

Article

The CREATE Strategy for Intensive Analysis of Primary Literature Can Be Used Effectively by Newly Trained Faculty to Produce Multiple Gains in Diverse Students

Leslie M. Stevens* and Sally G. Hoskins[†]

*Department of Molecular Biosciences, University of Texas, Austin, Austin, TX 78712; [†]Biology Department, City College and The Graduate Center, City University of New York, New York, NY 10031

Submitted December 12, 2013; Revised March 2, 2014; Accepted March 29, 2014
Monitoring Editor: Nancy Pelaez

The CREATE (Consider Read, Elucidate the hypotheses, Analyze and interpret the data, and Think of the next Experiment) strategy aims to demystify scientific research and scientists while building critical thinking, reading/analytical skills, and improved science attitudes through intensive analysis of primary literature. CREATE was developed and piloted at the City College of New York (CCNY), a 4-yr, minority-serving institution, with both upper-level biology majors and first-year students interested in science, technology, engineering, and mathematics. To test the extent to which CREATE strategies are broadly applicable to students at private, public, research-intensive, and/or primarily undergraduate colleges/universities, we trained a cohort of faculty from the New York/New Jersey/Pennsylvania area in CREATE pedagogies, then followed a subset, the CREATE implementers (CIs), as they taught all or part of an existing course on their home campuses using CREATE approaches. Evaluation of the workshops, the CIs, and their students was carried out both by the principal investigators and by an outside evaluator working independently. Our data indicate that: intensive workshops change aspects of faculty attitudes about teaching/learning; workshop-trained faculty can effectively design and teach CREATE courses; and students taught by such faculty on multiple campuses make significant cognitive and affective gains that parallel the changes documented previously at CCNY.

INTRODUCTION

The explosion of information in the biological sciences since the mid-20th century necessitates new approaches in the classroom (National Research Council, 2003, 2009; American Association for the Advancement of Science [AAAS], 2011).

DOI: 10.1187/cbe.13-12-0239

Address correspondence to: Sally G. Hoskins (shoskins@ccny.cuny.edu).

© 2014 L. M. Stevens and S. G. Hoskins. *CBE—Life Sciences Education* © 2014 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

“ASCB[®]” and “The American Society for Cell Biology[®]” are registered trademarks of The American Society for Cell Biology.

Recent studies suggest that undergraduates taught in traditional ways fail to achieve mastery of key concepts essential to working scientists (Shi *et al.*, 2011; Trujillo *et al.*, 2012). Textbooks may underemphasize the scientific process in figures (Duncan *et al.*, 2011), thereby contributing little to students’ understanding of how data should be represented or interpreted (Rybarczyk, 2011) and making it difficult for students to recognize that the material in their textbooks was largely derived from experimental or observational studies. Traditional teaching methods may not promote development of critical-reading, analytical, or thinking skills routinely employed by working scientists, especially when such teaching focuses on more basic skills (Momsen *et al.*, 2010). Traditional multiple-choice testing may even impair students’ cognitive development (Stanger-Hall, 2012).

Active-learning approaches that increase student engagement enhance learning in some situations (Knight and Wood, 2005; Freeman *et al.*, 2007) but not in others (Andrews *et al.*,

2011), with instructor training likely to be an important component of successful outcomes. Thus, although some new styles of teaching may be of significant value to students, it may be challenging for inexperienced faculty members to carry them out effectively. Even the highly motivated faculty members who attend professional development workshops can have difficulty applying the lessons learned (Silverthorn *et al.*, 2006; Ebert-May *et al.*, 2011; Henderson *et al.*, 2011).

These issues raise questions about the best ways to achieve pedagogical reform. In particular, are workshops an effective way to train faculty in the use of nontraditional teaching approaches? In addition, if a technique is effective when used by its developers, often science education specialists, will it also be effective when practiced by workshop-trained faculty members who do not have science education expertise? Finally, if a new teaching strategy produces cognitive and attitudinal gains in a particular cohort of students, will it also be effective when used by different faculty members with substantially different student cohorts? We have addressed these questions in relation to the CREATE (Consider Read, Elucidate the hypotheses, Analyze and interpret the data, and Think of the next Experiment) strategy, a teaching/learning approach that uses intensive analysis of primary literature to demystify and humanize science (Hoskins *et al.*, 2007).

CREATE was developed to address the shortcomings of traditional science teaching described above (Hoskins *et al.*, 2007) and utilizes primary literature, rather than textbooks, as its focus. CREATE guides students in a deep analysis of scientific papers that challenges them to interpret the data as if it were their own; students must understand both the design and the broader context of a given study. These students are likely to be motivated to think more deeply than when reading traditional textbooks, which typically summarize results or present simplified outlines of experiments in the absence of data (Duncan *et al.*, 2011). In addition, the recent *Discipline-based Education Research* (DBER) report states that DBER “clearly shows that research-based instructional strategies are more effective than traditional lecture in improving conceptual knowledge and attitudes about learning” (Singer *et al.*, 2012, p. 3).

CREATE was originally designed for upper-level students (Hoskins *et al.*, 2007) and has since been adapted for first-year students as well (Gottesman and Hoskins, 2013). In the original version of CREATE, students in an upper-level elective that was focused on “analysis of primary literature” read a series of papers that had been published in sequence from a single laboratory. The students learned to use a variety of new and adapted pedagogical tools that included concept mapping, cartooning (sketching “what went on in lab” to generate the data in each figure), annotation of figures (rewriting caption and narrative information directly onto figure panels), and “translation” (paraphrasing complex sentences from the narrative). At the conclusion of data analysis for each paper, and before the next paper in the series was revealed, students designed their own follow-up experiments and vetted these in grant panel exercises designed to mimic activities of bona fide panels. Late in the semester, students generated a set of questions for the authors of the papers, addressing both personal and professional motivations and experiences. These were compiled into a single survey that was emailed to

each author. The diversity of candid responses, from principal investigators (PIs), postdoctoral fellows, graduate students, and other collaborators, illuminated multiple aspects of “the research life.” These unique insights into the people behind the papers complemented students’ deep understanding of the papers and of the study design, affecting students’ views of “who” can become a scientist (see student postcourse interview data in Tables 1 and S1 of Hoskins *et al.*, 2007).

In previous studies at City College of New York (CCNY), we found that upper-level students in CREATE courses made gains in critical thinking, content integration, and self-assessed learning (Hoskins *et al.*, 2007), and also experienced positive shifts in science attitudes, self-rated abilities, and epistemological beliefs about science (Hoskins *et al.*, 2011). In a pilot first-year Introduction to Scientific Thinking CREATE cornerstone elective, CCNY freshmen also made significant gains in critical thinking and experimental design ability, exhibited positive shifts in science attitudes and self-rated abilities, and showed maturation of some epistemological beliefs about science (Gottesman and Hoskins, 2013). Thus, for both upper-level and first-year students at CCNY, CREATE courses promote cognitive as well as affective gains.

To investigate the applicability of CREATE approaches at other institutions and in other topic areas, we trained faculty members from multiple campuses in the CREATE approach in a series of intensive workshops. We then followed a subset of workshop participants as they applied CREATE methods at their home campuses (Figure 1). With the aid of an outside evaluator (OE), we tested three hypotheses: 1) that our workshops would be an effective way to convey CREATE pedagogies and change participants’ views about science education and their classroom approaches; 2) that workshop-trained faculty from diverse colleges and universities would teach CREATE effectively in their first attempt; and 3) that students on diverse campuses, in courses taught by CREATE workshop-trained faculty, would make cognitive and/or affective gains paralleling those seen previously at a minority-serving institution (MSI).

Our data indicate that the workshop series changed aspects of participants’ thinking about science teaching, that the subset of seven workshop-trained faculty participants who subsequently taught courses as part of our study enthusiastically and effectively implemented CREATE, and that students at every implementing campus demonstrated significant shifts in cognitive and affective categories similar to those examined earlier at CCNY. Gains varied on the different campuses, but it was notable that even short-term, partial-semester CREATE implementations produced some important changes. CREATE faculty development workshops are thus an effective way to convey CREATE teaching strategies, as faculty members trained in such workshops can successfully teach CREATE courses the first time they attempt to do so. Importantly, CREATE students from a wide range of campuses (public, private, liberal arts, research intensive) made gains in critical thinking and concurrently underwent positive shifts in self-assessed attitudes, abilities, and epistemological beliefs related to science. We conclude that our workshop model is an effective way to train faculty to teach CREATE and that the CREATE method produces cognitive and affective gains in diverse student cohorts.

