The Use of Interrupted Case Studies to Enhance Critical Thinking Skills in Biology

Tracy K. White,* Paul Whitaker, Terri Gonya, Richard Hein, Dubear Kroening, Kevin Lee, Laura Lee, Andrea Lukowiak, and Elizabeth Hayes

Department of Biological Sciences, University of Wisconsin Colleges

There has been a dramatic increase in the availability of case studies for use in the biology classroom, and perceptions of the effectiveness of case-study-based learning are overwhelmingly positive. Here we report the results of a study in which we evaluated the ability of interrupted case studies to improve critical thinking in the context of experimental design and the conventions of data interpretation. Students were assessed using further case studies designed to evaluate their ability to recognize and articulate problematic approaches to these elements of experimentation. Our work reveals that case studies have broad utility in the classroom. In addition to demonstrating a small but statistically significant increase in the number of students capable of critically evaluating selected aspects of experimental design, we also observed increased student engagement and documented widespread misconceptions regarding the conventions of data acquisition and analysis.

Traditional education in the sciences may not necessarily improve understanding of the nature of science: there remains only a weak association between students’ mastery of scientific facts and their understanding of how science works (9, 10). To some extent, this failure is not surprising. Students enter college with a dualistic vision of the world. Something is good or it is bad. It is true or it is false. It is fact or it is just a theory. It is within the scope of this dualism that many students see their instructor as an expert presenting the “proven” facts and fail to appreciate the iterative processes by which scientific knowledge advances and the tentative nature of the resulting explanations (1, 15). Ultimately, this dualism preempts the critical thinking required to acknowledge the possibility of competing explanations and, consequently, to evaluate those explanations in the light of available data.

This problem, however, probably does not stem entirely from the students’ level of cognitive development. There has been a long-standing call for the reform of science education in biology, particularly in introductory courses in which the traditional lecture format routinely disengages content from context and reinforces the role of students as passive learners (14). A variety of more student-centered instructional approaches have been proposed as alternatives to traditional lectures, including case studies (7, 14). Case studies often use narrative to provide intriguing background and generate open-ended questions in an attempt to rouse curiosity and motivate students to critically apply disciplinary concepts in an authentic context. In general, the use of case studies in teaching science enjoys an excellent reputation. Yadav et al. (17) recently reported on faculty perceptions of the instructional benefits to students when using case studies: an overwhelming majority believe that the use of case studies increases student engagement, increases critical thinking skills, and helps students to develop a deeper understanding of the concepts introduced by the case.

In addition to perceived improvements in student engagement and cognitive development, well-constructed case studies provide opportunities to effectively assess how students are using class material to construct understanding (or misunderstanding), and thereby provide effective means to “monitor” learning (2). Interrupted case studies are particularly useful in this regard, since student groups work through one section of the case, then share findings and questions with the class and instructor before moving on to the next section. This format allows instructors multiple opportunities to pose questions, review student responses, and use those responses to address student misconceptions and model answering questions appropriately. Herreid (8) refers to this as the “two-hat technique,” because the instructor alternately asks questions and then, following analysis of student responses, changes hats to answer them. Ultimately, this approach allows the instructor to model disciplinary approaches to the critical evaluation of a question or problem.

The goals of our research were to determine to what extent interrupted case studies could improve critical thinking skills with regard to the evaluation of evidence in scientific contexts, particularly the elements of sound experimental design and the conventions of data interpretation.

This work, a collaboration among eight biologists in the University of Wisconsin (UW) Colleges, represents the first step in a series of investigations designed to characterize student conceptions and effective teaching methods regarding the nature of science.

MATERIALS AND METHODS

Student population. The University of Wisconsin Colleges is an institution composed of 13 freshman-sophomore campuses located throughout the state. The mission of the institution is to provide both geographical and financial access to the first 2 years of a liberal arts education. The
The institution attracts many highly qualified students interested in the lower tuition costs or in the ability to remain close to home. It is also a “second chance” institution, and admission requirements are less stringent than at any of the comprehensive UW institutions. Following satisfactory completion of general education requirements, students are guaranteed transfer to the comprehensive UW campus of their choice, regardless of their original admission status. Students for this study were drawn from five of the 2-year campuses and from classes taught by eight members of the UW Colleges Department of Biological Sciences.

