A Modified CREATE Intervention Improves Student Cognitive and Affective Outcomes in an Upper-Division Genetics Course†

Stanley M. Lo1,2,3*, Tiffany B. Luu1†, and Justin Tran1‡

1Division of Biological Sciences, University of California San Diego, La Jolla, CA 92093; 2Section of Cell and Developmental Biology, University of California San Diego, La Jolla, CA 92093; 3Program in Mathematics and Science Education, University of California San Diego, La Jolla, CA 92093

Many national reports have called for undergraduate biology education to incorporate research and analytical thinking into the curriculum. In response, interventions have been developed and tested. CREATE (Consider, Read, Elucidate the hypotheses, Analyze and interpret the data, and Think of the next Experiment) is an instructional strategy designed to engage students in learning core concepts and competencies through careful reading of primary literature in a scaffolded fashion. CREATE has been successfully implemented by many instructors across diverse institutional contexts and has been shown to help students develop in the affective, cognitive, and epistemological domains, consistent with broader meta-analyses demonstrating the effectiveness of active learning. Nonetheless, some studies on CREATE have reported discrepant results, raising important questions on effectiveness in relation to the fidelity and integrity of implementation. Here, we describe an upper-division genetics course that incorporates a modified version of CREATE. Similar to the original CREATE instructional strategy, our intervention’s design was based on existing learning principles. Using existing concept inventories and validated survey instruments, we found that our modified CREATE intervention promotes higher affective and cognitive gains in students in contrast to three comparison groups. We also found that students tended to underpredict their learning and performance in the modified CREATE intervention, while students in some comparison groups had the opposite trend. Together, our results contribute to the expanding literature on how and why different implementations of the same active-learning strategy contribute to student outcomes.

INTRODUCTION

In the past few decades, many national reports have called for the transformation of undergraduate biology education. Following broad calls in Science for All Americans (1) and Reinventing Undergraduate Education (2), recommendations specifically for the biological sciences began to emerge: engage students in the excitement of discoveries (3); help students develop core concepts and competencies (4, 5); promote active learning in the classroom and incorporate research into the curriculum (6), especially for students across diverse undergraduate educational contexts (7, 8). These recommendations were echoed and paralleled by The Next Generation Science Standards in K–12 education (9). As a result, many educational interventions have been developed across science, technology, engineering, and mathematics (STEM) disciplines and their efficacies examined (10–13).

One instructional strategy developed in the biological sciences to achieve these recommendations is CREATE (Consider, Read, Elucidate the hypotheses, Analyze and interpret the data, and Think of the next Experiment) (14). CREATE builds upon research on how people learn and incorporates pedagogical tools that facilitate conceptual learning (15–20). Students develop core concepts and competencies in biological sciences through reading discussing primary literature and by systematically decoding figures, tables, and narratives in a scaffolded fashion (14, 21–25). CREATE has been successfully implemented by faculty across many institutional contexts (26) and has been shown to promote affective, cognitive, and epistemological development of diverse two- and four-year students (27–31).

A number of other instructional strategies have been developed to engage students in reading and understanding primary literature in the biological sciences. For example, Research Deconstruction engages students in dissecting...
Our modified CREATE intervention course had a limited enrollment of 48, equal to the enrollment of the small comparison course. Response rates are reported separately for CI,TOSLS (intervention only), affective survey and course evaluation. CI = concept inventory items; TOSLS = test of scientific literacy skills.

| TABLE 2. | Intervention and comparison groups. |
|----------|-------------------------------------|
| Intervention | CREATE | Small | Medium | Large |
| Enrollment | 48 x 2 = 96 | 48 | 48 | 48 |
| Response rate | CI: 73% | TOSLS: 90% | Survey: 85% | Evaluation: 87% |
| | CI: 75% | TOSLS: 90% | Survey: 71% | Evaluation: 90% |
| | CI: 66% | TOSLS: 90% | Survey: 74% | Evaluation: 59% |
| | CI: 67% | TOSLS: 90% | Survey: 73% | Evaluation: 58% |

Comparison groups

Multiple sections of the genetics course are typically offered within the same academic term at the study institution, and they were used as quasi-experimental comparison groups. Specifically, three other sections (n = 48, 148, and 356 students, respectively) were offered at the same time as the second iteration of the modified CREATE intervention course (Table 2). All three comparison sections had the traditional lecture-and-practice structure without the intervention. Each section had limited enrollment at the same number as the intervention course, and the other two had larger enrollments. Students self-selected to be enrolled in the CREATE intervention and the different comparison courses.
Cognitive outcomes

