THINK is a new initiative at NC State University focused on enhancing students’ higher-order cognitive skills. As part of this initiative, I explicitly emphasized critical and creative thinking in an existing bacteriophage discovery first-year research course. In addition to the typical activities associated with undergraduate research such as review of primary literature and writing research papers, another strategy employed to enhance students’ critical thinking skills was the use of discipline-specific, real-world scenarios. This paper outlines a general “formula” for writing scenarios, as well as several specific scenarios created for the described course. I also present how embedding aspects of the scenarios in reviews of the primary literature enriched the activity. I assessed student gains in critical thinking skills using a pre-/posttest model of the Critical Thinking Assessment Test (CAT), developed by Tennessee Technological University. I observed a positive gain trend in most of the individual skills assessed in the CAT, with a statistically significant large effect on critical thinking skills overall in students in the test group. I also show that a higher level of critical thinking skills was demonstrated in research papers written by students who participated in the scenarios compared with similar students who did not participate in the scenario activities. The scenario strategy described here can be modified for use in biology and other STEM disciplines, as well as in diverse disciplines in the social sciences and humanities.

INTRODUCTION

Robert Ennis (4) defines critical thinking as reflective and reasonable thinking that is focused on deciding what to believe or do. He argues that critical thinking is a practical activity that includes creative actions such as raising questions, formulating hypotheses, generating alternatives, and making plans about gathering information. Unfortunately, there is ample evidence that many adults, including college students, consistently fall prey to flawed and biased thinking (7, 10, 9). The good news is that critical and creative thinking skills are not limited to the few geniuses who are “born with it”; these higher-order thinking skills can be developed through practice, feedback, and reflection (11, 12).

The American Association for the Advancement of Science (AAAS) and the National Science Foundation (NSF) have issued a call to action in Vision and Change in Undergraduate Biology Education (1), recommending that all institutions of higher education focus on five core concepts for biological literacy and six core competencies. Interestingly, the definition of the first core competency, the “ability to apply the process of science” reflects the creative process in a vast array of disciplines. Much of the description, while written with a biological slant, is not unique to the biological sciences, nor even the natural sciences. Creative work in most disciplines requires the “creator” to raise vital questions, to gather relevant information, to generate multiple ideas (multiple possible hypotheses, multiple solutions, multiple interpretations, etc.), to interpret and evaluate information/data, and to draw appropriate conclusions.

Vision and Change recommends giving all students the opportunity to develop these competencies through authentic research experiences. Certainly, it would be ideal to provide every student with an authentic research experience, but there are real barriers to doing so. At many institutions, it is not feasible to provide every student majoring in the life sciences with an undergraduate research experience, let alone all of the non-major students who will be faced with biology-related decisions, such as whether to vaccinate their children, throughout their lives. Institutions of higher education should be striving to build inquiry into general biology lab courses, but not every school requires lab for non-majors. Furthermore, even laboratory activities that attempt to be inquiry-based do not always emphasize the higher-order thinking competencies, especially when huge numbers of students must be accommodated. While
we should promote authentic research for as many students as we can, we must examine ways to enhance the thinking skills developed during the experience, as well as developing strategies to enhance critical thinking across the curriculum.

An observation I have made in student research papers and lab reports is that although many of my students reach feasible conclusions, they almost always provide only one interpretation of their data and draw a single conclusion from it, even when there are alternative explanations to account for the data. On a related note, before embarking on this study, I had never witnessed a first-year student in my class independently question an interpretation of data or conclusion reached in a scientific journal article, even when I purposefully selected a paper with an overblown conclusion. These skills are critical for scientists and, in fact, critical for anybody who ever needs to make a decision based on evidence. Because of the lack of demonstrated use of these skills in my own students, I set out to help them engage in the habits of mind that would lead them to think more critically. One way I did this was through the use of targeted discipline-specific scenarios.