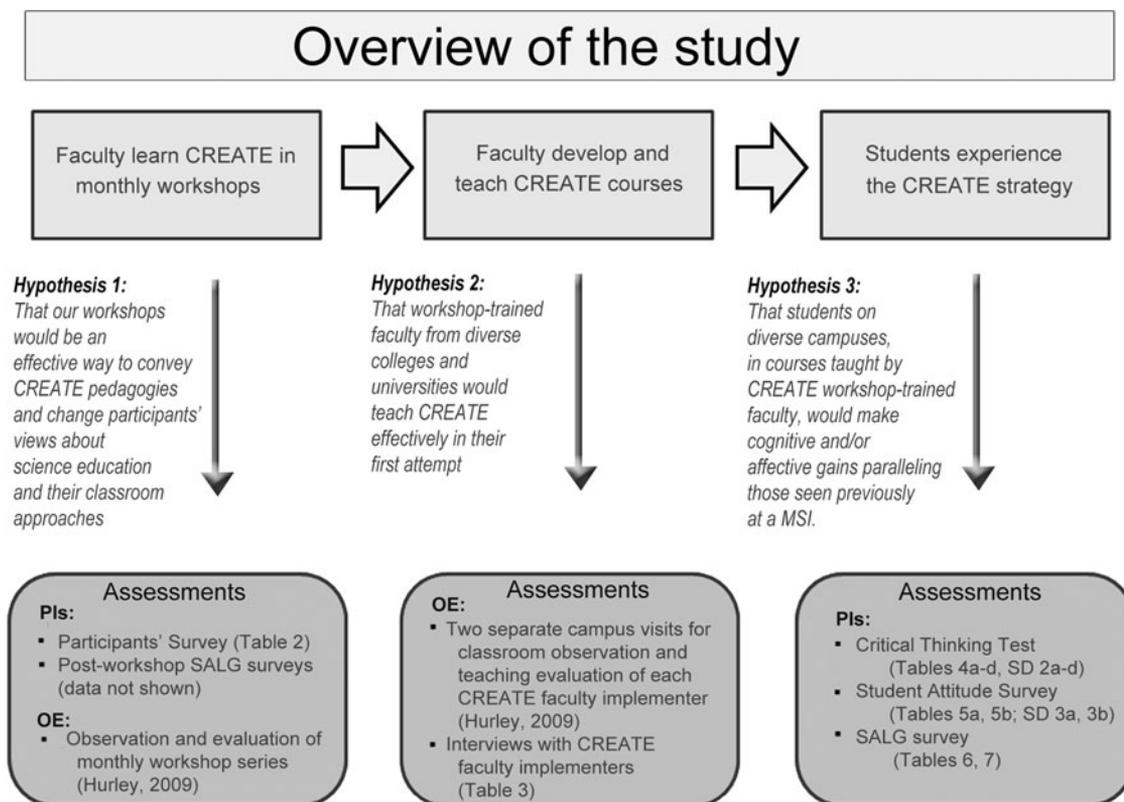


Figure 1. Outline of the three-part study. Faculty were recruited from the New York/New Jersey/Pennsylvania area and trained in workshops held in New York City monthly from August to December 2007. A subset of the faculty participants was subsequently followed as they implemented CREATE on their home campuses by using the strategy in a course they were already scheduled to teach. Each phase of the project was assessed by the PIs, an OE, or both.

METHODS

Structure of Workshops

We recruited 16 faculty workshop participants through presentations at both national and regional meetings of professional societies (including the American Society for Cell Biology, National Association of Biology Teachers, Society for Neuroscience, Society for Developmental Biology, Northeastern Nerve Net, Mid-Atlantic Regional Society for Developmental Biology, and SENCER [www.sencer.net]). Faculty participants met at the City University of New York (CUNY) Graduate Center in midtown Manhattan once per month from August to December 2007 for workshops aimed at presenting both theoretical and practical aspects of the CREATE strategy. Each workshop ran from 10 am to 4 pm and was focused on particular tools of the CREATE pedagogical tool kit (Hoskins and Stevens, 2009). As part of the training, workshop participants played the parts of students as a workshop leader (S.G.H.) taught using CREATE strategies. Some participants “student taught” by using CREATE methods, with their fellow participants acting as students. Faculty participants learned how to select and prepare appropriate primary scientific literature and related materials, to teach using CREATE approaches, and to assess students in CREATE courses. Some of the workshops also featured visits from an experienced CREATE teacher who learned the strategy from our publications and discussions with S.G.H.

and from a panel of seven students who were alumni of the 2006 CCNY upper-level CREATE class. The latter panelists answered participants’ questions about the experience of being a student in a CREATE class. Between workshop sessions, participants completed assignments that are typically given to CREATE students, using CREATE tools such as concept mapping, cartooning, figure annotation, and designing future experiments. Other assignments focused on selecting appropriate primary literature modules and designing classroom activities suitable for students working in groups.

Recruitment of CREATE Implementers

At the final workshop in December 2007, seven participants committed to teaching CREATE courses in the 2008 Spring or Fall semester (Table 1). Faculty participants who implemented CREATE did so by using CREATE strategies in courses they were already scheduled to teach on their home campuses. The PIs worked with the implementers to plan their integration of CREATE into a single existing course and to obtain institutional review board (IRB) approval for the education research project. Seven CREATE implementers (CIs) carried out implementations on five separate campuses. Two campuses had two CIs each, in different courses. Of the seven courses used for the implementations, four were in biology, two were in psychology, and one was in biochemistry. Each CI chose papers appropriate to his or her own student cohort

Table 1. Characteristics of CREATE professors, their campuses, and their CREATE courses^a

	Carnegie classification of campus where CREATE was implemented	Years of teaching experience	Weeks teaching CREATE	Class meeting times	Professor status	Had the implementer taught this course before?	Students' majors	Students' status and gender	Course type in catalogue
1	Very high undergraduate	3.5	Full semester	85-min class twice per week	Assistant	No	Biology	17 seniors (13 F, 4 M)	Topical seminar
2	High undergraduate, single doctoral program	24	10 wk	2 h, 20 min, once per week	Associate	Yes	Psychology	13 juniors–seniors (8 F, 5 M)	Seminar
3	Very high undergraduate, single doctoral program	6	Full semester	2 h, 45 min once per week	Assistant	Yes	Biology	14 seniors (9 F, 5 M)	seminar
4	Very high undergraduate, single doctoral program	6	4 wk	75 min twice per week	Assistant	Yes	Psychology	26 juniors (14 F, 12 M)	Lecture
5	High undergraduate, also graduate/professional programs	8.5	10 wk	3-h lab course twice per week	Assistant	Yes	Biochemistry	11 seniors (7 F, 4 M)	Laboratory
6	High undergraduate, also graduate/professional programs	7	4 wk	3h once per week	Assistant	Yes	Psychology	11 juniors (7 F, 3 M, 1 other)	Lecture
7	Majority graduate/professional	2	Full semester	2 h once per week	Assistant	Yes	Premed	13 seniors (8 F, 5 M)	Lecture

^aModified from OE final report (Hurley, 2009); order of courses adjusted and Carnegie classifications added. Note that implementations 3 and 4 and implementations 5 and 6 were on the same campus; thus, overall, we studied seven implementations on five campuses.

and course (Table 1); thus the implementations were individual case studies rather than exact duplications of the CREATE courses and modules of papers taught at CCNY.

Design of Implementations

Students self-selected into the courses on each campus; no effort was made to recruit students for a research study. Per IRB requirements, CIs invited their student cohorts to participate in a research study involving “CREATE, a new way to teach biology (or chemistry, or psychology)”; thus all students were aware that research on a novel teaching/learning strategy was taking place in their classrooms. Students were also informed that two course sessions would be observed and that on her second visit the OE would invite those who wished to participate to take an anonymous survey regarding their reaction to their course. All cohorts learned the tools of CREATE, but not all faculty participants took time in class to describe the CREATE project specifically. CIs administered anonymous pre- and postcourse surveys to those who chose to take part and forwarded these to the PIs for scoring. Each CI also coordinated with the OE to arrange for her to observe their CREATE class twice during the semester at an appropriate interval (observations 4 wk apart for full-semester implementations;

2–4 wk apart for shorter implementations). The PIs communicated informally (phone or email) with CIs during the semester and held a midsemester meeting with CIs at the CUNY Graduate Center to share experiences and troubleshoot any issues that arose. At the conclusion of their work, implementers received a stipend in recognition of their efforts. One hundred and eight students participated in the study in the seven implementations. Totals for particular surveys may vary slightly due to individual student absences on days that different surveys were administered. The majors of the participating students were psychology (40%), biology (32%), premed (12%), biochemistry (11%), and undeclared (5%). Of the participants 63% were female, and 30% described themselves as “ethnic minority” (Hurley, 2009).