Cognitive learning of genetics concepts was measured using selected items from existing concept inventories customized for the course material, given to students at the beginning and end (post) of a specific genetic term in the intervention and comparison courses. Sixteen items were chosen from established concept inventories in the literature (58, 59) by a different instructor who typically teaches the same course in another academic term and was not involved in the intervention or comparison courses (Appendix 2). In our dataset, the 16 items have a Cronbach’s alpha of 0.74. (In this paper, we will refer to these 16 selected concept inventory items collectively as “CI.”) While there are no universal guidelines for interpreting Cronbach’s alpha values, this value falls within the adequate-to-good range of reliability (60).

The Test of Scientific Literacy Skills (TOSLS) was given only pre- and post-course in the intervention course to measure potential changes in students’ ability to evaluate scientific information and arguments (61). TOSLS was given only in the intervention course, as the administration of the test would have taken too much class time in the comparison courses. In our dataset, the 28 TOSLS items have a Cronbach’s alpha of 0.94, indicating high or excellent reliability (60).

Pre- and post-course CI were analyzed across the intervention and comparison courses using a standard two-way analysis of variance (ANOVA) with a post-hoc Tukey’s honestly significant difference (HSD) test for multiple comparisons. For clarity, p values are only reported for comparisons between the intervention course and each of the other courses to establish baseline comparisons and for pre- versus post-course CI scores to determine statistical outcomes for learning gains. TOSLS scores were analyzed using a standard Student’s t-test. Effect sizes were calculated using Cohen’s d to measure the difference between the pre- and post-course means normalized to the standard deviation from the pre-course data. All statistical analyses in this study were performed in Microsoft Excel 2016 and JMP Pro version 13.0.

Perceived learning and course outcomes

To compare cognitive outcomes (measured by CI scores) with students’ perception of their own learning, we used the institution’s course evaluation data. Students rated the statements: “I learned a great deal from this course” on a standard 1-5 Likert scale (from “almost never” to “almost always”) (62). The CI score is a measure of learning process; the items were on a five-point, Likert-like scale. In our dataset, the 16 items have a Cronbach’s alpha of 0.74. (In this paper, we will refer to these 16 selected concept inventory items collectively as “CI.”) While there are no universal guidelines for interpreting Cronbach’s alpha values, this value falls within the adequate-to-good range of reliability (60).

The Test of Scientific Literacy Skills (TOSLS) was given only pre- and post-course in the intervention course to measure potential changes in students’ ability to evaluate scientific information and arguments (61). TOSLS was given only in the intervention course, as the administration of the test would have taken too much class time in the comparison courses. In our dataset, the 28 TOSLS items have a Cronbach’s alpha of 0.94, indicating high or excellent reliability (60).

To establish baselines for learning gains, we compared the pre-course CI scores between the modified CREATE intervention and each of the comparison courses. Students in the small and large comparison courses had a statistically equivalent cognitive baseline (measured by pre-course CI scores) to that of students in the modified CREATE intervention (ANOVA, p > 0.9999 and p > 0.95 respectively), and students in the medium comparison course had a higher baseline pre-course CI score on average (ANOVA, p < 0.001). While baseline test scores alone do not eliminate all possible biases in students self-selecting to enroll in the CREATE intervention and the different comparison courses, they are the best predictors for student outcomes in K–12 science education research (63). Therefore, the small comparison course is a suitable statistically equivalent comparison (66), given the equivalent course size and pre-course CI scores. The other two comparison courses, while not fully statistically equivalent, are included as additional data source triangulation.