The use of critical thinking scenarios is flexible both in terms of disciplinary content, as well as in terms of courses of different structure and class size. I utilized the strategy described in this paper in a first-year undergraduate research course focused on bacteriophage biology. However, the strategy could be easily implemented in a lecture course, a seminar course, or any other course structure or academic club. I used the activity as a springboard for discussion, but it could be used as a graded activity. Depending on the complexity of the scenario, this strategy could also be used at virtually any academic level.

Student learning outcomes

The course had many desired outcomes. The student learning outcomes described and assessed in this paper are:

A. Students will recognize inappropriate inferences and will question overstated conclusions in the works of others.
B. Students will avoid making inappropriate inferences from data.
C. Students will provide multiple explanations/interpretations of data.
D. Students will determine ways to assess the best explanation or interpretation of data.

METHODS

Use of human subjects

This study was reviewed and granted exempt status by the North Carolina State University Institutional Review Board (IRB).

Student demographics

The test group students were enrolled in the SEA-PHAGES (8) first-year research course in fall 2014. This test group participated in critical thinking scenarios, two journal article discussions where elements of the scenarios were embedded in the discussion prompts, and a final research paper; the test group also was assessed pre and post by the non-discipline-specific Critical Thinking Assessment Test (CAT). The control group students were enrolled in the SEA-PHAGES first-year research course in fall 2013. This control group participated in two journal article discussions where elements of the scenarios were not embedded in the discussion prompts and a final research paper; the control group also was not assessed by the CAT.

Test group. The test group was composed of fourteen first-year, first-semester undergraduate students; five females and nine males. Eleven of the fourteen students in the course participated in the University Honors Program (UHP); on average, UHP students enter the university with a higher degree of academic achievement than our typical first-year student and have demonstrated intellectual curiosity in essays during the selection process. The minimum high school grade-point average (GPA) of entering UHP students is 3.75 unweighted or 4.5 weighted; the minimum scholastic aptitude test (SAT) score is 1,300 (critical reading and critical math only) and the minimum ACT is a 30 composite. For the fall 2014 incoming UHP class, the average unweighted GPA was 3.80 and the average SAT score (critical math/reading) was 1,376.

Control group. The control group was also composed of fourteen first-semester, first-year students enrolled in the same course the previous year, twelve females and two males. In this cohort, nine out of fourteen students participated in the UHP. The final research papers of this group were compared with those of the test group.

Critical thinking scenarios

Implementation. Over the course of the semester, students in the test group engaged in five critical thinking scenarios. Some scenarios were written before the start of the course and planned in the syllabus, and some were created spontaneously during class to guide students when they were not thinking through experimental plans or data interpretation fully.

I organized the class activity similarly to a classic “think-pair-share.” In each case, the instructor provided the scenario, and students were given 10 to 15 minutes in class to write their response independently. Students then formed small groups of three to four and discussed their responses. The class then came back together to share their responses, and the instructor provided feedback to the class as a whole. Students were then permitted to flip
their papers over and add to their response on the reverse side. Papers were not graded but were collected to ensure thoughtful participation.

Although we did the activity fully within the class period, it could be modified such that students completed the individual write-up, or even the paired/small group discussion, outside of class time. The important feature is that the open discussion and feedback from the faculty member occur in class. These modifications would allow the activity to be used in courses with higher enrollment or where more content had to be covered during classroom time.

**Structure.** The structure I used for the in-class scenarios is outlined below.

* Lay out scenario, provide image, or present data.
  * Q1. “What does the author want you to infer?” or “Does the evidence strongly support the conclusion?”
  * Q2. “Provide alternative explanations.”
  * Q3. “What additional information would you need to draw a conclusion?”

I based scenarios on overstated conclusions I found in primary literature related to the course content, current events relevant to the course or to student life, or potential outcomes of experiments that the students designed.

The first scenario used in class, with aggregated student responses, is outlined in Figure 1 (16) and the text below. Additional scenarios used, without student responses, are provided in Appendix 1.

