The CREATE Method Does Not Result in Greater Gains in Critical Thinking than a More Traditional Method of Analyzing the Primary Literature †

Miriam Segura-Totten* and Nancy E. Dalman
Biology Department, University of North Georgia, Dahlonega, GA 30597

Analysis of the primary literature in the undergraduate curriculum is associated with gains in student learning. In particular, the CREATE (Consider, Read, Elucidate hypotheses, Analyze and interpret the data, and Think of the next Experiment) method is associated with an increase in student critical thinking skills. We adapted the CREATE method within a required cell biology class and compared the learning gains of students using CREATE to those of students involved in less structured literature discussions. We found that while both sets of students had gains in critical thinking, students who used the CREATE method did not show significant improvement over students engaged in a more traditional method for dissecting the literature. Students also reported similar learning gains for both literature discussion methods. Our study suggests that, at least in our educational context, the CREATE method does not lead to higher learning gains than a less structured way of reading primary literature.

INTRODUCTION

Over the past two decades, there has been a shift towards engaging students in scientific discovery through activities that mirror scientific research, like inquiry-based laboratories and dissection of the literature (1–3, 5, 15, 18, 23, 30, 31). These techniques may increase student learning by compelling students to examine their level of understanding through active participation (6, 36).

Actively discussing the primary literature increases students’ confidence, scientific literacy, science process skills, and helps to shape their epistemological beliefs (9, 17, 20). Different approaches exist for analyzing primary literature (9, 18, 19, 24, 32–34). While previous studies highlight the importance of reading the literature to help students understand science, it is important to test different methods of literature analysis using controlled studies. One approach whose effectiveness has not been determined in comparison to a control is CREATE (Consider, Read, Elucidate hypotheses, Analyze and interpret the data, and Think of the next Experiment) (16–18). Through a single-instructor uncontrolled study, Hoskins and colleagues (2007) posited that use of CREATE results in a heightened ability to critically analyze data and a deeper connection to science (18). The CREATE method was attractive to us because it allows students to “think like scientists.” We felt that this approach would be effective in helping students critically analyze the literature because it is relatable to the way in which science is done.

For these reasons, we tested the effects of CREATE on critical thinking compared to a more traditional method within a required cell biology course. Traditional students took a more passive approach by answering instructor-generated questions; the answers to these questions guided article discussions (Table 1). CREATE students dissected an article using diverse active learning techniques (CREATE Assignments, Table 1). Because Hoskins’ (2007) results showed that CREATE heightens critical thinking and because of the active, student-centered nature of the method when compared to the traditional method, we hypothesized that CREATE students would have higher critical thinking gains compared to traditional students. Surprisingly, our results showed that all students experienced gains in critical thinking, regardless of the literature analysis method used. Additionally, there were no significant differences in critical thinking between the CREATE and traditional groups. Our data suggest that exposing students to the primary literature leads to gains in critical thinking, regardless of whether they follow a highly structured approach like CREATE.

METHODS

Institutional and course background

The University of North Georgia (UNG) is a public, primarily undergraduate institution with an enrollment of about 15,000 students. Four sections of required course BIOL 3240 (Cell Biology; 91 students) participated in this study in fall 2010 and spring 2011. We replaced a 3-hour-per-week laboratory with article discussions (Table 1). During
the study, this course consisted of about 25% seniors, 50% juniors, and 25% sophomores. This study followed a quasi-experimental approach: during the fall of 2010, students knew the instructor but not the discussion approach for a particular section. In the spring of 2011, both sections were labeled “STAFF” during pre-registration, so students could not select a course section based on instructor.

**Article discussions**

Each semester, one section followed a traditional approach and the other followed CREATE. Both sections discussed the same articles and spent approximately the same time on-task. Instructors involved in the study taught using the traditional approach one semester and CREATE the other.

**Traditional.** The traditional method was developed based on conversations with faculty experienced in leading article discussions, and from our own experience.