Evaluation of Workshops (Figure 1, Hypothesis 1)

PIs' Evaluation. Workshop participants' reactions to the workshops were assessed by the responses of faculty participants to 1) Likert-style participant surveys prepared by the PIs and administered before (August) and after (December) the series of monthly workshops (Table 2), and 2) surveys regarding individual workshops. Workshop participants completed the participant surveys using code numbers that allowed the PIs to pair the pre- and postworkshop surveys

Table 2. Shifts in faculty workshop participants' attitudes and beliefs preworkshops vs. postworkshops^a: Survey of faculty workshop participants

Statements presented in modified format pre- and postworkshop series	% agree, pre/post
I often have/will have my students work in small groups in my <i>lecture</i> classes.	56/77
I often have/will have my students work in small groups in my <i>seminar</i> classes.	71/100
In preparing for class, I consider/will consider whether my students have misconceptions about the subject that could interfere with their ability to learn the material.	77/90
I refer/will refer to the literature on science education in developing my teaching strategies.	52/75
At the beginning of each course, I assess/will assess how much the students already know/understand about the topic.	61/79
I give/will give assignments that challenge students to examine data closely and represent it in multiple ways.	69/81
I teach/will teach concept mapping in class to help students relate new content to what they learned previously.	43/88
I have/will have students in my lecture classes read articles from the primary literature.	70/81
I have/will have students in my seminar classes read articles from the primary literature.	91/99
I make/will make a special effort to humanize the science that I teach by describing my own research experiences or inviting guest speakers involved in the work being studied	84/87

Statements presented in identical format pre- and postworkshop series	Average (SD)		Significance	Effect size
	Pre	Post		
Students need to have completed introductory course work in science before they can read and understand primary scientific literature.	3.00 (0.91)	2.59 (1.13)	$p < 0.05$	0.4
The time required to learn new styles of teaching is prohibitive.	3.76 (0.73)	3.18 (1.20)	$p < 0.04$	0.6
I find it difficult to understand the literature on science education.	3.82 (0.77)	3.29 (0.89)	$p < 0.04$	0.6
Only the most talented students can learn to think critically about science.	1.62 (0.72)	1.37 (0.50)	ns	—
Science classes are different from courses in other subjects in that students must learn the content before they can think about it analytically.	3.47 (0.95)	3.00 (0.98)	ns	—
Learning how to think critically does not require advanced knowledge of the subject matter and therefore can be taught in college as early as the freshman year.	4.00 (1.01)	4.20 (0.73)	ns	—
It is necessary to teach most science courses using lectures, in order to cover enough content.	3.65 (0.72)	3.29 (1.03)	ns	—
Lab classes provide students with hands-on experience that gives them a good idea of how scientific research is carried out.	3.50 (1.15)	3.00 (1.15)	ns	—

^aAll workshop participants ($n = 16$) took a Likert-style survey before the first workshop (August 2007) and after the final workshop (December 2007). Surveys were anonymous but coded with numbers known only to the participants to allow statistical analysis of paired pre/postsurveys. Ten statements aimed at assessing aspects of participants' current or future classroom practice were presented with slight modification (e.g., "I will . . .") in the postworkshop survey; eight others aimed at assessing participants' beliefs about pedagogical matters were identical in the two surveys. Scores for the latter cohort were analyzed by paired t test, and ESs were calculated for statements showing significant change.

while preserving anonymity. The participant surveys asked workshop participants to respond to 18 statements about pedagogical practice and teaching/learning issues. Phrasing of statements varied so that the "optimal" postcourse response was not always the one indicating stronger agreement. For example, following the workshops we would expect *less* agreement with the statement "Science classes are different from courses in other subjects in that students must learn the content before they can think about it analytically." The seven statements that were presented identically on the pre- and postsurveys were scored on a 5-point scale (I strongly disagree; I disagree; I'm not sure; I agree; I strongly agree), and pre/post workshop differences were calculated using a paired t test (Excel). For 11 statements, participants were asked (preworkshop) to respond regarding the extent to which they already engaged in particular practices characteristic of the CREATE classroom (e.g., having students work in small groups, spending class time closely analyzing data, teaching concept mapping, making efforts to humanize science), whereas the postworkshop survey asked participants to indicate the likelihood that they would implement such practices in the future. For these 11 statements, we calculated

the overall percentage of respondents who agreed or strongly agreed.

For evaluation of each individual workshop, we adapted the Student Assessment of Their Learning Gains (SALG) anonymous evaluation tool provided free to the college-level teaching community (www.salgsite.org). At the conclusion of each individual workshop session, the participants completed a SALG survey that posed questions directed at that month's workshop activities.

OE's Evaluation. Independent evaluation of the workshop series was provided by OE Marlene Hurley, Ph.D. (in science education). Dr. Hurley attended each workshop, tracked numerous aspects of interactions of workshop participants and workshop faculty using observation protocols designed for this purpose, observed individual workshop participants as they "student taught" the rest of the group, and observed question/answer sessions between the workshop participants and visiting CREATE students or faculty. During each workshop session, the OE tracked the extent to which individual aspects of the CREATE process were introduced, discussed, modeled, and/or practiced. After each workshop,

the OE provided immediate formative feedback to the PIs in a half-hour discussion of the workshop day. At the conclusion of the workshop series, the OE provided written summative feedback. This evaluation and the OE's observations and evaluations of the individual CREATE implementations (outlined below) were compiled by the OE in a 127-page final report and provided to the PIs (Hurley, 2009)

OE's Evaluation of Teaching by CIs (Figure 1, Hypothesis 2)

For evaluation of CREATE faculty implementers on individual campuses, the OE modified two existing observation protocols to adapt them specifically for the CREATE classroom. One, the Weiss Observation Protocol for Science Programs, was originally devised for evaluation of National Science Foundation (NSF)-funded projects (Hurley, 2009; see also Weiss *et al.*, 1998). The OE adapted this instrument so that it could be used during each observation session to gather evidence on four focus areas and key indicators related to each area (Supplemental Material). The focus areas were: design of session (12 key indicators), instruction of session (18 key indicators), science content (seven key indicators), and the nature of science (three key indicators). Each focus area was tracked during the observation with reference to these measures. For example, design of session included "The design of the session appropriately balanced attention to multiple goals within the CREATE structure" and "Adequate time and structure were provided for reflection." During the teaching session, performance in each area was tracked on a numeric scale, where 1 = "not at all"; 2, 3, and 4 were unlabeled; and 5 = "to a great extent." The scale also included 6 (don't know) and 7 (not applicable). Each focus area was summarized in a synthesis rating based on a scale of 1–5, where 1 = "Design of the session was not at all reflective of reflective of best practices for CREATE"; 2, 3, 4, were unlabeled; and 5 = "Design of the session was extremely reflective of best practices for CREATE." This synthesis rating in turn was supported by additional written comments. For each of the four focus areas, the observer also rated faculty members' teaching on a scale of 1 to 5 (1 = not successful, 2 = slightly successful, 3 = moderately successful, 4 = highly successful, 5 = extremely successful).

The OE also used a Flanders Observation Protocol (Flanders, 1963) that was adapted to broaden its applicability in college teaching situations with diverse formats. This was used to collect quantitative data on interactions between CREATE implementers and their students; these data were used by the OE to determine the extent to which implementers ran student-centered CREATE classrooms.

Owing to scheduling constraints, not every observation was made by the primary OE, Dr. Hurley. Some observations were made by her colleague, Dr. Fernando Padró. To ensure consistency, the two evaluators pilot-tested the observation protocol instruments during initial observations in Spring 2008 to establish interrater reliability (Hurley, 2009). To supplement the data gleaned from the instruments described above, the OEs took descriptive notes during observations. Late in the semester, the primary OE interviewed each faculty member to elicit overall reactions to teaching with the CREATE strategy for the first time as well as his or her sense of potential student gains.