Students were divided into control or target cohorts based on whether or not their class was administered teaching case studies. All students completed both pre- and postcourse assessment case studies. In addition, target groups completed two additional teaching case studies during the semester.

Pre- and postcourse assessments. The primary goal of this study was to evaluate the efficacy of interrupted case studies as instructional tools to improve student ability to critically evaluate experimental design and data interpretation. Student understanding of these concepts was evaluated using assessment case studies administered at the beginning and end of a biology course. The assessment case studies were written to reveal the degree to which students could recognize and articulate problematic elements of experimental design. Thus, these assessment tools had a short essay format that required students to elaborate on their observations and evaluations. In addition, we strived for assessment tools that could evaluate student understanding of these concepts while not presuming an extensive understanding of biological content.

Each assessment case began with a prologue describing the goals of a particular research study being conducted by a college professor. In each case, the professor had charged four groups of students with collecting data to address these goals. The remainder of the case briefly summarized each group’s work, including their approach to collecting data, a tabular summary of their data, and a statement of their conclusions. Throughout this paper, we will refer to these fictional summaries as “vignettes.” The four data collection and interpretation vignettes incorporated in the assessment case studies included various methodological flaws related to sample size, controlled variables, randomization, and the use of descriptive statistics in data interpretation. Students were asked to carefully read these vignettes and then answer the following questions:

1) Clearly state the hypothesis being tested in this study.
2) If the hypothesis is correct, what do you expect to see in the field?
3) Which study best addresses the goals of the professor’s research?
4) What, if anything, is wrong with each study?

Teaching notes were developed for each assessment so that instructors of target groups could lead uniform discussions of the precourse assessment after student responses were collected. The discussions were intended to introduce the important concepts embodied in the assessment case.

The precourse assessment was based on the impact that goldenrod stem galls have on the growth and physiology of goldenrod plants, while the postcourse assessment was based on the effect of soil phosphate levels on bacterial diversity in soil. For ease of data analysis, precourse vignettes and postcourse vignettes mirrored one another with regard to the problematic design elements that they contained. Thus, both the pre- and postcourse assessment vignettes reflected precisely the same types of errors in data collection and interpretation. In addition, the assessments were designed to include all essential biological content in the introductory paragraph of the case study so that students would not need to have studied plant biology or soil microbiology in order to successfully complete the assessments.

A total of 366 students completed the precourse assessment early in the term (usually the first day of classes), and 322 students completed the postcourse assessment near the end of the term. The difference between pre- and postcourse enrollment was primarily due to student withdrawals during the course of the academic term.

Because this work is directed at the comparison of instructional techniques in an established educational setting, it is exempt from review by the Instructional Review Board of the University of Wisconsin Colleges (General Institutional Policy #404). Informed consent was obtained for all student responses summarized in this manuscript.

The actual assessment case studies and associated information can be found at the University of Wisconsin System Case Portal at http://www.cfkeep.org/html/gallery.php?id=91594149816403.

Teaching case studies. In addition to completing the pre- and postcourse assessment case studies, 150 students in “target courses” also practiced evaluating experimental designs via in-class completion of two interrupted case studies. These teaching case studies were designed to reiterate the important elements of experimental design and the use of descriptive statistics in data interpretation, and were administered at different points in the term, at the instructor’s discretion. Each case study was organized into sections that became increasingly more complex with regard to data collection and analysis, and students were challenged to critique the validity of both the experimental designs and associated conclusions as the case study became more involved. Students worked in small groups for each section of the case study and subsequently reported their analyses back to the class, at which point the instructor addressed misconceptions and clarified problematic approaches and conclusions. We developed four different teaching case studies and students in target groups worked through two of these. Each case study involved a different biological context appropriate to introductory biology courses, but all focused on the elements of sound experimental design and the conventions of data interpretation. The teaching case studies and associated information are available at the University of Wisconsin System Case Portal at http://www.cfkeep.org/html/gallery.php?id=91594149816403.
Summary of student responses. The assessment case studies had a short essay response format that asked students to elaborate on their observations and evaluations. After we read through these written responses, a number of patterns became obvious, and a data summary worksheet was developed to summarize and score the pre- and postcourse assessments based on these patterns. First of all, hypotheses were identified as scientific, statistical, or “other,” and predictions were categorized as correct or incorrect given the background information provided.