Prior to the first discussion, students received the article lacking the abstract and identifying information and answered three to five questions, such as explaining how specific experiments tested the hypothesis, or how experiments related to one another, or how the results of the study support authors’ conclusions. The instructor led discussions by prompting students to provide their answers for these homework questions. At times, the instructor asked follow-up questions to further deepen the article discussion.

### TABLE 1.
Comparison of the activities associated with the discussion methods used in this study.

<table>
<thead>
<tr>
<th>Component</th>
<th>Traditional Discussions (Comparison Group)</th>
<th>CREATE Discussions (Treatment Group)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approach design</td>
<td>Based on conversations with other faculty with experience leading article discussions, and from our own experience.</td>
<td>Inspired by the CREATE (Consider, Read, Elucidate hypotheses, Analyze and interpret data, and Think of the next Experiment) method (17).</td>
</tr>
<tr>
<td>Article</td>
<td>Students received entire article except for abstract, title, and identifying information.</td>
<td>Like comparison group, but students did not receive discussion section until after the first discussion.</td>
</tr>
<tr>
<td>Assignments</td>
<td>Instructor-generated questions. First discussion: 3–5 questions about the article, including explaining how specific experiments tested the authors’ experimental hypothesis, how experiments related to one another, or how the results of the study support authors’ conclusions. Second discussion: 2–3 questions about the “big picture” of the study described in the article. Questions might ask students to propose future hypotheses and experiments to test them or apply findings to medicine or subject matter from lecture.</td>
<td>First discussion: 1) Concept map using concepts from the article’s introduction, 2) figure-by-figure diagrams of experiments performed to obtain the data presented, 3) annotated copy of the article. Second discussion: 1) A list of the authors’ conclusions, 2) a summary concept map integrating results and conclusions, 3) two follow-up experiments.</td>
</tr>
<tr>
<td>Discussion format</td>
<td><strong>First and second discussions:</strong> Instructor-led; sparked by student answers to assignment questions. First discussion: Instructor-led; sparked by student annotations. Instructor asked specific questions to guide students through discussions (e.g., what hypotheses drove a particular experiment, go through and explain a particular figure to the rest of the class, etc.). Students often worked in groups to answer challenging questions. At the end of the discussion, student groups generated conclusions and handed them in to the instructor. Second discussion: 1) Comparison of students’ and authors’ conclusions, 2) peer evaluation of student-generated conclusions, 3) discussion of follow-up experiments and selection of top three experiments, 4) discussion of importance and implications of results.</td>
<td>First discussion: Instructor-led; sparked by student annotations. Instructor asked specific questions to guide students through discussions (e.g., what hypotheses drove a particular experiment, go through and explain a particular figure to the rest of the class, etc.). Students often worked in groups to answer challenging questions. At the end of the discussion, student groups generated conclusions and handed them in to the instructor. Second discussion: 1) Comparison of students’ and authors’ conclusions, 2) peer evaluation of student-generated conclusions, 3) discussion of follow-up experiments and selection of top three experiments, 4) discussion of importance and implications of results.</td>
</tr>
</tbody>
</table>
CREATE. We adapted the original CREATE method (18) from a stand-alone upper-level course to fit into a required cell biology course (Table 2). Briefly, these were the changes made: 1) because of time limitations, students received the Introduction, Methods, and Results sections of the article at one time, instead of during separate class sessions, 2) students analyzed data without using the CREATE analysis templates (we noticed in the pilot study that students were confused by these), 3) groups of three to four students picked top experiments, instead of forming more formalized grant panels (time limitations), and 4) students did not not a suite of articles from the same research group. Instead, they read articles from different research groups. Before the first discussion, instructors lectured on formulating a hypothesis, constructing concept maps, and identifying controls. Students received detailed instructions for the article dissection process, the rubric used to grade homework (Appendix 1), and the Introduction, Methods, and Results sections of an article. To prevent students from looking up the Discussion section ahead of time, articles were scrubbed of identifying information. At the start of the first discussion, students handed in several assignments (Table 1). The instructor led discussions by asking students to identify hypotheses, explain data, and discuss potential flaws or inconsistencies in data. Students sometimes worked in groups to answer questions requiring data interpretation. At the end of the discussion, student groups generated a list of conclusions. Before the second meeting, students received the Discussion section of the article. They prepared: 1) a list of authors’ conclusions, 2) a summary concept map integrating results and conclusions into the initial concept map, and 3) two follow-up experiments. During the second discussion, they: 1) compared students’ and authors’ conclusions to determine the significance of similarities and differences between lists, 2) assigned and justified a score of 1, 2 or 3 to conclusions drafted by a different group, 3) discussed follow-up experiments and voted for the top three, and 4) discussed the importance and implications of article results. Article discussions comprised similar percentages of the class grade for traditional and CREATE sections (39% and 37%, respectively).