Evaluation of Potential Cognitive and Affective Gains by Students in CREATE Implementations (Figure 1, Hypothesis 3)

The PIs coordinated with the CIs to administer: 1) a critical-thinking test (CTT) adapted from the Field-tested Learning Assessment Guide (www.flaguide.org), which was administered pre- and post-CREATE implementation; 2) the Student Attitude Survey (SAS), a Likert-style survey examining students' self-assessed abilities, attitudes, and beliefs, also administered pre- and post implementation, with scores analyzed by Wilcoxon signed-rank test (<http://vassarstats.net/wilcoxon.html>); and 3) a postcourse online SALG survey. All surveys were anonymous; on the CTT and SAS surveys, students used secret code numbers that allowed their pre- and postcourse responses to be paired for statistical analysis. No coding was used with the SALG, which was administered postcourse only.

CTT. As one goal of our study was to determine whether the critical-thinking gains seen at CCNY would be replicated on campuses with different student cohorts, the CIs administered a CTT that included four questions that also appeared on the CTT used in our original study (see Figure 2, Q 1–4, of Hoskins *et al.*, 2007). Questions A, C, and D were from the "General Science/Conceptual Diagnostic Test/Fault Finding and Fixing/Interpreting and Misinterpreting Data" section of the Field-tested Learning Assessment Guide website and were presented in an identical form on the pre- and postimplementation CTT. Question B was designed by the PIs to focus on biological data analysis and, as in our previous study, was presented in isomorphic form (identical form with different contexts and data pre- and postimplementation).

Each question challenges students to interpret charts or graphs, find trends in data, and determine whether the conclusions stated in the narrative accompanying each question follow logically from the data shown. The question asks students whether they do or do not agree with the conclusion stated and prompts them to explain their thinking with reference to the relevant chart, graph, or table, for example, "Do you agree or disagree? Explain why, using data to support your answer." Students wrote short responses to each question. We used a rubric developed in our previous study (Hoskins *et al.*, 2007) to score the following measures: Whether a student agreed, disagreed, or took no stance on the conclusion stated in the question; the number of logical justifications presented as part of the student's explanation; and the number of illogical justifications presented. Total numbers of logical and illogical justifications in each response were compiled separately. Responses that appropriately disagreed with an illogically stated conclusion received an additional "logical" point, while those that inappropriately agreed with an illogical conclusion received an additional point in the "illogical" category. Students who took no stance on the "agree/disagree" prompt had no points added to their totals. Initially, 20 CTT responses were scored independently by a research assistant and one of the authors (S.G.H.). Scores were compared and discrepancies reconciled. An additional 15 tests were then scored by each scorer, with 87% of scores identical. The research assistant then scored all of the surveys. Not all students were present for both the pre- and the postimplementation test, and we scored responses only for

which we had both assessments. The CTT scores were analyzed by paired *t* test (Excel).

SAS. In the previous CCNY studies, we documented a positive effect of the CREATE approach on students' confidence in their ability to think like scientists and to read/analyze primary literature, as well as on their attitudes about their own abilities, the nature of science, and the nature of knowledge, or epistemological beliefs (Hoskins *et al.*, 2011). To determine whether similar changes occurred in the student cohorts in the implementations, CIs administered a SAS to their students pre- and postimplementation. The 26-statement SAS is derived from the 37-statement Survey of Student Attitudes, Abilities, and Beliefs (SAAB), the development of which has been described (Hoskins *et al.*, 2011; Gottesman and Hoskins, 2013). Factor analysis of SAAB statements was used to derive broad categories addressed by that survey. Statements in the SAS are identical to those in the SAAB, but the SAS contains fewer statements in some categories. Students responded to SAS statements on a 5-part Likert scale to which we assigned the following point values: 1 = "strongly disagree," 2 = "disagree," 3 = "I'm not sure," 4 = "agree," 5 = "strongly agree." We calculated post- versus preimplementation differences using the Wilcoxon signed-rank test, a nonparametric statistical test considered appropriate for Likert-style data (Lovelace and Brickman, 2013).

SALG. The SALG surveys were constructed to assess reaction to multiple aspects of the CREATE implementations, including students' sense of whether particular tools enhanced their learning, whether the course structure enhanced their learning, and whether their views of scientists and the research process changed during the CREATE semester. This survey was given postcourse only. Spring 2008 implementations (CI1–CI5) used the original version of the site, which was redesigned and altered slightly in Summer 2008 by the survey developers. As the original SALG version was no longer available, the Fall 2008 implementations, CI6 and CI7, used the new version, potentially affecting survey validity. We designed surveys to be as parallel as possible and aligned questions for scoring.

Most questions were phrased "How much did [a particular tool or course aspect] help your learning?" and answered with a number ranging from 1 to 5 (no help; little help; some help; much help; great help). We analyzed the data by pooling responses to the set of statements within individual categories (e.g., grouping the multiple individual statements that addressed categories including "the class overall"; "class activities"; or "your understanding of course content"). Means and SDs were calculated in Excel. The SALG also allows construction of open-ended questions to which students can respond in writing. Here, we asked questions about students' confidence in their scientific thinking ability and their postcourse views of scientists.

RESULTS

I. Test of hypothesis 1: Our workshops would be an effective way to convey CREATE pedagogies and change participants' views about science education and their classroom approaches.

PIs' Evaluation—Participant Survey

Table 2 summarizes data from the 18-statement participant survey addressing workshop participants' teaching approaches and attitudes about science pedagogy pre- and postworkshop. With regard to statements examining participants' preworkshop practices compared with what they planned to do postworkshop, we saw changes in 10 out of 11 survey statements. The largest (more than 20% different postworkshop) changes were in response to statements regarding the use of small groups in lecture classes, on assessing students' precourse knowledge, on referring to the published literature on science education, and on using concept mapping. The response to the latter increased to almost 90% from less than 50% preworkshop. Statements showing smaller changes tended to have a strong response even preworkshop, indicating that many of the highly motivated workshop participants were already engaging in some activities that encourage student engagement, although it was clear from discussions in the first workshop that all faculty participants used lecture as their most typical classroom approach. We used the "pre" survey to gain insight into the extent to which participants were also using some of the approaches typical of the CREATE classroom.

With regard to the statements presented in identical form pre- and postworkshop, we saw significant change with moderate effect size (ES) on two statements: "The time required to learn new styles of teaching is prohibitive" and "I find it difficult to understand the literature on science education." In each case, faculty participants agreed significantly *less* with the statement postworkshop. An additional statement, "Students need to have completed introductory course work in science before they can read and understand primary scientific literature" showed significant change (less agreement postworkshop) with small ES.

PIs' Evaluation—SALG Survey

In the SALG surveys given at the conclusion of each workshop session, participants ranked multiple aspects of each workshop on a scale of 1–5, with 5 = "great help"; 4 = "much help"; 3 = "moderate help"; 2 = "some help"; and 1 = "little help." Overall, workshop rankings averaged 4, "much help," for each of the five workshops (data not shown). In written comments, participants expressed particular enthusiasm for the November workshop (#4), which included a discussion with the panel of CCNY CREATE students.

OE's Evaluation—Observation of Workshops

The OE's assessment of the effectiveness of workshop sessions focused on the question "Was CREATE successfully introduced, modeled, and disseminated in the faculty development workshops in order to adequately prepare college faculty for methods of science teaching that are beyond the traditional lecture and lab?" (Hurley, 2009, p. 3). These data complemented those collected by the PIs in the surveys described above. As part of her evaluation, the OE tracked the extent to which individual aspects of CREATE were introduced, taught, and reinforced in each workshop, the degree to which the PIs modeled student-centered teaching, how much "student" (faculty participant) participation ensued, the nature of classroom discussions, and the degree to

which faculty participants seemed to be learning the CREATE strategy. Based on naturalistic notes taken during the workshops, including specific tracking of the extent to which key aspects of CREATE (constructivism, active engagement, inquiry, concept mapping, cartooning and annotating, “design the next experiment,” grant panels, and email contact with science researchers) were included in each session (data not shown), the OE concluded: “In response to Evaluation Question 1 [above], and in light of all the evidence presented and considered, CREATE was successfully introduced, modeled, and disseminated in the faculty development workshops in order to adequately prepare college faculty for methods of science teaching that are beyond the traditional lecture and lab” (Hurley, 2009, p. 81).