The term “hypothesis” is used in many ways, creating considerable confusion, and not just among students. In analyzing and scoring student hypothesis statements we relied heavily on Giere’s (6) “four box model” which clearly delineates the relationships among the natural world, hypotheses, predictions, and data. According to this model, a tentative explanation of how some aspect of the natural world works is a scientific hypothesis. Various predictions may be deduced from a scientific hypothesis, and the predictions can be directly compared to data that we collect from the natural world. Thus, we can prove predictions to be true or false, but because there is no direct connection between data and a scientific hypothesis, the hypothesis can be supported or refuted, but not proven. Unfortunately, many students have been taught that predictions are scientific hypotheses, and this can make it difficult for them to understand that hypotheses are tentative. Readers with backgrounds in statistics will be aware that Giere’s predictions are also correctly known as statistical hypotheses, but few introductory-level students have been exposed to this complication.

When students were asked to state the hypothesis being tested in the research being conducted in the assessment case study, several patterns were identified in their constructions, although the bulk of their hypotheses fell into one of two categories (13):

(i) The student provided a scientific hypothesis, i.e., the tentative description of a mechanism from which specific predictions can be made for use in testing the hypothesis.

(ii) The student provided a statistical hypothesis, i.e., a prediction that can be proven true or false by direct comparison of the data.

In addition to characterizing hypothesis construction, the student comments on the experimental design and data interpretation of each vignette were categorized as relating to sample size, statistical analysis, randomization, incorrectly controlled variables, and other. If a student identified an irrelevant problem, the nature of that problem was recorded.

The data reported in Results reflect absolute numbers of students as a percentage of total students completing the assessment (366 students precourse, 150 students and 172 students in postcourse target and control cohorts, respectively). Changes in the types of responses from precourse to postcourse assessments between target and control cohorts were tested for statistical significance using chi-square analysis with one degree of freedom. Where statistical significance is reported in Results, we have also indicated the chi-square and \( P \) values.

RESULTS AND DISCUSSION

The nature of student hypotheses. The prologue of each assessment case study explained the broad research question driving the experiments described in the case. Students were asked to carefully read this information and compose a hypothesis that reflected the goals of the research. Most students composed a statistical hypothesis, one that is stated in a way to statistically confirm a pattern, e.g., goldenrod with galls will be shorter than those without galls. Giere (6) calls such statements predictions, since they lack explanatory value. Some students, however, composed more sophisticated scientific hypotheses, e.g., a gall redirects a plant’s resources so it will not grow as tall, and there was a significant increase in the number of students posing scientific hypotheses in the postcourse assessments (chi-square = 30.4; \( P < 0.0001 \)) (Fig. 1). It is of interest that this improvement was slightly greater among students who had not used the teaching case studies (chi-square = 4.84; \( P \leq 0.05 \)).

Our initial motivation for asking students to compose a hypothesis was to determine if they understood the correct form of the hypothesis and whether they appreciated the function of the hypothesis in directing research. As we evaluated student responses, we became aware of additional issues, particularly differences in the structure and role of statistical and scientific hypotheses (13) and whether or not introductory-level students could or should understand this distinction. We determined that this dichotomy was stressed more in some classes than in others due to the teaching methods of individual instructors. Since an introduction to hypothesis construction is a nearly universal topic in introductory biology, it seems likely that differences in teaching pedagogies may also be a primary influence that contributed to our control cohort improvement over the target group with respect to hypothesis construction.

Student ability to recognize problems with experimental design. Students were asked to evaluate the experimental design and the validity of the conclusions for four different approaches to the same research question. In each of these four vignettes, the experimental design contained one or two critical flaws in some combination of design elements including sample size, controlled variables, randomization, and/or data interpretation.