Assessment methods

Student materials were made anonymous by assigning each student a random number and using that identifier on materials. We used a pretest/posttest design to compare methods. Since the analysis, evaluation, and synthesis Bloom’s levels are associated with critical thinking (4, 8), we measured performance at these levels.

Exam questions. We designed two questions each at the analysis, synthesis, and evaluation levels for the first (pretest) and final (posttest) lecture exams (Appendix 2)

---

**TABLE 2.** Comparison of the original and adapted CREATE methods.  

<table>
<thead>
<tr>
<th>Original CREATE Method</th>
<th>Adapted CREATE Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Create:</td>
<td>Create:</td>
</tr>
<tr>
<td>- Students work through data as though they generated it themselves.</td>
<td>- Students work through data as though they generated it themselves.</td>
</tr>
<tr>
<td>- Students receive Introduction, Methods and Results, and Conclusions/Discussion sequentially.</td>
<td>- Students receive Introduction, Methods and Results at one time. They receive Conclusions/Discussion after in-class data analysis.</td>
</tr>
<tr>
<td>- Students create concept map using article introduction.</td>
<td>- Students create concept map using article introduction.</td>
</tr>
<tr>
<td>Read:</td>
<td>Read:</td>
</tr>
<tr>
<td>- Students work backwards from data to understand how it was obtained.</td>
<td>- Students work backwards from data to understand how it was obtained.</td>
</tr>
<tr>
<td>- Students annotate figures, create diagrams of experiments, and write descriptive titles for figures.</td>
<td>- Students annotate figures, create diagrams of experiments, and write descriptive titles for figures.</td>
</tr>
<tr>
<td>Elucidate hypothesis:</td>
<td>Elucidate hypothesis:</td>
</tr>
<tr>
<td>- Students determine hypothesis for each experiment.</td>
<td>- Students determine hypothesis for each experiment.</td>
</tr>
<tr>
<td>Analyze data:</td>
<td>Analyze data:</td>
</tr>
<tr>
<td>- Students use CREATE analysis templates to analyze data.</td>
<td>- Students are asked to analyze data in class without using CREATE analysis templates.</td>
</tr>
<tr>
<td>- Class discussion centers around data analysis.</td>
<td>- Class discussion centers around data analysis.</td>
</tr>
<tr>
<td>- Students create summary concept map.</td>
<td>- Students create summary concept map.</td>
</tr>
<tr>
<td>Think of the next experiment:</td>
<td>Think of the next experiment:</td>
</tr>
<tr>
<td>- Students design follow-up experiments.</td>
<td>- Students design follow-up experiments.</td>
</tr>
<tr>
<td>- Student grant panels deliberate on experiments.</td>
<td>- Students pick the top experiments and provide rationale for their choices without using grant panels.</td>
</tr>
</tbody>
</table>