**Scenario 1: MMR (measles, mumps, rubella) vaccine and autism**

1. What does the author of the graph want you to infer?

   **Student response:** The author wants me to infer that the MMR vaccine causes autism.

   ![Graph showing MMR vaccine and autism](image)

   **FIGURE 1.** Data for MMR/autism critical thinking scenario. MMR = mumps, measles, rubella.

   2. Give an alternate explanation for the data shown in the graph.

   **Student response:** The increase in autism shows only occurrence, not incidences per X number in the population. Perhaps the increase in cases is explainable by population growth.

   **Student response:** Could there have been gradual changes in the frequency with which autism was diagnosed or reported?

   **Student response:** If MMR was given to almost all children by some point in time, shouldn't new cases of autism have leveled off once that had occurred?

   **Student response:** Even if neither population growth nor diagnosis/reporting can account for the increase in the incidence of autism, introduction of the MMR vaccine is definitely not the only change that occurred in our society during that time frame. The possibilities to investigate are endless. Was there a gradual increase in paternal age? Was there an increase in the use of anti-flammable chemicals in baby and toddler pajamas? Did pregnant mothers gradually eat less fish? Was the use of a pesticide in apples increased over time? Was there a new type of material used in food packaging? The list goes on and on.

   **Student response:** Research whether there were any changes in the definition or diagnoses of autism during the given time period.

   **Student response:** Even if a correlation can be made, causation is not proven. Experimental evidence in laboratory animals, or something like that, would need to be investigated to show causation.

**Primary literature reviews**

We previously described a method to systematically introduce primary scientific literature to first-year undergraduate students (3). In that paper, we provided a template for a written summary of journal articles that students brought to class to facilitate their active participation in classroom discussion. In the present work, I followed a similar strategy for introducing students to primary literature, but I revised the instructions to explicitly place additional
emphasizes on critical thinking goals. The additional prompts that students responded to were:

Did the conclusions follow logically from the data? Provide an example.

Did the authors consider alternate conclusions of the data? Provide an example. Are there any interpretations that you thought about that the author did not consider?

Given the findings and approach taken in this work, does it lead you to any questions you would like to address with respect to our research project? List as many research questions as you can think of. Select one research question and provide a possible methodology to pursue (briefly). Explain why you selected the above research question from among your alternatives.

In essence, the additional prompts in the journal article write-ups mimic the structure of the critical thinking scenarios and provide further opportunities for students to practice and receive feedback in the desired student learning outcomes that this study focuses on.

Students in the control group received the instructions described in Carson and Miller (3) without the additional above prompts.

Research papers

All students (test and control groups) wrote a final research paper in the format of a scientific journal article. The discussion section of the research papers provided an opportunity for students to demonstrate student learning outcomes B, C, and D in a real-world situation.

Assessment

We included three measures of student learning, two of which are direct assessments of student learning. The Critical Thinking Assessment Test (CAT) is useful in that it allows for a pre/post assessment to look at higher-order thinking gains within the semester. Assessment at the start of a course (pre) can be difficult to implement in a discipline-specific way because, at the beginning of the semester, a student does not always have the knowledge to achieve higher-order thinking in the discipline. We complemented the use of CAT with a discipline-specific activity that requires higher-order thinking at the end of the semester, scored using a rubric. Another feature of the CAT is that it does show whether skills were transferred beyond disciplinary content.

Critical thinking assessment test. The Critical Thinking Assessment Test (CAT) was developed at Tennessee Technological University with funding by the National Science Foundation (13). The problem scenarios used in the CAT instrument are very general and do not require specialized discipline-specific knowledge. Therefore, using the CAT allows for assessment of higher-order thinking without being conflated by the presence or lack of disciplinary knowledge.

The CAT instrument has been assessed for validity and reliability (15). It contains 15 open-ended non-discipline-specific questions that assess higher-order thinking skills in the realms of evaluating and interpreting information, problem solving, creative thinking, and effective communication. The specific skill measured by each question is included in the results section in Table 1.