To compare cognitive learning outcomes across the intervention and comparison courses, we examined pre- and post-course CI scores (Fig. 2A, left y-axis). For the intervention, CI scores were higher (ANOVA, p < 0.0001) at the end of the course (average ± standard deviation = 9.54 ± 3.31) compared with the beginning of the course (5.93 ± 2.40) with an effect size of 1.51 (Fig. 2B). In contrast, the comparison courses (small, medium, and large) had statistically significant effect sizes of 0.87, 0.56, and 0.46 respectively, all of which were lower than the effect size in the intervention course. Furthermore, we used TOSLS to examine changes in students’ ability to evaluate scientific information and arguments in the intervention course (Fig. 2B). TOSLS scores were higher (t-test, p < 0.0001) at the end of the course (19.7 ± 4.1) than at the beginning of the course (17.3 ± 3.0), with an effect size of 0.78, which is considered large (67). We evaluated perceived learning in relation to the cognitive outcomes measured by CI scores (Fig. 2A, right versus left y-axis). Students reported a lower perceived learning score in the intervention course (3.66 ± 1.19) versus the three comparison courses (small, medium, and large), which had perceived learning scores of 4.28 ± 0.91 (ANOVA, p < 0.001), 4.27 ± 1.00 (ANOVA, p < 0.001), and 4.23 ± 0.73 (ANOVA, p < 0.001), respectively. In contrast to the comparison courses, students in the intervention course reported statistically lower perceived learning, even though the effect size in CI scores showed much higher cognitive learning gains. In addition, we compared students’ expected versus actual grades (Fig. 2C). Consistent with the perceived learning and CI results, we found that students in the intervention course underpredicted how well they had done (predicted average grade point out of 4.00 = 3.12, actual = 3.36; Fisher’s exact test, p < 0.05), whereas students in the comparison courses either made accurate predictions

Percentage of class time with multiple voices from the DART profiles is tabulated across the three types of classes in the intervention: CREATE only; interactive clicker lecture, and mixed.
Affective outcomes

We compared affective outcomes across the intervention and comparison courses (Fig. 3). In all six affective dimensions measured, students in the intervention had statistically higher outcomes than at least two of the three comparison courses. In contrast to the small and large comparison courses, students in the intervention reported higher personal relevance of the subject matter (ANOVA, p < 0.05 and p < 0.0001, respectively) and had more comfort with the uncertainty of science (ANOVA, p < 0.01 and p < 0.0001, respectively), a core aspect of the nature of science (5, 62). Students in the intervention reported having more critical voice and shared control in the classroom, as well as more peer negotiation and affective support in the learning process, than students in all three comparison courses (ANOVA, p at least < 0.05).

DISCUSSION

In this study, we showed that a modified CREATE intervention in an upper-division genetics course led to improved student outcomes. Triangulation across multiple sources of data (CI, survey, and course evaluation) provides potential explanations and mechanisms for why we observed these results. For example, a higher survey score on peer negotiation (in the modified CREATE intervention versus the comparison courses) indicates that students perceived more interactive discussions, which are critical for learning (16, 20); correspondingly, we observed higher learning gains as measured by pre- and post-course CI scores. Somewhat ironically, the higher score on peer negotiation could also explain why students in the modified CREATE intervention tended to underpredict their learning and performance. Students typically hold a combination of teacher-centered and knowledge-centered conceptions of teaching and learning, where the instructors are expected to clearly present the correct knowledge through examples, in contrast to a student-centered conception of teaching and learning.
First, we observed that students in the modified CREATE intervention on average underpredicted their learning and course performance, whereas in at least one of the comparison courses, students on average overpredicted their learning and course performance. This pattern is consistent with the existing literature on interleaved versus blocked learning (72) and productive failure (55) and can be explained by the lack of metacognitive awareness on the effectiveness of different learning strategies (73). Thus, our results argue—as do previous studies in the literature—that it is important to measure both student self-reported perceived learning and actual cognitive learning gains to triangulate findings in a research study.

Second, the use of different outcome measurements can lead to varying results across studies that make comparisons difficult. (13). Even though a previous study observed no difference in cognitive gains in critical thinking between CREATE and a comparison (36), we do not necessarily see our results being contradictory to that study. This is in part because the two studies had different outcome measurements, as we did not measure critical thinking. Our comparison courses also did not explicitly engage students in primary literature in an extensive fashion, thus potentially widening the difference between the modified CREATE invention and the comparisons. When situated in the context of other existing studies, our work contributes to the expanding literature on how and why different implementations of the same active-learning strategy contribute to student outcomes.

Finally, recent studies have shown that instructors who report using active-learning strategies often do not use them as suggested or originally designed by researchers (74, 75), which may lead to different student outcomes (76). Therefore, it is critical to align the intended curriculum (designed by the researchers based on learning principles) and the enacted curriculum (implemented in different educational contexts by practitioners). More broadly speaking, the intended and enacted curricula, as well as what students learn, can often be substantially different (77, 78). In this study, our modified CREATE intervention (intended curriculum) was designed with a course structure that is supported by existing literature on how people learn, and we observed the implementation (enacted curriculum) using DART. Implementing established active-learning strategies with integrity based on fundamental learning principles can have an important positive impact on student outcomes; this is in contrast to directly copying the instructional strategy (high fidelity) without adapting it to the local educational context (39). The increasing calls for widespread and large-scale implementation of active learning across STEM disciplines (11) raise important questions for how best to support the professional development of current and future faculty (79) and what mechanisms of propaganda would be best for ensuring the integrity of implementation (40, 80).

where students actively participate in the learning process by providing feedback to one another (66). It is conceivable that peer negotiation is inconsistent with what students may perceive to be effective teaching and learning, thus resulting in a lower predicted level of learning.