*a Assignments are arranged according to the aspect of the CREATE method that they target (according to 17). Italicized text denotes deviations from the original method.*
using the Blooming Biology Tool (BBT; 8). One question from each category was included in each exam. Miriam Segura-Totten (MST) received training on designing questions from Mary Pat Wenderoth (University of Washington), one of the developers of the BBT (8) and she then trained Nancy E. Dalman (NED). To measure question difficulty, we asked three biology faculty members who teach or do research in the question topic and three past cell biology students to rate questions as easy (1), medium (2), or difficult (3). We looked at average scores across categories: expert averages for pre- and posttest questions are 2.33 and 2.33, while student averages are 2 and 1.67. Identical questions were used in all study sections. Questions were graded using a detailed rubric developed by both instructors. Each instructor scored her particular section. The rubric was revised during a norming session based on the range of student answers collected. Questions were rescored using the revised rubric.

**Article critique exercise.** Students in traditional and CREATE groups did pre- and post-critique exercises on weeks four and fourteen of the semester, respectively (Table 3). A week before the exercise, students received the Introduction, Methods, and Results sections of an article scrubbed of identifying information. The articles used for critiques were different from those used in discussions. To avoid advance preparation, students did not receive instructions for the article critique (Appendix 3) before completing the exercise. Students worked individually for two hours to complete the critique. We designed a scoring instrument based on student instructions. Elements of the instrument were categorized according to the level of Bloom’s targeted.

**Scoring of the critiques.** In the fall of 2010, instructors met for a norming session, where they resolved concerns or disagreements regarding the scoring instrument. Instructors scored critiques following the resulting modified instrument (Appendix 4). While each instructor graded critiques for her section, instructors graded the fall 2010 critiques together, to ensure uniformity in scoring. Since instructors were confident with scoring in the spring of 2011, assignment of points was done individually, with each instructor scoring her own section. To normalize questions with varying point values, student scores for each category were divided by the total possible number of points, resulting in a scale of 0–1. Normalized scores were used for statistical analysis. To determine the reliability of the scoring rubric, thirty-two critiques were scored by the instructors and an independent rater blinded to the experimental condition. Strong interrater reliability was demonstrated, ICC(2,2) = 0.753 (p < 0.001), suggesting that scores obtained represent student performance that is not rater-specific.

We chose articles for the pre- and post-critique that contained techniques and concepts familiar to students, were similar in reading difficulty, and had similar numbers of figures. To compare the readability of the articles, we used the Flesch Reading Ease (FRE) and Flesch-Kincaid Grade Level (FGL) analyses in Microsoft Word. Both tests evaluate readability by looking at average sentence length and syllables per word (11). The FGL is based on FRE but rates text on a U.S. school grade level; a rating of 8 means that an eighth grader should understand the document (22). In the fall of 2010, pre- and post-articles had a FRE score of 20.6 and 20.4, and a FGL score of 11.6 and 12.0, respectively. In the spring of 2011, pre- and post-articles had a FRE score of 35.8 and 31.7, respectively, and an identical FGL score of 12.0. Additionally, two colleagues whose research specialties include cell biology determined that, for each semester, the two articles chosen were similar in difficulty. Importantly, we did not tell our colleagues what article would be used as a pretest or posttest.

**Student equivalence across semesters**

Because we wanted to compare student performance across semesters and sections, we determined whether students came into the course with similar critical thinking skills. To do this, we performed a 2(pretest vs. posttest) × 2(control vs. experimental group) ANCOVA on each measure of critical thinking while controlling for student GPA. This analysis showed that students in the traditional group