In general, our students were unfamiliar with descriptive statistics (mean, standard deviation, and standard error), why they are used, and how they should be interpreted (Fig. 2, data analysis bars). In particular, students failed to take descriptive statistics into account when determining the validity of conclusions. Most students struggled with this skill throughout the term, although the target cohort (exposed to two teaching case studies) demonstrated a highly significant increase (chi-square = 50.48; \( P < 0.001 \)) in the number of students capable of correctly incorporating descriptive statistics into their evaluation of the conclusions reported in the vignettes (Fig. 2). Our data also indicate that the target cohort was bet-
terable than the control cohort to recognize sample size problems in the postcourse assessment. However, neither target nor control groups were as successful at recognizing this problem at the end of the term as they had been at the beginning (Fig. 2, sample size bars). Students appear to enter our classrooms with a clear conception that sample size is an important factor in experimental design. We attribute the overall drop in student success at recognizing this factor to a potential problem with the content of the postcourse assessment, and not to a loss of understanding. Specifically, the postcourse assessment was based on the effect of soil phosphate levels on bacterial diversity, and we suspect that many students did not recognize that multiple bacterial species would be identified from one soil sample. Hence, these students confounded the number of bacterial species with the number of samples. When seeking insight into the ways

FIG. 1. Students generating scientific hypotheses.

FIG. 2. Student ability to recognize existing problems with experimental design and data interpretation pre- and postcourse. Student comments about problems with the experimental design and data interpretation in the assessment vignettes were categorized as relating to sample size, statistical analysis, randomization, incorrectly controlled variables, and “other.” The graph shows the number of students with comments relating to problems that actually existed in the vignettes.
in which students construct understanding, a lack of content knowledge can confound attempts to assess cognitive thinking skills, and this can be a difficult issue to resolve when designing assessment tools. It seems likely that, despite our efforts to write assessments in which understanding was divorced from content, differences in prior content knowledge and experience probably impacted these data. Clearly, future studies will need to reconcile this difficult issue.

Both target and control cohorts demonstrated a significant (chi-square = 7.76; P < 0.01) increase in the number of students capable of recognizing problems with sampling methods (Fig. 2, randomization bars). However, students in our control group appeared to improve slightly more than students in the target cohort (chi-square = 5.53; P < 0.05). While it might be simple to conclude that this reflects an increase in the number of students who had a firm grasp on the role of randomization in experimental design, the confusion evident in student responses makes it difficult to interpret these data. The variety of responses that addressed the manner in which data were collected seems to indicate that students appreciate that there are rules for collecting data, but they are confused by the role of random sampling in this process. For example, the pre- and postcourse assessment case studies both included a vignette in which nonrandom sampling was included as an experimental design flaw. In the precourse assessment case study, student researchers are described as sampling goldenrod plants along a roadside, whereas in the postcourse assessment case study, soil samples are collected during a canoe trip down a river bordering the research site. Students completing the assessment made a variety of revealing comments on this latter approach to sampling. For example, a number of our students noted that the researchers described in the case study were having too much fun, apparently in direct contrast to what is allowed when scientists collect data. Indeed, student responses revealed a tendency to equate randomization with randomness, and both terms appear to imply a lack of the control they assume should be associated with scientific endeavors. In fact, both pre- and postcourse assessments included one vignette in which appropriate random sampling was explicitly described, and a number of students indicated that data collection was “too random.” Students often seemed concerned that properly done randomization would introduce unnecessary variability into the experiment. In some instances, students articulated a problem with the averaging of data that were collected randomly. Clearly, we have identified an element of experimental design that is rife with misconceptions, and this topic will serve as the focus of further research.

In both the pre- and postcourse assessments, students were presented with a vignette in which conclusions were drawn based on the comparison of data that had been collected in two separate years. Most of the students assessed found this problematic, and many articulated the issue correctly appreciating the lack of controlled variables (Fig. 2, controlled variables bars), though it is important to note that their language did not necessarily incorporate that term. Despite their ability to identify this problem, their responses often made it clear that this concept was slippery for them. One of the research vignettes in the precourse assessment compared goldenrod plants with and without galls that had been raised in a greenhouse, and we were struck by the number of students who found the relatively controlled environment of the greenhouse to be “unnatural” and therefore problematic. Ironically, while many students objected to the greenhouse not being natural enough, other students expressed concern regarding the lack of control over certain variables (light, water, temperature, etc.) that must impact the field studies described in subsequent vignettes and suggested that this also invalidated the data.

**CONCLUSION**

Our experience in the execution and analysis of this work has convinced us that the use of case studies has broad potential impacts on both teaching and learning. In addition to demonstrating a small but statistically significant increase in the number of students capable of critically evaluating selected aspects of experimental design, we also observed increased student engagement and documented widespread misconceptions regarding the conventions of data acquisition and analysis.