The CAT was employed to directly measure gains in higher-order thinking skills in a pre/post assessment approach where students took the exam in both the first week and the last week of the course. Both pre and posttests were coded and scored after the end of the semester by faculty at Tennessee Tech in order to minimize any unconscious biases. This assessment was used for the test group only.

Common rubric. I utilized a rubric to assess mastery of critical thinking skills in the final research paper for both the test group and control group. Students were not provided this rubric, although the instructions to students indicated that the discussion section must provide complete and thoughtful interpretation of data. The two rubric items relevant to this study were:

Does the student generate multiple potential ideas about the issue at hand?

1. Explores only a singular idea
2. Uses a few elements but with limited exploration
3. Experiments with multiple elements and variables but shows difficulty in addressing their appropriateness
4. Experiments with multiple elements and variables and successfully chooses an appropriate idea

How does the student make judgments about benefits and drawbacks of various ideas about the issue at hand?

1. Shows limited or no awareness of the benefits and drawbacks of ideas or defends them with unrelated criteria
2. Recognizes relevant benefits and drawbacks of ideas
3. Weights the value of relevant benefits and drawbacks of ideas
4. Weights the value of relevant benefits and drawbacks and selects appropriate ideas

Student reflections. In the last week of the semester, students in the test group responded to the following question prompt: Please describe the single most important assignment, strategy, or other course aspect that made a meaningful difference to your higher-order thinking abilities.
RESULTS

Direct assessment of critical thinking skills gains in test group

I directly measured gains in students’ critical thinking skills using the CAT in a pre/post testing model. Table 1 indicates the skill assessed by each CAT question, the maximum point value of each question, the mean score for each question on the pre and post assessment, probability of difference, and effect size. Because questions have different maximum point values, results are also displayed as a percent of total point value in Figure 2.

I observed gains in the mean scores of 12 out of 15 items on the CAT from the beginning of the semester to the end of the semester. Even with the small number of students in the course (14), there was a statistically significant large effect on the skills measured by the CAT overall, and statistically significant large effects on the specific items of “summarizing a pattern of results without making inappropriate inferences” and “providing alternative explanations for spurious associations.”

Notably, students in this study performed highly in many of the questions on the pretest. The average student score was greater than 85% on four of the questions in the pretest, and greater than 70% on seven of the questions in the pretest. Because of this, a ceiling effect may have occurred where no statistically significant increase of skills could be measured in several areas because students already performed at the highest level measurable by the CAT instrument at the start of the semester.

Student reflections. Admittedly, student reflections do not directly measure ability, and individuals do not always reliably identify which experiences truly enhanced learning. However, gathering student reflections served

### Table 1. Mean scores on pre/post critical thinking assessment test, N = 14.

<table>
<thead>
<tr>
<th>Skill Assessed by CAT Question</th>
<th>Max Points</th>
<th>Pre Mean</th>
<th>Post Mean</th>
<th>Probability of Difference</th>
<th>Effect Size a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1 Summarize a pattern of results without making inappropriate inferences.</td>
<td>1</td>
<td>0.71</td>
<td>1.00</td>
<td>*</td>
<td>+0.86</td>
</tr>
<tr>
<td>Q2 Evaluate how strongly correlational-type data supports a hypothesis.</td>
<td>3</td>
<td>1.83</td>
<td>2.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q3 Provide alternative explanations for a pattern of results that has many possible causes.</td>
<td>3</td>
<td>1.36</td>
<td>1.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q4 Identify additional information needed to evaluate a hypothesis.</td>
<td>4</td>
<td>1.57</td>
<td>1.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q5 Evaluate whether spurious information strongly supports a hypothesis.</td>
<td>1</td>
<td>0.93</td>
<td>0.93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q6 Provide alternative explanations for spurious associations.</td>
<td>3</td>
<td>1.93</td>
<td>2.36</td>
<td>*</td>
<td>+0.69</td>
</tr>
<tr>
<td>Q7 Identify additional information needed to evaluate a hypothesis.</td>
<td>2</td>
<td>0.50</td>
<td>0.43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q8 Determine whether an invited inference is supported by specific information.</td>
<td>1</td>
<td>0.86</td>
<td>0.93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q9 Provide relevant alternative interpretations for a specific set of results.</td>
<td>2</td>
<td>1.07</td>
<td>1.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q10 Separate relevant from irrelevant information when solving a real-world problem.</td>
<td>4</td>
<td>3.71</td>
<td>3.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q11 Use and apply relevant information to evaluate a problem.</td>
<td>2</td>
<td>1.00</td>
<td>1.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q12 Use basic mathematical skills to help solve a real-world problem.</td>
<td>1</td>
<td>0.86</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q13 Identify suitable solutions for a real-world problem using relevant information.</td>
<td>3</td>
<td>2.14</td>
<td>2.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q14 Identify and explain the best solution for a real-world problem using relevant information.</td>
<td>5</td>
<td>3.57</td>
<td>3.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q15 Explain how changes in a real-world problem situation might affect the solution.</td>
<td>3</td>
<td>1.29</td>
<td>1.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAT Total Score</td>
<td>39</td>
<td>23.33</td>
<td>26.02</td>
<td>*</td>
<td>+0.67</td>
</tr>
</tbody>
</table>