<table>
<thead>
<tr>
<th>Week</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NO CLASS (classes start mid-week, so one section does not meet)</td>
</tr>
<tr>
<td>2</td>
<td>Overview of paper reading process and discussion format, consent forms, students asked to research specific cell biology techniques.</td>
</tr>
<tr>
<td>3</td>
<td>Discussion of cell biology techniques, what is a hypothesis, experimental controls. CREATE only: concept maps.</td>
</tr>
<tr>
<td>4</td>
<td>Article critique #1</td>
</tr>
<tr>
<td>5</td>
<td>Article Discussion #1 part 1</td>
</tr>
<tr>
<td>6</td>
<td>Article Discussion #1 part 2</td>
</tr>
<tr>
<td>7</td>
<td>Article Discussion #2 part 1</td>
</tr>
<tr>
<td>8</td>
<td>Article Discussion #2 part 2</td>
</tr>
<tr>
<td>9</td>
<td>Article Discussion #3 part 1</td>
</tr>
<tr>
<td>10</td>
<td>Article Discussion #3 part 2</td>
</tr>
<tr>
<td>11</td>
<td>NO LAB – SPRING BREAK</td>
</tr>
<tr>
<td>12</td>
<td>Article Discussion #4 part 1</td>
</tr>
<tr>
<td>13</td>
<td>Article Discussion #4 part 2</td>
</tr>
<tr>
<td>14</td>
<td>Article critique #2</td>
</tr>
<tr>
<td>15</td>
<td>NO CLASS (classes end mid-week)</td>
</tr>
</tbody>
</table>
did not differ from the CREATE group on their pretest critique scores, \( t(82) = 0.527, p = 0.599 \), or exam scores, \( t(82) = 1.174, p = 0.244 \).

**Data analysis**

Both instructors involved in the study taught the course using the traditional approach one semester and CREATE the other. This allowed differences between instructors and teaching methods to be separated statistically. Results were analyzed with a four-way mixed analysis of variance (variables: teaching method, instructor, pretest vs. posttest, Bloom’s category). An ANCOVA on student scores further controlled for student GPA.

**SALG coding**

Using the SALG survey (http://www.salgsite.org/), we asked students to comment on their present level of interest in the subject, how the class heightened or dampened their interest, how material in the course integrates with their studies, career, and life, and the skills they developed by the end of the course. NED and MST first independently determined salient themes in student answers for both sections. They met to refine the list of themes and compare their coding. This led to a single list of themes (scientific literacy, confidence, nature of science, transfer of critical thinking skills, and critical thinking) and one group of coded student answers. Surveys were stripped of information pertaining to section and semester prior to analysis.

**Institutional review board protocols**

Approval to conduct this study (exempt status, application 201023) was granted by the Institutional Review Board, UNG.

**RESULTS**

**Discussion of the literature improves students’ ability to critique articles**

Because we defined critical thinking as higher-order cognitive skills necessary for analysis and evaluation of data, as well as synthesis of information to create new knowledge or inferences (8, 38), we designed assessments that measure student performance in analysis, synthesis, and evaluation (4, 8). We first assessed student performance in an article critique. Students in the CREATE group showed a 65% increase in performance while traditional students showed a 40% increase (Fig. 1). A mixed 4-way ANOVA of scores combined across categories showed that scores increased significantly at the end of the semester for all groups on the critique, \( F = 2.931, p = 0.036 \). Figure 2 shows similar trends for students in traditional and CREATE groups: analysis and synthesis skills increased, while evaluation did not.

While we did not detect a significant interaction between Bloom’s category and pretest/posttest differences in scores by method (\( F = 1.456, p = 0.236 \)), we wanted to look at trends for student learning in particular categories. Figure 2 shows similar trends for students in traditional and CREATE groups: analysis and synthesis skills increased, while evaluation did not.

We also designed exam questions at the analysis, synthesis, and evaluation levels in the first (pretest) and final (posttest) exams. While CREATE students showed a modest increase in performance compared to traditional students (Fig. 3), this gain is not significant (\( F = 2.887, p = 0.093 \)). Additionally, scores for all students increased significantly at the end of the semester, regardless of method used (\( F = 4.190, p = 0.044 \)). Because the increase in performance was so modest and because there was no significant difference between Bloom’s categories (\( F = 1.290, p = 0.278 \)), we did not examine trends in individual categories.