II. Test of hypothesis 2: Workshop-trained faculty from diverse colleges and universities would teach CREATE effectively in their first attempt.

OE's Evaluation of Implementer Teaching

The conclusions of the primary OE regarding CI effectiveness relied on data from multiple sources, including 1) student activities during the two observations of each implementing classroom, 2) faculty classroom performance as rated on the observation protocols, 3) interviews with the CIs, 4) naturalistic notes taken during the observation periods, and 5) her overview of student survey data, which included a statement asking students to rate their teacher's performance on a 1–5, low→great scale (data not shown). Two CIs were rated “extremely successful,” one “highly to extremely successful,” two “highly successful,” one “moderately to highly successful,” and one “moderately successful” by the OE. There was no obvious relationship between the success of each implementation and the CI's years of teaching experience, whether the campus was public or private, or whether the overall implementation was full semester or partial semester.

The response of the CIs to using the CREATE approach was uniformly positive. In their 2008 interviews with the OE (Table 3), all CIs stated that they would use the teaching strategy again and would recommend it to colleagues. Further, multiple CIs remarked that they would recommend CREATE for courses beyond the ones they had taught, including graduate courses, courses for non–science majors, and courses in psychology, history, and social science. Implementers noted their enjoyment of the group activities, their sense that students were building metacognitive awareness, their students' growing ability to discuss and argue productively with peers, the students' increased confidence in their reading/analysis ability, and overall high student engagement in the classroom.

III. Test of hypothesis 3: Students on diverse campuses, in courses taught by CREATE workshop-trained faculty, would make cognitive and/or affective gains paralleling those seen previously at an MSI.

Assessments of Implementers' Students

CTT. For many of the individual implementations, when pre- and postimplementation scores were compared, significant increases and decreases were seen in the number of logical and illogical justifications, respectively (Table 4, A–D). To diminish the variability arising from differences between the

effectiveness of individual CIs, we pooled the data for all the implementations and found that, for all four questions, the students exhibited statistically significant increases in their logical statements. These are consistent with an increase in the ability of the students to think critically and analytically about data, that is, in their ability to “think like a scientist.”

We also examined the CTT data for differences between full-semester and partial-semester implementations and between public and private campuses by pooling the relevant categories. In partial-semester implementations, faculty members typically presented material in a traditional format in the course's early weeks and then shifted to CREATE-style teaching later in the semester. For the pooled full-semester implementations ($n = 45$ students), all of the questions showed significant shifts in the number of logical justifications, with large to moderate effect size (ES). For the pooled partial-semester cohort ($n = 63$), significant change was seen for three questions, with a large, moderate, or small ES. Thus, while significant change is seen in all groups, more and larger gains are made by the full-semester CREATE implementations. A similar pattern is seen for illogical statements, for which the full-semester implementations changed significantly (fewer illogical justifications postcourse than pre-course) on three questions, while the partial-semester implementations changed on only one. Taken together, these findings suggest that full-semester implementations produce more changes with larger overall ESs than do partial-semester implementations.

Comparison of private ($n = 42$) and public ($n = 66$) student cohorts shows significant gains in logical statements made by both groups on all four questions (Table 4, A–D). ESs were large or moderate for private campuses and large, moderate, or small for public universities. Illogical statements decreased significantly on all questions in the private school cohort, and on a single question in the public school cohort. Overall, private school cohorts changed significantly in more categories and with larger ESs than did public school cohorts, but as the numbers overall are small and may be compounded by effects related to implementation duration, this conclusion must be considered tentative.

SAS. With respect to SAS factors (1 = decoding primary literature, 2 = interpreting data, 3 = active reading, 4 = visualization, and 5 = thinking like a scientist), all cohorts changed significantly on factors 1, 3, 4, and 5, and all but the partial-semester cohorts also made significant gains on factor 2 (Table 5A). Comparing full- and partial-semester implementations shows larger ESs in full-semester cohorts, arguing for more significant gains (Coe, 2002; Maher *et al.*, 2013) in the longer-duration implementations. Pooled private or pooled public implementations showed gains on each of factors 1–5 with more comparable ESs. No significant change was seen in any subgroup in a sixth category, “research in context,” which deals with students' sense of the use of model systems and controls in experiments, suggesting that students were already familiar with these concepts (data not shown).

When we examined individual implementations with respect to these factors (Supplemental Material, SD 3a), we found that students in the majority of implementations changed significantly on statements assessing self-rated ability to decode scientific literature, science reading ability, and visualization, with large ESs. Multiple campuses also showed

Table 3. Implementers' comments on teaching with CREATE and whether they would use the strategy in future courses or recommend the strategy to colleagues

Implementation number	OE interview question to implementing faculty ^a : "What are the benefits of using CREATE?"	Implementers' perceptions	
		Would you use CREATE again?	Would you recommend CREATE to other faculty?
1	OE notes: Having students collectively engage in materials in class, bringing everyone together to discuss the articles in sections and spend time on in-depth examination of the articles. C mapping is a valuable tool students don't necessarily like. Implementer's favorite part: doesn't have to lecture. Also, CREATE is a confidence builder for peer learning. Implementer: "It pushes them to do something they haven't done before and there's not a single right answer."	Yes	Yes (including non-science faculty)
2	OE notes: Increased participation of students is a benefit. Implementer: "This is an approach that frees you from standard routines and allows exploration and innovation." "[Post-CREATE] I'm more open to experimenting. . . ." "It made teaching more fun—not scripted—that happened from the first day . . . My favorite aspects are that students get to investigate on their own. Taking time to go over an article is a luxury. Planning well makes it really pay off."	Yes—	Yes, would recommend for science, anthropology, and sociology courses
3	OE notes: Implementer said a big benefit was that in small-group work, even students who would not talk in a large group scenario get involved. Implementer: "Cartooning provided a fantastic way to visualize what's going on—the student has to understand experimental design to cartoon, whereas writing means that the student knows what's in the article, not if he or she understands it." "[I would use it again because] it [CREATE] is an effective way to teach about science, to propose experiments, and to read the science literature. These are things we hardly ever teach students to do"	Yes	Yes—including social sciences and psychology History colleague has incorporated some of the ideas
4	Implementer: "CREATE is a much better method to walk through the literature in a systematic manner. It is easy to understand and helps [students] to also understand articles in other fields." "[The benefits] include knowing when students understand the issue/concepts/ideas and when they do not. Another benefit is that students help each other out and it makes the class more interactive than it had been before." "[I would use it again because] CREATE provides a way to work with articles that is different from anything else I have seen."	Yes	Yes
5	OE notes: A major benefit is that students realize "they can do it" (read primary literature in the field). Student understanding in biochemistry and nature of science were below average to start; improved "dramatically." Implementer: "Getting students to realize that they could actually understand the primary literature in the field" is the major benefit of CREATE. "Previously, students tended to become aggravated because they had no sense of how to evaluate themselves on their progress in understanding the material. Now they have a better sense of how to get where they need to be."	Yes	Yes and would also recommend it to non-science colleagues
6	OE notes: CREATE helped to establish a more learner-centric environment rather than the traditional teacher-centric one. Implementer: "Students seem to be more engaged because they now have a mechanism through which they can talk about what they read, generating good class dynamics." "[I would use it again] because it is an informative way to get students to analyze work from the ground up and to think critically about their subject matter."	Yes	Yes, with some modifications/yes also for non-science colleagues
7	Implementer: "The discovery focus is the most exciting part. They own their ideas and argue ideas with their peers. Revealing information in small bursts drives them through curiosity. We have the most fun with grant panels . . . students really get into it. They suggest good experiments." "I am a total convert to group activities and how they [the students] feed off each other."	Yes	Yes, including graduate educators

^aOE interviews with implementers were conducted on the OE's second campus visit in 2008.