aMean difference divided by pooled group standard deviation. (0.1–0.3 = small effect; 0.3–0.5 = moderate effect; >0.5 = large effect).

CAT = Critical Thinking Assessment Test.

*p < 0.05.
two valuable purposes. First, it encouraged students to engage in metacognitive behavior, which was promoted over the course of the semester in various ways. And second, together with the direct measures of ability, student responses shed light on their perception of what worked in the classroom. In the last week of the semester, students were asked to respond to the following question prompt: Please describe the single most important assignment, strategy, or other course aspect that made a meaningful difference to your higher-order thinking abilities. Complete and unmodified student responses are displayed in Appendix 2.

Two students specifically identified the critical thinking scenarios as the most impactful activity that enhanced their higher-order thinking skills, four students identified the primary literature reviews, and three students identified the research papers. Five students identified other aspects of the course, such as experimental design, keeping of lab notebooks, or general reference to class discussion. There was quite a bit of overlap in most of the specified activities. For example, critical thinking scenarios were utilized in guiding students in the experimental design process as described in the “Recombinant Bacteriophage Lysin Scenarios,” and scenarios were also embedded within the journal article write-ups and discussions. Based on the quality of the responses, it seems likely that students thoughtfully considered their responses.

Comparison with other high-impact learning interventions. Because we know that high-impact practices associated with undergraduate research are likely to affect higher-order thinking skills, we wanted to see whether the use of scenarios increased the effect. We compared our observed gains with other high-impact learning interventions such as the CREATE (Consider, Read, Elucidate hypothesis, Analyze and interpret data, Think of the next Experiment) strategy (6) and another undergraduate research initiative that focused on critical thinking (5). The overall pre and post scores between this study and these two other high-impact courses in the biological sciences that also assessed critical thinking using the CAT are compared in Figure 3. The overall effect sizes (mean difference divided by pooled group standard deviation) of each study are as follow: this study (+0.67 = large effect), Gasper and Gardner (+0.3 = small/moderate effect) (personal communication S. Gardner, 26 February 2015), Gottesman and Hoskins (0.97 = large effect) (6). Effect sizes were categorized as follows: 0.1 to 0.3 = small effect; 0.3 to 0.5 = moderate effect; greater than 0.5 = large effect. The CREATE project had a very large effect, but based on pretesting, it is clear that the students in that initiative had a much lower baseline of critical thinking skills than the students in the initiative described in the present paper. In the study described in the present paper, ceiling effects occurred in several of the questions where students scored very highly on average in the baseline testing, leaving no opportunity to measure improvement. Comparison with the Gasper and Gardner authentic research experience is therefore more meaningful since students in both studies participated in an authentic research experience, and the baseline (pretest) critical thinking abilities of the two groups of students are comparable. Overall, our students benefitted, with a larger effect in their critical thinking skills than the students in the Gasper and Gardner study. Because of this difference, I hypothesize that when discipline-specific scenarios are included in an authentic research experience, students’ critical thinking skills are even more enhanced.