**Grade point average does not influence student critical thinking gains**

To determine if GPA influenced student performance, we performed a 2(pretest vs. posttest) \( \times \) 2(traditional vs. CREATE) ANCOVA on exam and critique scores while controlling for GPA. Students’ scores increased significantly from pretest to posttest on both the critique, \( F(1,81) = 8.113, p = 0.006, \eta^2 = 0.091 \), and the exam questions, \( F(1,81) = 364.20, p < 0.001, \eta^2 = 0.818 \), even when the effect of GPA was controlled for. However, there was no significant difference in improvement between CREATE and traditional groups on the critique, \( F(1,81) = 0.006, p = 0.938, \eta^2 < 0.001 \), or the exam, \( F(1,81) = 0.106, p = 0.746, \eta^2 = 0.001 \). Therefore, significant learning gains on the critique from pretest to posttest did not differ between the traditional (\( M = 0.25 \), \( SE = 0.02 \) and \( M = 4835.21, SE = 412.23 \), respectively) and...
CREATE groups (M = 0.25, SE = 0.02 and M = 4789.91, SE = 412.23, respectively). Likewise, the significant learning gains on the exam questions from pretest to posttest did not differ between the traditional (M = 0.55, SE = 0.03 and M = 3.10, SE = 0.07, respectively) and CREATE groups (M = 0.69, SE = 0.03 and M = 3.21, SE = 0.07, respectively).

**Student perceptions of gains**

Students received an end-of-semester survey asking them to comment on: present level of interest in the subject, how the class heightened or dampened their interest, how they felt the material in the course integrates with their studies, career, and life, and the skills they developed by the end of the course. Students also rated their understanding, skills, attitudes, and ability to integrate knowledge. Four-way ANOVA analysis of scores for questions related to critical thinking (Appendix 5) did not show a difference between students in traditional and CREATE groups (F = 0.148, p = 0.863).

In the traditional group, 83% of students gave positive comments about their current interest in the subject, and 86% of students said the class heightened their interest (Fig. 4). In the CREATE group, 61% of students gave positive
comments regarding their level of interest, while 65% felt the class heightened their interest (Fig. 4). The difference in positive comments between groups was made up mostly by students with neutral feelings (Fig. 4). Both sections had low numbers of students with negative feelings.

We then coded student answers to determine perceptions of learning gains. Both groups felt they were more critical when looking at information (Critical thinking, red bar, and Transfer of CT skills, green bar, Fig. 5). However, more traditional students felt they had increased critical thinking than CREATE students (16 vs. 11, Skills gained, red bar, Fig. 5).

Slightly more traditional students felt they gained scientific literacy skills (12 vs. 9, Skills gained, blue bar, Fig. 5), while slightly more CREATE students felt confident at the end of the course (10 vs. 8, Skills gained, orange bar, Fig. 5). Traditional students mentioned having a better understanding of the process of science more often than CREATE students (purple bars in Fig. 5). Interestingly, 50% more of the CREATE students directly mentioned feeling they could read and analyze literature more proficiently than traditional students (12 vs. 8, Skills gained, blue bar, Fig. 5).

Table 4 contains examples of student comments.

DISCUSSION

To our knowledge, this is the first study to perform a controlled, multi-instructor comparison of learning gains from CREATE. Because a previous single-instructor, “pre-experimental” (i.e., lacking a control group) study of CREATE revealed that this method increases critical thinking (18), we hypothesized that use of CREATE in our context would lead to gains in critical thinking skills beyond those of students engaged in a more traditional method.

