engage in self-testing during studying, it is important that they do so without immediately going to their notes and textbook for help. Practiced as described, self-testing can be a very effective and cognitively

active form of learning (Karpicke, 2009; Karpicke *et al.*, 2009), a prerequisite for critical thinking that helps students review information while providing a context for this information at the same time. In contrast, looking up information or asking for answers without reflection only leads to passive memorization and helpless behavior in a testing situation.

It is noteworthy that the plant life cycle self-test in the workshop group took place during the week after students drew the life cycle (guided by the instructor) in class. There was a wide range in recall (0–20 of 20 points) between students during the first (preview) self-test, but on average students struggled to remember (5.46 of 20 points), and many students gave up rather than trying other approaches. This helpless behavior seems to be a characteristic response to a challenge by students without self-regulation skills, who tend to be low-achieving students (Isaacson and Fujita, 2006). Isaacson and Fujita (2006) suggested that such learned helplessness might be the consequence of high performance expectations that remain unadjusted by actual performance. By overestimating their abilities and not adjusting their study strategies, these students will decrease their efforts over time and ultimately give up and fail (Isaacson and Fujita, 2006).

The motivation of students to implement self-testing as a study strategy was positively correlated with how well they did in their first (preview) and second (postview) attempt to draw and label the plant LC. Difficult tests are better at differentiating between low and high achievers, and self-tests that also assess higher-level skills are generally more challenging than self-tests that only require lower-level skills (Isaacson and Fujita, 2006). This suggests that, although it is important to challenge students during a self-testing exercise, some success should be built in to convince students of the value of the exercise, and to move from helpless behavior to filling in as many pieces and connections as possible, even if skipping of steps is required. This can be best achieved by using recent material from class (as in this exercise) or by asking students to review their notes (or a textbook passage) on previously untested material. We recommend asking students whether the material "makes sense" to them before the start of a self-testing exercise. Students routinely will state that the notes (text) make sense but then struggle when asked to replicate the information on paper during a (closed-book) self-testing exercise. This apparent conflict between self-assessment and performance leads to the insight that "making sense" is not the same as understanding the material, which is a key step in convincing students to be receptive to making changes in their study routine.

The reviewing and self-testing skills are not only important for students during studying (Karpicke, 2009) but also during assessments (quizzes and exams). Rather than giving up immediately when not remembering an answer to a question, or when having to apply, analyze, synthesize, or evaluate the learned material to answer a higher-level question, these self-testing skills give students the opportunity during an exam (or in their everyday life) to figure out what they do not remember, or how to answer a critical thinking question. Because critical thinking is a life-long learning skill, active self-testing should be an integral part of studying and practicing for life-long learning, rather than immediately giving up and looking to others (notes, text, teacher) for answers when faced with a "hard" or "unfair" question. Unfortunately, the latter approach is far more common but promotes only

short-term memory, not long-term retention and higher-level thinking.

The benefit of this self-testing exercise for students was at least threefold: 1) it helped students review the class material actively, thereby helping them distinguish what they knew and what they did not; 2) it helped them place what they remembered in context, a prerequisite for identifying mistakes or misconceptions; and 3) it provided them with the tools to deduce missing pieces and to practice their critical thinking skills.

The written (documented) aspect of reviewing and self-testing is in our opinion crucial to achieve these outcomes. Although many students insist that they "do self-testing in their head" (Stanger-Hall, unpublished data), it is key for the success of this self-testing routine that a student documents her/his knowledge and knowledge gaps on paper (as an external visual representation) so they can be organized visually and their (existing and missing) connections (context) becomes apparent to the student. Students who self-test "in their head" usually just go through lists of terms and processes without recognizing their overall relationships and connections (Stanger Hall, personal observation), whereas students who document their self-testing on paper better recognize possible connections and can inspect them for flaws in logic afterwards.