Our data demonstrate that following the repeated use of interrupted case studies, there is a small but significant increase in the number of students who can accurately characterize problematic interpretations of descriptive statistics compared to traditional instruction. Additionally, when compared to the control group, the target cohort demonstrated a significant increase in the number of students capable of recognizing inadequate sample size. While these changes are small, it is possible that they reflect very large leaps in cognitive development for a selected subset of students. Constructivist explanations of learning suggest that students learn by adding new information to existing mental scaffolds, and that learning, in some circumstances, may initially require significant reorganization of that mental scaffolding if students maintain faulty preconceptions. Solid understanding of a new concept often requires “multiple generations” of exposure to material, and students are likely to traverse a series of relatively unsophisticated stepping stones before complete understanding is achieved (5, 16). Bloom’s taxonomy (3) provides a putative hierarchy of cognitive skills, beginning with knowledge and comprehension, e.g., factual recall, and peaking with evaluation, which requires judgment using critical analysis of evidence. Ultimately, both foundational knowledge and conceptual mastery are required to achieve the most sophisticated levels of understanding (5). Because our assessments required critical evaluation and elaboration rather than recall or simple identification, it is likely that they were unable to measure more modest cognitive gains among students, the gains which are more likely to occur in the early stages of learning. While interpretation of these early data remains difficult, the observation that the interrupted case studies appear to have moved a small cohort of students to fairly sophisticated levels of understanding may attest to their power in the classroom.
Clearly, it would be of value to proceed with a comparative analysis of assessment approaches with regard to the impact of case studies on student learning. However, as Lundeberg and Yadav (11, 12) point out, while perceptions of student learning and engagement have often been measured, actual learning gains are more difficult to assess and have seldom been measured in research on case-study-based instruction. The extent to which case studies succeed in improving student skills likely rests on the intersection of a number of parameters, including individual instructor’s approaches, student experience and motivation, the structure of the case studies, and the novelty and complexity of the content included in the case studies. For example, Brickman et al. (4) assert that the success of case studies and the discussions that stem from them are related to the manner in which cases are introduced by the instructor. And, as is probably reflected in this study, the ability to recognize success will also depend on the degree to which assessment methods take a learning continuum into account. It is an important focus for future research to determine which subpopulation of students most benefited from this approach, why they benefited, and how their success might be extended to a larger proportion of their classmates. Additionally, future studies in which assessments have been designed to evaluate learning at multiple levels will also likely provide further insight into the ways in which understanding is constructed. Indeed, as case-study-based learning enjoys further assessment, a great deal of useful information will likely be obtained by closely examining the multiple factors that contribute to student success.

It is important to note that our experience with case studies revealed benefits that go beyond their use as simple teaching instruments, and it is likely these observations contribute to the potential effectiveness of case studies in the classroom. For example, the majority of instructors in our group who used the interrupted case studies to introduce experimental design would concur with the findings of Yadav et al. (17): the use of case studies dramatically improves student engagement and directs student focus in a way that is difficult to achieve using lecture formats. And interrupted case studies provide opportunities to model appropriate disciplinary approaches to critical analysis that are difficult to achieve using more traditional formats.

We found that one of the most powerful attributes of case studies was their ability to provide insight into student conceptions and misconceptions, particularly those surrounding the perceptions of “control” and “randomness” in the context of scientific investigations. In addition, students demonstrated consistent (though not universal) confusion regarding how to determine adequate sample size and how to calculate and interpret descriptive statistics. For example, a number of students objected to the comparison of averages that were based on samples of different sizes. Instructors within our research group have used this information as incentive to design focused tutorials to help students challenge these misconceptions, and these areas should be productive areas for future study.

In summation, it is worth noting that the questions raised by this ongoing collaborative project have prompted continuing discussions within our department regarding how to help students best understand the nature of science, causing many department members to clarify instructional goals for this topic in our courses. Thus, this exercise had important overarching results for our approaches to assessment, our appreciation of the breadth of case study impacts, and our development as instructors.

ACKNOWLEDGMENTS
We are pleased to acknowledge the support of the University of Wisconsin System Office of Professional and Instructional Development for financial support, as well as Lil Tong and the University of Wisconsin System Center for Biology Education for technical support. In addition, numerous colleagues both administered case studies and reviewed our work throughout this study, and we are grateful for their thoughtful contributions.

REFERENCES