CREATE and traditional discussions lead to similar gains

Our findings reveal that, while discussion of the literature helps to develop critical thinking, CREATE does not lead to significant gains over a more traditional method, at least in our educational context. While it is possible that our adaptation of CREATE dampened development of critical thinking, the changes we made were minor and unlikely to drastically reduce student learning (Table 2; see Methods for a description of changes). One notable difference between the original and adapted CREATE methods is that our students did not read a suite of articles from the same research group. While this might diminish students’ ability to see the progression of a research project, our assessments did not require students to appreciate the continuity of research. Thus, it is unlikely that this modification had a marked effect on students’ performance in our study.

Importantly, the CREATE method is not a prescribed course—rather it is a combination of pedagogical tools, which our adapted method also utilized (16, 18; Table 2). This adds validity to the comparison between our adapted method and the traditional discussions. It is worth noting that Hoskins and colleagues (2007) used a different assessment of critical thinking from the one we used. It is possible that this led to the difference in results. However, because our definition of critical thinking is widely used (8, 38) and effectively describes skills necessary for scientific thought, our results contribute important information to what is currently known about CREATE.

Why does CREATE not lead to improved learning when compared to traditional discussions? One reason might be that CREATE uses tools like concept maps and diagramming to prompt “meaningful learning,” where the student must reflect on the information provided to actively incorporate concepts into his or her knowledge structure (10, 27). It is
conceivable that traditional students also had to practice meaningful learning to answer questions at higher levels of Bloom’s, since these questions require higher order cognitive reasoning. Additionally, CREATE utilized student-centered tools that compel students to take control of the article analysis process. These inquiry-based and student-centered tools can help students become more sophisticated thinkers (21). It is possible that students in the traditional group also had to independently navigate the direction of article analysis to answer complex questions, thus becoming more sophisticated in the way they approach the literature. It would be interesting to determine if crafting discussion questions at lower levels of Bloom’s would negate the gains that the traditional group experienced.

We observed greater gains in the article critique than in exam questions. One potential reason for this is the practice effect inherent in the repeated measures design for critiques. We attempted to reduce the possibility of practice effect in our study by: showing students the instructions for the critique only while they worked on it; having students only perform the critiques once on week 4 and then on week 14; and ensuring that discussions did not include components that mimicked the critique. Therefore, we believe that improvements because of practice effect are unlikely. Another explanation for the lower gains in exam questions is that many other variables may affect student performance on an exam. Performance at higher levels of Bloom’s is dependent on the lower levels of knowledge and comprehension. Thus, if students did not understand a concept that was used in our assessment, or if they did not study a concept, they may not have performed well. On the other hand, the article critique is knowledge-independent, because students had access to the critique article during the exercise, and they could ask the instructor basic conceptual or vocabulary questions.

Interestingly, students in CREATE and traditional groups had similar perceptions of their learning gains. This reflects the results of our two direct assessments of critical thinking. However, CREATE students were more confident about their abilities than traditional students, which mirrors the gains in confidence previously reported (17). This boost in confidence may be beneficial in other aspects of students’ college careers.

### Reading the literature as scientific inquiry

Engaging students in scientific discovery leads to important learning gains (5, 6, 12–14, 18, 30, 31, 36). However, the question of how reading the scientific literature fits into the student experience remains. Is reading the primary literature akin to scientific inquiry? Should we take time away from other educational experiences to focus on the dissection of primary sources? Several groups argue that student immersion in scientific reading mimics the process of science (18, 29, 37). Given the amount of time scientists spend reading literature (35), and the importance of understanding the literature for the process of science (25), an argument can be made for cultivating scientific literacy. If we strive to forge individuals who can make informed decisions after analyzing scientific data, then engaging them in activities that develop critical thinking and scientific literacy is of the outmost importance.