In our modified CREATE intervention, students reported more comfort with the uncertainty of science, which is a key aspect of the nature of science (5, 62). Our results are consistent with a previous study showing significant changes in students’ self-assessed understanding of the nature of science from an upper-division course taught using the original CREATE instructional strategy (29). In addition, higher personal relevance, critical voice, shared control, and affective support would likely result in increased persistence in biological sciences majors, especially for underrepresented minority or first-generation college students, by potentially providing additional information. More importantly, one of the comparison courses was the same size as the intervention course and this comparison course had equivalent pre-course CI scores. We were also not able to fully disentangle the effect of course size on student outcomes in the comparison courses; however, the main goal of the study is to compare the modified CREATE intervention with other active-learning strategies. Despite these limitations, our study highlights important considerations that are of interests to the growing field of biology education research and practice.

LA et al.: CREATE- AND ARGUMENTATION-BASED GENETICS

FIGURE 3. Affective outcomes. Results on the six affective dimensions from our survey are plotted: (A) personal relevance, (B) uncertainty of science, (C) critical voice, (D) shared control, (E) peer negotiation, and (F) affective support. Error bars indicate standard deviation, and statistical differences (by two-way ANOVA) are indicated by brackets and the following notation: * p < 0.05; ** p < 0.01; *** p < 0.001; **** p < 0.0001.

**REFERENCES**

1. American Association for the Advancement of Science. 1989. Science for all Americans. Oxford University Press, New York, NY.
2. Boyer Commission. 1998. Reinventing undergraduate education: a blueprint for America’s research universities. Commission on Educating Undergraduates in the Research University, Stoney Brook, NY. https://eric.ed.gov/?id=ED424840
3. National Research Council. 2003. BIO2100: Transforming undergraduate education for future research biologists. The National Academies Press, Washington, DC.
4. Association of American Medical Colleges, Howard Hughes Medical Institute. 2009. Scientific foundations for future physicians. Washington, DC.
5. American Association for the Advancement of Science. 2011. Vision and change in undergraduate biology education: a call to action: a summary of recommendations made at a national conference organized by the American Association for the Advancement of Science, July 15–17, 2009. Washington, DC.
6. President’s Council of Advisors on Science and Technology (PCAST). 2012. Engage to excel: producing one million additional college graduates with degrees in science, technology, engineering, and mathematics. Executive Office of the President, Washington, DC.
7. National Academy of Sciences, National Academy of Engineering, and Institute of Medicine. 2011. Expanding...
underrepresented minority participation: America's science and technology education: the Mylonas. The National Academies Press, Washington, DC.

10. National Academies of Sciences, Engineering, and Medicine. 2016. Barriers and opportunities for 2-year and 4-year STEM degrees: system change to support student diverse pathways. The National Academies Press, Washington, DC.

11. National Research Council. 2012. A framework for K-12 science education: Practices, crosscutting concepts, and core ideas. The National Academies Press, Washington, DC.

12. Bowen CW. 2000. A quantitative literature review of student performance in science, engineering, and mathematics. Proc Natl Acad Sci 97:1273–1279.

13. Ruiz-Primo MA, Briggs D, Iverson H, Talbot R, Shepard LA. 2016. Introducing conceptualizing concepts and core ideas. The National Academies Press, Washington, DC.

14. Bowman CW, Martin JD. 2008. A diagnostic assessment for introductory biological and molecular cell biology. CBE Life Sci Educ 7:436–441.

15. Tiber SB. 2018. The use of Cronbach’s alpha when developing and reporting research instruments in science education. Res Sci Educ 48:1273–1296.

16. National Academies of Sciences, Engineering, and Medicine. 2014. Practice makes pretty good: assessment of primary school mathematics. The National Academies Press, Washington, DC.

17. National Research Council. 2000. How people learn: brain, mind, experience, and school, expanded ed. The National Academies Press, Washington, DC.