Table 4. CTT scores for pooled implementer campuses^a

Student population	<i>n</i>	Logical justifications			Illogical justifications				
		Pre mean (SD)	Post mean (SD)	Significance ES	Pre mean (SD)	Post mean (SD)	Significance ES		
Question A									
CCNY ^b	48	1.9 (1.4)	2.0 (1.9)	ns	—	0.8 (0.9)	0.9 (0.8)	ns	—
Full semester	45	2.4 (1.4)	3.3 (1.4)	0.000	0.7	0.3 (0.6)	0.2 (0.5)	0.040	0.3
Partial semester	63	1.8 (1.2)	2.5 (1.4)	0.002	0.4	0.4 (0.7)	0.3 (0.6)	ns	—
Private	42	2.6 (1.3)	3.2 (1.3)	0.011	0.5	0.3 (0.6)	0.1 (0.4)	0.020	0.4
Public	66	1.8 (1.3)	2.6 (1.6)	0.000	0.6	0.5 (0.6)	0.4 (0.6)	ns	—
Pool of all	108	2.1 (1.3)	2.8 (1.5)	0.000	0.5	0.4 (0.6)	0.3 (0.6)	0.020	0.2
Question B									
CCNY ^b	48	0.8 (1.2)	3.4 (1.9)	0.001	1.6	1.8 (1.2)	0.4 (1.2)	0.001	1.2
Full semester	45	1.3 (1.1)	3.0 (1.6)	0.000	1.2	1.0 (1.0)	0.3 (0.6)	0.003	1.0
Partial semester	63	0.8 (1.0)	2.4 (1.6)	0.000	1.2	1.3 (0.8)	0.6 (0.7)	0.000	0.9
Private	42	1.3 (1.1)	3.0 (1.3)	0.000	1.4	1.1 (1.0)	0.2 (0.6)	0.000	0.9
Public	66	0.8 (1.0)	2.4 (1.7)	0.000	1.1	1.3 (0.8)	0.6 (0.7)	0.000	0.9
Pool of all	108	1.0 (1.1)	2.6 (1.6)	0.000	1.2	1.2 (0.9)	0.5 (0.7)	0.000	1.2
Question C									
CCNY ^b	48	1.5 (0.7)	2.9 (1.2)	0.021	1.4	0.7 (0.4)	0.3 (0.5)	0.023	0.9
Full semester	45	1.7 (1.0)	2.5 (1.4)	0.000	0.7	0.8 (0.9)	0.3 (0.6)	0.001	0.7
Part semester	63	1.3 (1.3)	1.6 (1.2)	ns	—	0.6 (0.7)	0.5 (0.7)	ns	—
Private	42	1.7 (1.0)	2.5 (1.3)	0.000	0.7	0.7 (0.8)	0.1 (0.4)	0.001	0.8
Public	66	1.4 (1.3)	1.7 (1.4)	0.030	0.2	0.5 (0.7)	0.5 (0.7)	ns	—
Pool of all	108	1.5 (1.2)	2.0 (0.5)	0.000	0.5	0.6 (0.7)	0.4 (0.6)	0.012	0.3
Question D									
CCNY ^b	48	0.8 (1.3)	1.1 (1.9)	ns	—	1.7 (1.0)	1.1 (0.8)	0.001	0.7
Full semester	45	1.3 (0.8)	1.7 (0.9)	0.012	0.5	1.3 (1.0)	1.7 (0.9)	ns	—
Partial semester	63	0.9 (0.8)	1.3 (0.9)	0.000	0.5	1.1 (0.9)	1.0 (0.8)	ns	—
Private	42	1.2 (0.6)	1.7 (0.9)	0.004	0.6	1.3 (1.0)	0.9 (0.9)	0.029	0.4
Public	66	1.0 (0.9)	1.4 (0.9)	0.002	0.4	1.1 (0.9)	1.1 (0.7)	ns	—
Pool of all	108	1.1 (0.8)	1.5 (0.9)	0.000	0.5	1.2 (0.9)	1.0 (0.8)	ns	—

^aCTT questions A, C, and D were from the Field-tested Learning Assessment Guide; question B was written by the PIs for the CTT. Each question was used previously at CCNY (Hoskins *et al.*, 2007). CIs 1, 3, and 7 were full-semester implementations; the others ranged from 4 to 10 wk. CIs 1, 2, and 7 are private institutions; the others are public. Pool of all = CI 1–7. Significance determined using paired *t* test (Excel); *p* values < 0.05 were considered significant.

^bOriginal CCNY cohort; pool of three CCNY implementations; all were full-semester and taught by S.G.H. in 2004–2005.

gains on self-rated data interpretation ability and ability to think like a scientist. The changes seen are in the same range as those seen in a comparison single class of CCNY students taught in a full-semester CREATE course by an experienced CREATE teacher (S.G.H.; see Supplemental Material, “CCNY 2009” in SD 3a).

The factors addressing epistemological issues showed fewer overall changes and more differential outcomes between cohorts (Table 5B). Four of the seven factors defined in the original SAAB survey as addressing epistemological issues, (referred to here as 1 = creativity; 2 = sense of scientists; 3 = sense of scientists’ motivations; 4 = research as collaborative), showed significant change in the implementations. The pooled group of all institutions changed significantly with small ES on factors 1 and 2. Full-semester implementations changed on factors 1, 2, and 3 (large to moderate ES), while partial-semester implementations did not change significantly on any epistemological factor.

Outcomes in public versus private implementations were comparable, with each cohort changing with moderate to small ES on two of the four categories (Table 5B). Looked at individually, four of the seven cohorts changed significantly

on at least one epistemological factor (Supplemental Material, SD 3b). Figure 2 plots precourse versus postcourse responses for pooled cohorts on factor 1, decoding primary literature. This set of scatter plots shows a pattern seen throughout the SAS responses, with full-semester and private school cohorts less dispersed overall than partial-semester and public school cohorts.

The changes seen at individual campuses on the SAS are in the same range as those seen in a comparison single class of CCNY students taught by an experienced CREATE teacher (S.G.H.) in a full-semester CREATE course (see Supplemental Material, “CCNY 2009” in SD 3). CCNY students made gains in three epistemological categories, with moderate to large ES. These findings are consistent with our hypothesis 3, that student groups at various types of institutions will respond to the CREATE method by undergoing positive shifts in self-rated science abilities and attitudes as well as epistemological beliefs.

SALG. The SALG survey contained 45 prompts distributed across 10 categories. Table 6 summarizes data for each category, with scores reflecting combined averages for the

Table 5A. SAS: Statements addressing students' self-assessed attitudes and abilities^a

	Pre average (SD)	Post average (SD)	Significance	ES
Decoding primary literature				
Full semester	13.4 (2.7)	15.7 (2.0)	0.0001	1.0
Partial semester	13.3 (3.1)	15.0 (3.2)	0.0001	0.5
Private	14.2 (2.5)	16.2 (2.1)	0.0001	0.8
Public	12.8 (3.0)	14.7 (3.0)	0.0001	0.6
All institutions	13.3 (2.9)	15.3 (2.8)	0.0001	0.7
Interpreting data				
Full semester	10.5 (1.1)	11.8 (1.6)	0.0001	0.7
Partial semester	10.6 (2.2)	11.2 (2.3)	ns	–
Private	11.0 (1.7)	11.9 (1.6)	0.0411	0.5
Public	10.3 (2.2)	11.9 (1.6)	0.0016	0.4
All institutions	10.5 (2.1)	11.4 (2.1)	0.0001	0.4
Active reading				
Full semester	14.0 (1.7)	15.8 (1.9)	0.0001	1.0
Partial semester	14.1 (2.4)	15.8 (2.1)	0.0001	0.7
Private	14.7 (1.8)	16.4 (1.8)	0.0001	0.9
Public	13.7 (2.2)	15.4 (2.0)	0.0001	1.9
All institutions	14.1 (2.1)	15.8 (2.0)	0.0001	0.9
Visualization				
Full semester	9.9 (2.2)	11.8 (1.4)	0.0001	1.1
Partial semester	10.1 (2.2)	12.0 (2.0)	0.0001	0.6
Private	10.1 (2.1)	12.0 (1.1)	0.0001	0.9
Public	9.9 (2.2)	11.3 (2.1)	0.0001	0.6
All institutions	10.0 (2.2)	11.5 (1.8)	0.0001	0.7
Thinking like a scientist				
Full semester	10.7 (1.7)	12.5 (1.5)	0.0000	1.1
Partial semester	10.9 (2.1)	11.5 (2.1)	0.0285	0.3
Private	11.4 (1.7)	12.5 (1.3)	0.0012	0.6
Public	10.5 (1.9)	11.5 (2.1)	0.0001	0.4
All	10.9 (1.9)	11.9 (1.9)	0.0001	0.5

^aThe SAS comprises a subset of statements from the SAAB survey described in Hoskins *et al.* (2011). Survey responses were scored and summed in categories. We took “strongly disagree” = 1, “strongly agree” = 5. Some statements (e.g., “scientists usually know what the outcomes of their experiments will be” were reverse-scored for analysis. Significance calculated by Wilcoxon signed-rank test. *p* values are shown; those < 0.05 were considered significant. On statements addressing a sixth factor, “research in context,” no significant changes were seen in cohorts other than “All.” See Supplemental Material, SD 3a, for results of individual implementations.

substatements in that category. Notably, students in all implementations reported self-assessed learning gains in all categories. Responses to two open-ended SALG survey questions, on students' views of scientists and of their ability to “think like a scientist,” are also discussed below (Table 7).