STUDENT FEEDBACK

Overall, students perceived both workshop exercises positively but rated the information-processing exercise as more useful than the self-testing exercise. In addition, they reported that they were somewhat more likely to implement what they learned during the information-processing exercise in their own learning (Table 1). This preference may have been influenced by several factors. For example, the information-processing exercise may simply have been more fun and was therefore better received by the students (as suggested by the 56% preference rating). It could also in part be due to the results being immediate for the information-processing exercise during the workshop, whereas the self-testing exercise and its results were not as immediate since it required grading by the workshop instructor, despite obvious improvements between the first and second self-tests (Figure 3).

Regardless of overall preference, students generally valued the visualization techniques learned in this workshop. For example, several students contacted the workshop instructor to relay how they had tried implementing one or both of these techniques into their own note-taking and study practices. According to those students, there was an immediate positive effect in terms of their performance on the next exam.

SUMMARY AND CONCLUSIONS

We feel that the workshop format, working in multiple, smaller groups (up to 12 students), is particularly useful for demonstrating the efficacy of these techniques and for promoting discussion of how these techniques might be used to enhance student learning. Although both exercises could potentially be done in larger groups, in our experience, larger groups of students take more time to lead through the exercises, and large group sizes are less effective at promoting individual interaction and discussion than smaller groups. We also feel that using a separate workshop instructor adds credibility to the lecture instructor's efforts to teach students critical thinking skills, by providing a second "independent" proponent of this approach. However, there is an inherent increase in implementation cost with large college classes because multiple workshop instructors would need to be hired and trained for this small-group approach. This cost needs to be justified by significant gains in student learning, the ultimate purpose of college instruction.

Our small-scale study (within a large class) convincingly showed that students who use visualization during information processing remember more during a recall test immediately following the exercise than students who do not. In addition, our assessment of student learning on the cumulative final exam demonstrated that: 1) self-testing led to long-term learning gains for the self-testing topic; 2) these learning gains apply to both lower- and higher-order thinking skills; and 3) students across achievement levels (GPA) benefit. Thus, all the previously stated criteria for large-scale implementation and further evaluation have been met.

Ideally, the self-testing exercise (taught to all students in class) would be followed by a series of weekly study groups where visualization and self-testing are practiced by students and applied to all class topics. These weekly study groups could even be facilitated by trained peers, recruited from the same class or from previous classes (Crouch and Mazur, 2001; Stanger-Hall *et al.*, 2010). This practice will help achieve the ultimate goal of teaching students self-regulating techniques, encouraging them to use what they learned from the workshop exercises for all other class topics and to do so while studying on their own.

ACKNOWLEDGMENTS

REFERENCES

Agarwal PK, Karpicke JD, Kang SHK, Roediger HL, McDermott KB (2008). Examining the testing effect with open- and closed-book tests. Appl Cogn Psychol 22, 861–876.

Benjamini Y, Hochberg Y (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol 57, 289–300.

Bloom, BS (1956). Taxonomy of Educational Objectives: The Classification of Educational Goals, Chicago: Susan Fauer Co.

Bolt M (1996). Instructor's Resources to Accompany Exploring Psychology by David G. Myers, New York: Worth.

Boyer Commission on Educating Undergraduates in the Research University (1998). Reinventing Undergraduate Education: A Blueprint for American Research Universities.

http://naples.cc.sunysb.edu/Pres/boyer.nsf (accessed 9 September 2007).

Coffield F, Moseley D, Hall E, Ecclestone K (2009). Learning styles and pedagogy in post-16 learning: a systematic and critical review. Learning and Skills Research Centre, UK. www.hull.ac.uk/php/edskas/learning%20styles.pdf.

Crouch CH, Mazur E (2001). Peer instruction: ten years of experience and results. Am J Phys *69*, 970–977.

Crowe A, Dirks C, Wenderoth MP (2008). Biology in Bloom: implementing Bloom's taxonomy to enhance student learning in biology. CBE Life Sci Educ *7*, 368–381.

Fleming ND, Mills C (1992). Not Another Inventory, Rather a Catalyst for Reflection. To Improve the Academy 11, 137. www.varklearn.com/english/page.asp?p=questionnaire (accessed August 2009).

Gilbert JK (2005). Visualization: a metacognitive skill in science and science education. In: Visualization in Science Education, Vol. 1, ed. JK Gilbert, Dordrecht; The Netherlands: Springer, 9–27.