Comparison of student research papers between the test group and the control group. In both fall 2013 and fall 2014, students in the SEA-PHAGES first-year research course wrote comprehensive research papers that were in the standard format of a scientific journal article in life science disciplines. Students were instructed that the discussion section must provide complete and thoughtful...
interpretation of data but were not provided with the critical thinking rubric utilized for this study.

The skills relevant to this study that were assessed by the rubric, with further explanations of the questions, are set out below. The ratings for each question are described in the materials and methods. Although this is a Likert-like scale, I utilized specific descriptions of the skill level in order to avoid subjective scaling.

A. Does the student generate multiple potential ideas about the issue at hand? This was assessed by whether the student explored more than one possible explanation or interpretation of data before drawing a conclusion where appropriate.

B. How does the student make judgments about benefits and drawbacks of various ideas about the issue at hand? It is important to note that if a student scored the lowest skill level on Question A, s/he could not score above the lowest level on this question. This was assessed by whether the student systematically and logically weighed the benefits and drawbacks of explanations/interpretations and his or her ability to reach an appropriate conclusion or suggest information that would be needed in order to reach a final conclusion.

The increase in the test group students’ likelihood to explore more than one potential explanation or interpretation of data compared with the control group was striking (Fig. 4). The control group students generally explored only a singular idea or used a few elements with limited exploration, while the test group of students who participated in the critical thinking scenarios generally explored multiple explanations/interpretations of data and recognized relevant benefits and drawbacks of ideas, thus demonstrating student learning outcomes C and D.

DISCUSSION

Generally, few gains in higher-order thinking skills are accomplished in courses in the absence of a specific, targeted intervention. In the majority of college courses, there is no change pre to post on the Critical Thinking Assessment Test (13). In this study, a statistically significant large effect on students’ overall critical thinking skills was measured by the CAT in a single semester. It is likely that high-impact activities including participating in and discussing multiple critical thinking scenarios across the semester, together with analyzing and discussing primary literature, were the key factors leading to the increase. Research papers also provided opportunities for students to practice and receive feedback on skills.

In the introduction to this paper, I outlined the four student learning outcomes that I wanted students to achieve through the use of the critical thinking scenarios. Questions 1 to 9 in the CAT directly measure the skills that students practiced in the critical thinking scenarios and questions 5 and 6 in the CAT directly measure skills practiced in the literature write-ups. Table 2 maps the student learning outcomes with the assessment questions on the CAT.

The skills described in learning outcomes A and B are overlapping and are assessed by four questions on the CAT. Students demonstrated a statistically significant large gain on the specific skill of “summarizing a pattern of results without making inappropriate inferences (Q1),” and though not statistically significant, I observe increasing trends in the closely related skills of “evaluating how strongly correlational-type data supports a hypothesis (Q2)” and “determining whether an invited inference is supported by specific information (Q8).” I observed no change in “evaluating whether spurious information strongly supports a hypothesis (Q5),” likely due to a ceiling effect since students performed at a very high level on the pretest. I believe these gains were due to the practice and feedback students received through the scenarios and the literature write-ups and discussions. Perhaps also impactful was the fact that these activities gave students permission to question authority. As one student explained: “The article reviews also encouraged me to critically evaluate another person’s thinking. I realized the blind trust I gave to authors, and how little I stopped to question their methods and logic” (Appendix 2).