Of the SALG summary categories (Table 6), “the course overall,” “class activities,” and “support for you as an individual learner” were rated by two implementations each as the category in which students felt they made the largest learning gains. In one implementation, the score was highest for “your understanding of class content.” Within individual categories, we explored the substatements that received high scores from more than one implementation. For the majority of implementations, “the way the material was approached” (the course overall); “discussions in class” (class activities) and “opportunities for in class review” (assignments, graded activities, and tests) received the highest ratings.

Regarding the category “aspects of the class that might have affected learning,” students in multiple implementations singled out “the relationship between classwork and assigned reading,” “the modified scientific papers,” or “templates for analyzing figures” as the most helpful. With regard

to the categories “support for you as an individual learner” and “your understanding of class content,” the majority of implementations rated “the quality of contact with the instructor” and “the relationships between concepts” as the most influential.

With regard to “increases in your skills” (15 substatements), we tracked the top three self-reported gains for each implementation. “Critically reviewing articles,” “understanding the ‘methods’ sections in papers,” “relating methods used to data obtained,” “reading about science,” “ability to explain concepts,” “understanding experimental design,” and “critical analysis” were among the top choices in multiple implementations. Regarding “impact on your attitudes,” most implementations scored “appreciating this field” as the area in which they had made the largest overall gains, with others noting their sense that they had gained transferable skills. Regarding “integration of your learning,” the majority chose “understanding the relevance of this field to real-world issues,” while responding students of the remaining implementations considered their largest gains to be in “understanding how ideas in this class relate to those in other science classes.”

On SALG open-ended questions, students were asked about their views of scientists and their confidence that they

Table 5B. SAS: Statements addressing students' epistemological beliefs^a

	Pre average (SD)	Post average (SD)	Significance	ES
Creativity				
Full semester	4.2 (0.7)	4.7 (0.5)	0.0006	1.0
Partial semester	4.2 (0.7)	4.4 (0.7)	ns	—
Private	4.2 (0.8)	4.7 (0.5)	0.0012	0.7
Public	4.2 (0.6)	4.4 (0.7)	ns	—
All institutions	4.2 (0.7)	4.5 (0.6)	0.0005	0.4
Sense of scientists				
Full semester	3.1 (1.0)	3.7 (0.8)	0.0220	0.7
Partial semester	3.0 (1.0)	3.2 (0.9)	ns	—
Private	3.6 (0.9)	3.7 (0.7)	ns	—
Public	2.9 (0.9)	3.3 (0.9)	0.0117	0.4
All institutions	3.2 (1.0)	3.4 (0.9)	0.0131	0.3
Sense of motives				
Full semester	3.7 (0.8)	4.2 (0.7)	0.0032	0.6
Partial semester	3.7 (0.8)	3.7 (0.8)	ns	—
Private	3.9 (.80)	4.2 (0.7)	0.0434	0.4
Public	3.6 (.79)	3.7 (0.8)	ns	—
All institutions	3.7 (.80)	3.9 (0.8)	ns	—
Collaboration				
Full semester	4.5 (0.6)	4.6 (0.9)	ns	—
Partial semester	4.4 (0.7)	4.5 (0.6)	ns	—
Private	4.7 (0.4)	4.6 (0.9)	ns	—
Public	4.2 (0.7)	4.5 (0.6)	0.0285	0.5
All institutions	4.4 (0.7)	4.5 (0.7)	ns	—

^aIn additional categories addressing the certainty of knowledge and the innateness of ability, no changes were seen in pooled groups. Significance = *p* values calculated by Wilcoxon signed-rank test ; values < 0.05 were considered significant. Supplemental Material, SD 3b, shows outcomes in these categories for individual implementations.

could think like scientists. Fifty-two percent reported that their views of scientists had changed postcourse. Open-ended comments associated with this SALG question indicated that views had changed favorably. In addition, 74% of student respondents described themselves as feeling "more confident" of their scientific thinking ability. Table 7 presents representative open-ended comments from each implementation that are related to these two questions. Undergraduates on multiple campuses noted a new awareness of scientific creativity as well as confidence that analytical skills they developed during the semester would be applicable in future academic situations.

DISCUSSION

We will discuss the results of our study with respect to the three hypotheses proposed at the outset of our study (Figure 1). We note, however, that our findings related to hypothesis 2 reflect on hypothesis 1, in that effective teaching by the CIs implies the workshops were successful. Similarly, gains made by the CIs' students (hypothesis 3) reflect on the effectiveness of the CIs' teaching (hypothesis 2).

Hypothesis 1: Our workshops would be an effective way to convey CREATE pedagogies and change participants' views about science education and their classroom approaches.

Workshops have long been used to disseminate pedagogical improvements (Emerson and Mosteller, 2000; American

Association of Physics Teachers, 2013; Gregg *et al.*, 2013), but relatively few studies have evaluated workshop efficacy with respect to faculty participants, beyond participant self-report. The increasingly common "develop and disseminate" model, in which science educators develop novel pedagogical approaches and then train other faculty members to implement them, a strategy also used in this study, has to date not been supported strongly by data (Henderson and Dancy, 2007, 2011; Henderson *et al.*, 2010, 2011). Even if a workshop successfully changes faculty attitudes and intentions, the process of changing how teachers teach and students learn is surprisingly difficult. For example, workshop-trained faculty members may self-report that they are implementing new approaches, yet not actually use the new methods as their developers intended (Ebert-May *et al.*, 2011). In some cases, it may take several years postworkshop before faculty self-report applying the workshop training effectively (Pfund *et al.*, 2009). Finally, it is unusual for projects involving faculty development workshops to combine faculty participant self-assessment with both independent evaluation of "workshop alumni performance," such as repeated outside evaluation of their postworkshop teaching, and with separate anonymous assessment of participants' students using cognitive as well as affective measures. As a result, there are few data on cognitive and affective changes in the student cohorts taught by workshop-trained faculty. While our workshops and subsequent implementations were relatively small scale, our multiple levels of assessment allow us to address both the effects of the workshop training on faculty

Table 6. SALG—the online survey addresses 10 broad topics with varying numbers of statements related to each topic^a

Summary topic (number of substatements)	1	2	3	4	5	6	7
Q1. The course overall (3)	3.75 (0.8)	3.18 (1.67)	4.17 (0.63)	3.59 (0.86)	4.36 (0.61)	4.00 (0.85)	4.83 (0.41)
Q2. Class activities (3)	4.29 (0.20)	3.13 (0.44)	4.50 (0.17)	3.59 (0.86)	4.47 (0.19)	4.33 (0.38)	4.28 (0.35)
Q3. Assignments, graded activities, and tests (2)	3.76 (0.05)	2.79 (0.06)	4.21 (0.30)	3.36 (0.10)	4.13 (0.05)	4.12 (0.18)	4.47 (0.18)
Q4/5. Aspects of the class that might have affected learning (5)	3.64 (0.19)	3.02 (0.21)	4.06 (0.20)	3.33 (0.25)	4.48 (0.14)	4.5 (0.52)	3.73 (0.92)
Q6. Support for you as an individual learner (3)	4.24 (0.21)	3.67 (0.19)	4.11 (0.61)	3.3 (0.63)	4.73 (0.08)	4.25 (0.66)	4.40 (0.36)
Q7. Your understanding of class content (4)	3.83 (0.12)	3.14 (0.37)	4.25 (0.15)	3.57 (0.28)	4.46 (0.05)	4.75 (0.20)	4.58 (0.17)
Q8. Increases in your skills (15)	3.81 (0.37)	3.21 (0.49)	3.88 (0.44)	3.19 (0.45)	4.41 (0.29)	4.37 (0.38)	4.41 (0.41)
Q9. Impact on your attitudes (8)	3.99 (0.24)	3.31 (0.41)	3.98 (0.20)	3.17 (0.27)	4.45 (0.12)	4.56 (0.29)	4.75 (0.86)
Q10. Integration of your learning (2)	4.20 (0.38)	3.47 (0.23)	4.25 (0.11)	3.39 (0.05)	4.41 (0.12)	4.63 (0.18)	4.51 (0.24)
<i>n</i> respondents	15	13	12	14	12	4	6
Duration	Full	Partial	Full	Partial	Partial	Partial	Full
Type	Private	Private	Public	Public	Public	Public	Private