18. Piaget J. 1973. To understand is to invent: the future of mental process. Harvard University Press, Cambridge, MA.

19. National Academies of Sciences, Engineering, and Medicine. 2014. How people learn II: learners, contexts, and cultures. The National Academies Press, Washington, DC.

20. Nix RK, Fraser BJ, Ledbetter CE. 2005. Evaluating the fidelity of implementation of research-based instructional strategies. Instr Sci 33:219–243.

21. Hoskins SG. 2008. Using a paradigm shift to teach scientific thinking, a new course for STEM-intended freshmen, demystifying scientific thinking through analysis of scientific literature. CBE Life Sci Educ 7:125–139.

22. Pieper CW. 2004. Does active learning work? A review of the research. J Eng Educ 93:223–231.

23. Stevens LM, Hoskins SG. 2014. The CREATE strategy for introducing scientific thinking, a new course for STEM-intended freshmen, demystifying scientific thinking through analysis of scientific literature. CBE Life Sci Educ 13:219–231.

24. Sisco BK, Kandukuri V, Murayama MH, Wisher M, Wunschauer M. 2014. Practice makes pretty good: assessment of primary literature reading abilities among large multi-enrollment biology laboratory courses. CBE Life Sci Educ 13:677–686.

25. Taylor K, Rohrer D. 2010. The effects of interleaved practice. Appl Cogn Psychol 24:355–367.

26. Taber KS. 2018. The use of Cronbach’s alpha when developing and reporting research instruments in science education. Res Sci Educ 48:1273–1296.

27. Taylor K, Rohrer D. 2010. The effects of interleaved practice. Appl Cogn Psychol 24:355–367.

28. Taber KS. 2018. The use of Cronbach’s alpha when developing and reporting research instruments in science education. Res Sci Educ 48:1273–1296.

29. Taber KS. 2018. The use of Cronbach’s alpha when developing and reporting research instruments in science education. Res Sci Educ 48:1273–1296.

30. Taber KS. 2018. The use of Cronbach’s alpha when developing and reporting research instruments in science education. Res Sci Educ 48:1273–1296.

31. Stevens LM, Hoskins SG. 2014. The CREATE strategy for introducing scientific thinking, a new course for STEM-intended freshmen, demystifying scientific thinking through analysis of scientific literature. CBE Life Sci Educ 13:219–231.

32. Stevens LM, Hoskins SG. 2014. The CREATE strategy for introducing scientific thinking, a new course for STEM-intended freshmen, demystifying scientific thinking through analysis of scientific literature. CBE Life Sci Educ 13:219–231.
72. Yan VX, Bjork EL, Bjork RA. 2016. On the difficulty of mending metacognitive illusions: a priori theories, fluency effects, and misattributions of the interleaving benefit. J Exp Psychol Gen 145:918–933.
73. Karpicke JD, Butler AC, Roediger III HL. 2009. Metacognitive strategies in student learning: do students practise retrieval when they study on their own? Memory 17:471–479.
74. Ebert-May D, Derting TL, Hodder J, Momsen JL, Long TM, Jardeleza SE. 2011. What we say is not what we do: effective evaluation of faculty professional development programs. BioScience 61:550–558.
75. Henderson C, Dancy MH. 2009. Impact of physics education research on the teaching of introductory quantitative physics in the United States. Phys Rev Spec Top Phys Educ Res 5:020107 doi:10.1103/PhysRevSTPER.5.020107.
76. Andrews TM, Leonard MJ, Colgrove CA, Kalinowski ST. 2011. Active learning not associated with student learning in a random sample of college biology courses. CBE Life Sci Educ 10:394–405.
77. Bussey TJ, Orgill M, Crippen KJ. 2013. Variation theory: a theory of learning and a useful theoretical framework for chemical education research. Chem Educ Res Pract 14:9–22.
78. Lloyd GM, Cai J, Tarr JE. 2017. Issues in curriculum studies: evidence-based insights and future directions, p 824–852. In Cai J (ed), Compendium for research in mathematics education. National Council of Teachers of Mathematics, Reston, VA.
79. Amundsen C, Wilson M. 2012. Are we asking the right questions? A conceptual review of the educational development literature in higher education. Rev Educ Res 82:90–126.
80. Henderson C, Beach A, Finkelstein N. 2011. Facilitating change in undergraduate STEM instructional practices: an analytic review of the literature. J Res Sci Teach 48:952–984.