Students also met learning outcome C. Students demonstrated a statistically significant large gain in “providing alternative explanations for spurious associations (Q6).” They also displayed a positive trend in “providing alternative explanations for a pattern of results that has many possible causes (Q3)” and “providing relevant alternative interpretations for a specific set of results (Q9).” Every scenario asked students to provide multiple possible explanations for data or evidence, and students identified where authors gave multiple possible interpretations when discussing primary

FIGURE 4. Mean rubric scores with standard error of control group versus test group. N = 14 for both groups. p = 0.0064 and p = 0.016 (2-tailed) for each question respectively.
literature. One student stated, “The Critical Thinking Experiment Scenarios made me consider alternate explanations for patterns, trends, and experimental results. I had to think about how to explain certain observations and how to make sure my assumptions and conclusions were right. This helped me when I was writing my midterm and final lab reports” (Appendix 2).

Learning outcome D was the most difficult for students to achieve. Disappointingly, but not entirely surprisingly, there was no improvement in Question 7, “identify additional information needed to evaluate a hypothesis.” While students practiced this skill in scenarios throughout the semester, they still made the error of leaving out an important experimental control toward the end of the semester, and only figured it out upon the instructor’s creation of a specific scenario (see results section, recombinant bacteriophage lysin scenarios) and significant classroom discussion. Question 4 assesses the same skill, and students scored comparatively low on that item as well, although there was a small apparent gain. Overall, this was the least developed skill at the beginning of the semester, and it remained a challenge for students even though they had as many opportunities to practice it in the class as the others. Not surprisingly, this skill falls at the highest level of Bloom’s taxonomy: create (2). This outcome points to an increased need to provide students both in college and in primary and secondary school with more practice and feedback in creative thinking skills.

The CAT data demonstrated that students who participated in the scenario activities made large significant gains that were transferable to contexts outside of the realm of biology. Admittedly, analysis of primary literature and feedback on research papers also likely contributed to the gains. However, in previous semesters, prior to implementation of the scenarios, students were far less likely to question authors’ claims in the primary literature. Students who participated in the scenarios were also far more likely to give multiple explanations for results in research papers than the test group, indicating critical thinking gains in a domain-specific manner. Additionally, the difference in critical thinking gains of students in this study compared with the Gasper and Gardner study (5) points to the use of real-world scenarios as adding significant value to an authentic research experience.

One concern regarding this work is the number of students involved in the study. Unfortunately, it is not possible to increase the number of students in this course due to a change in teaching assignment. While it would be ideal to gather data on additional students for this study, the fact that there was a large statistically significant effect size in results on some items on the CAT as well as the rubric suggests quite large gains. There was an increasing trend on many other items in the CAT for which we could not show statistical significance. It is possible that with a larger number of subjects, these gains would have been demonstrated to be significant; the small N may be underestimating the effects of the strategies used in the course.

Since the use of scenarios is more readily transferable to different course structures than either discussion of primary literature or multiple research papers, I would like to glean how much of an effect the scenarios may have independently. Future plans include engaging faculty in diverse courses (i.e., lab, no lab, multiple disciplines, different course enrollment sizes, different student demographics) in developing and implementing discipline-specific scenarios to determine how well this strategy works in various types of courses and in isolation from the primary literature reviews, experimental design activities, and written lab reports used in this course. This follow-up study will include a much larger number of students. As an institution, we are also investigating means, including the use of a common rubric, to provide students with more opportunities to practice and more opportunities for feedback on their creative thinking skills.

**SUPPLEMENTAL MATERIALS**

Appendix 1: Additional critical thinking scenarios used during the course

Appendix 2: Student responses to “Please describe the single most important assignment, strategy, or other course aspect that made a meaningful difference to your higher-order thinking abilities”
ACKNOWLEDGMENTS

Bayer CropScience Bee Care Center provided financial support for this course. Tennessee Technical University provided a training workshop for designing discipline-specific, real-world scenarios (https://www.tntech.edu/cat/training/). A potential conflict of interests associated with this paper is that the author, Dr. Susan Carson, was the developer and instructor of the course. However, scoring of the Critical Thinking Assessment Test, the instrument used for direct measure of student learning outcomes, was done at Tennessee Technological University by faculty who did not know which tests were pretests vs. posttests.

REFERENCES