Our study shows that article discussions improve students’ ability to think critically. Given the value that faculty,

### TABLE 4.
Sample student comments on the SALG survey.

<table>
<thead>
<tr>
<th>Question</th>
<th>Sample Student Comment</th>
</tr>
</thead>
</table>
| Change in interest     | “I am aware of what will happen in future biology courses. I will be more prepared when having to do research from scientific papers.”  
“I am also able to tell if the article is a good one or if I need to look for another one.”  
“Thanks to critically thinking and analyzing while reading articles in cell biology topics has increased my interest in science a great deal!” (sic).  
“It heightened it, because through the articles […] we could see cellular biology being researched and the knowledge gained in the class applied in the actual world.” |
| Present level of interest | “I really liked cell biology, but mainly because it’s just as cool as all the other biology classes we take. I just like it all, really.”  
“My current interest is high, such that I would like to read more articles pertaining to cell biology if I had the time.” |
| Integration of material | “What can I say? I’m more critical!”  
“It doesn’t.”  
“Study wise, I feel like I am better prepared for undergraduate research. There are many articles that I have to read for my undergrad research to prepare me for it. I am now able to tell if the article is a good one or if I need to look for another one.”  
“It has been useful in other biology classes.”  
“The material discussed in this class has tied many things together from other subjects and things that come up even at work.” |
| Skills learned         | “Efficiently critique a lab article.”  
“Understand scientific journal articles.”  
“I didn’t learn anything in this course, to be honest.”  
“Scientific papers intimidated me and I avoided them at all costs. While taking this lab, I learned how to actually READ, understand and evaluate [articles]. I learned how to evaluate the experimental procedures and interpret data and figures.”  
“I feel as though I’m better at solving complicated conceptual problems after cell.” |
educational organizations, and the government place on thinking critically as an educational outcome for STEM majors (2, 7, 23), and given that high school and college students show weak interpretation skills (25, 26, 28), including literature discussions within a required course is an effective way to improve STEM education. While we did not see significant differences in critical thinking between CREATE and traditional students, a preexperimental study of CREATE suggested that this approach leads to a shift in students’ attitudes about science and their epistemological beliefs (17). For this reason, we suggest that instructors carefully weigh the learning goals for their course against the time investment required by the particular literature discussion method.

Concluding remarks

We show that students engaged in literature discussions within a required cell biology course achieve gains in critical thinking in one semester. It will be interesting to investigate if these gains persist or increase in subsequent semesters. As others have described previously (17), CREATE students in our study also reported gaining scientific literacy skills, and they were more confident than traditional students about their ability to read the literature. It remains to be seen if students who engage in discussing the literature develop other facets of scientific literacy, such as understanding the implications of science, knowing what counts as science, the ability to participate in science-based social issues, and knowledge of the risks and benefits of science (25).

SUPPLEMENTAL MATERIALS

Appendix 1: Student instructions. Prior to the start of the CREATE discussions, students received detailed instructions that guided them through the process of dissecting the literature (A). Students also received the rubric that instructors used to grade their assignments prior to the due date for the first set of homework. The instructors gave back the graded homework along with the completed rubric (B).

Appendix 2: Sample assessment questions. Bloom’s level for each question is specified in italics. Question #2 was adapted from the “Test Bank for Biology,” fifth edition (Campbell, Reece, and Mitchell), Daniel Wivagg, editor, ©1999, Menlo Park, CA.

Appendix 3: Student instructions for completing the article critique.

Appendix 4: Instrument used to score student article critiques.

Appendix 5: Questions that were used in the student end-of-semester survey. Survey was designed using the SALG template (http://www.salgsite.org).

ACKNOWLEDGMENTS

We thank Frank Corotto and Steve Lloyd for help with statistical analyses, and Tom Nelson for early discussions that helped shape the study. We are thankful to Jenny Knight and Mary Pat Wenderoth for stimulating discussions and critical reading of the manuscript, and to the American Society for Microbiology’s Biology Scholars Program for help with study design. We are grateful to Peggy Brickman and the University of Georgia’s Biology Educators group for their advice and guidance. This study was funded in part through the Department of Biology and the Harry B. Forester Fund – NGCSU Foundation. The authors declare that there are no conflicts of interest.

REFERENCES