Gilbert JK (2008). Visualization: an emergent field of practice and enquiry in science education. In: Visualization: Theory and Practice in Science Education, Vol. 3, eds. JK Gilbert, M Reiner, and M Nakhleh, Dordrecht; The Netherlands: Springer, 3–24.

Gobert JD, Clement JJ (1999). Effects of student-generated diagrams versus student-generated summaries on conceptual understanding of causal and dynamic knowledge in plate tectonics. J Res Sci Teach *36*, 39–53.

Gurung RAR, Weidert J, Jeske A (2010). Focusing on how students study. JoSOTL *10*, 28–35.

Holschuh JP (2000). Do as I say, not as I do: high, average and low-performing students' strategy use in biology. J Coll Read Learn *31*, 94–108.

Irwin D, Simons J (1993). Slippery Snakes (exercise contrasting memory strategies using imagery vs. pronunciation). Developmental Educational Learning Institute in Des Moines, IA, 1993. Available online through the Wisconsin Association of Academic Librarians at http://docs.google.com/viewer?a=v&q=cache:c9mb3d3bypwJ:www.wla.lib.wi.us/waal/conferences/2006/handouts/WAAL%2520Exercises-%2520Citations.doc.

Isaacson RM, Fujita F (2006). Metacognitive knowledge monitoring and self-regulated learning: academic success and reflections on learning. JoSOTL *6*, 39–55.

Karpicke JD (2009). Metacognitive control and strategy selection: deciding to practice retrieval during learning. J Exp Psychol Gen *138*, 469–486.

Karpicke JD, Butler AC, Roediger HL (2009). Metacognitive strategies in student learning: do students practice retrieval when they study on their own? Memory *17*, 471–479.

Lord T (2008). "Darn it, professor. Just tell us what we need to know to pass your course." JCScT *37*, 71–73.

Lord T, Baviskar S (2007). Moving students from information recitation to information understanding: exploiting Bloom's taxonomy in creating science questions. JCScT *36*, 40–44.

Mathewson JH (1999). Visual-spatial thinking: an aspect of science overlooked by educators. Sci Educ *83*, 33–54.

Matt GE, Pechersky B, Cervantes C (1991). High school study habits and early college achievement. Psychol Rep *69*, 91–96.

National Research Council (2003). BIO2010: Transforming Undergraduate Education for Future Research Biologists, Washington, DC: National Academies Press. www.nap.edu/catalog/10497.html (accessed September 2009).

Nist-Olejnik SL Holschuh JP (2008). College Success Strategies, 3rd ed., Zug, Switzerland: Penguin Academics, Longman/Prentice Hall.

Paulk W (2000). How to Study in College, 7th ed., Boston, MA: Houghton-Mifflin Company.

Pressley GM (1976). Mental imagery helps eight-year-olds remember what they read. J Educ Psychol *68*, 355–359.

Revak MA, Porter DB (2001). The toothless bathing beauty and the t-test. Teach Stat *23*, 22–23.

Schönborn KJ, Anderson TR (2006). The importance of visual literacy in the education of biochemists. Biochem Mol Biol *34*, 94–102.

Stanger-Hall KF, Lang S, Maas M (2010). Facilitating learning in large lecture classes: testing the teaching teams approach to peer learning. CBE Life Sci Educ *9*, 489–503.

Tobias S, Everson H (2002). Knowing What You Know and What You Don't: Further Research on Metacognitive Knowledge Monitoring. College Board Report No. 2002–3, New York: College Board.

van Blerkom DL (2008). College Study Skills: Becoming a Strategic Learner, 6th ed., Belmont, CA: Wadsworth Publishing.

Weimer ME (2002). Learner-Centered Teaching. Five Key Changes to Practice, San Francisco: Jossey-Bass.

Wu H-K, Shah P (2004). Exploring visuospatial thinking in chemistry learning. Sci Educ *88*, 465–492.

Yip MCW, Chung OLL (2005). Relationship of study strategies and academic performance in different learning phases of higher education in Hong Kong. Educ Res Eval *11*, 61–70.