^aImplementers’ students were invited to take the postcourse online SALG survey at the conclusion of their respective courses. The SALG invites students to respond on a 5-point scale. Depending on the syntax of the SALG question, some responses were in the form of “no help/little help/moderate help/much help/great help.” On questions of “how much did you gain . . .?”, responses included “no gain/a little/some/much/a great deal.” Seventy-six students completed the SALG. Mean (SD) scores for pooled statements in each category are presented for the seven CREATE implementations (CIs). Implementers who taught CREATE in Spring 2008 used the original SALG; those who taught in Fall 2008 (implementations 6 and 7) used a new version, redesigned and altered slightly by the SALG site developers in Summer 2008. We aligned categories of the original and new SALG surveys and combined original SALG category 4 and new SALG category 5, as they addressed the same issue.

Table 7. Representative open-ended responses of students from different cohorts to prompts on the SALG survey regarding their sense of scientists and their confidence in their ability to think like a scientist^a

Cohort	Personal connection to science/scientists—	Confidence in ability to “think like a scientist”—
	“Do you think of scientists same as before or differently? Please explain”	“Are you more/less/equally confident in your ability to “think like a scientist” now, compared to precourse? Please explain”
1	“I have a better understanding of the creative aspect of their work”	“I feel much more confident in my ability to ‘think like a scientist’ since I can now sit down with a scientific paper, pick it apart, and understand it. I understand the methods, why those methods were chosen, and the results without being overwhelmed.”
2	“I feel that scientists are generally creative people who want to understand how things in the world work and why they work that way”	“More confident, because I know how to design an experiment”
3	“I definitely think differently. I think that scientists are much more creative than I had originally thought coming into this class.”	“More confident I will use these skills in the future and am sure that I will do good.”
4	“I have more respect for scientists as dedicated, passionate people beyond being just intelligent.”	“I feel more confident. The paper we discussed should have been way over my head, but through breaking it down into small pieces, it was much easier to understand what the authors were trying to convey.”
5	“[Precourse] I always felt scientists were very boring people and I learned a lot about them throughout the semester that they are just people like us. I learned to respect them and their research”	“This class has given me an opportunity to realize all that I have been taught and organize it in such a way that I am able to access this knowledge without much trouble. This class has also given me a way to be able to think about the questions that need to be answered, and the possible questions that arise with new discoveries.”
6	“I used to be intimidated when I heard researchers but this class really put some things into [a new] light. I feel I could reach that level.”	“More confident. I used to get intimidated but I really understand [now] how to better read research papers.”
7	“I definitely [think] of scientists as more creative in designing elegant experiments, and not just applying the same techniques to new things.”	“More confident. I’ve been able to criticize their [scientists’] work in the classroom this semester and that is a confidence that I can carry with me to other situations as well.”

^aFifty-two percent of respondents self-described as thinking differently (and more positively) about scientists; 74% stated their confidence had increased. See text for details.

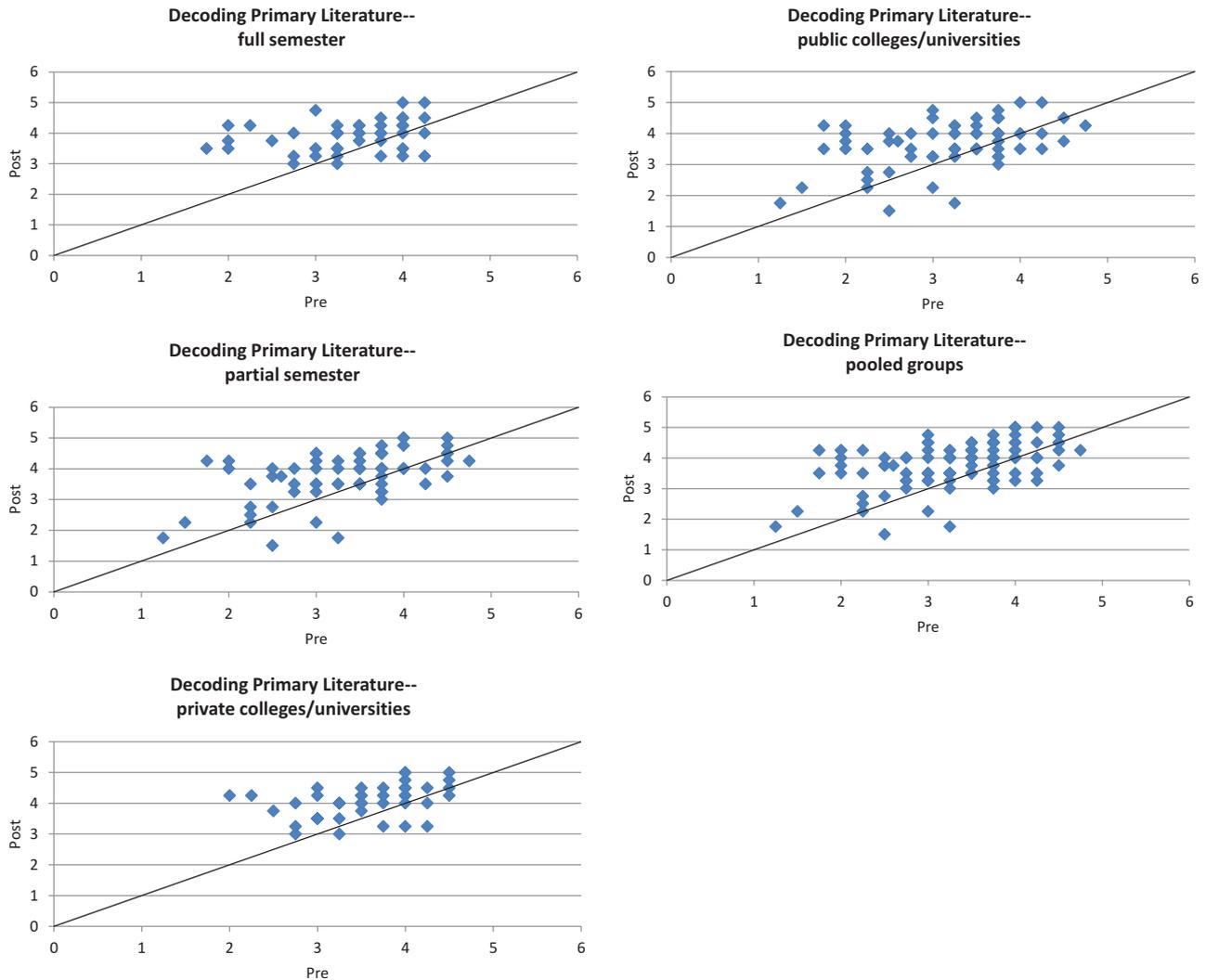


Figure 2. Scatter plots of data from the “decoding primary literature” factor. The factor encompasses four substatements, thus we divided pooled scores by 4 to yield average precourse and postcourse values. The numbers of dots are less than the numbers of students per cohort due to duplication of scores (multiple students had pre = 3 and post = 4, for example). The $y = x$ trend line represents hypothetical scores that were identical pre- and postcourse; thus points below the line represent students whose postcourse scores were lower than precourse, and conversely. While all cohorts made significant postcourse gains (Wilcoxon signed-rank test), we saw a broader distribution of scores and smaller effect sizes for students in the public and the partial-semester groups. Scores were less dispersed and effect sizes were larger in the full-semester and private school cohorts. The pattern represented here was typical of all factors (data not shown); see Tables 5A and 5B and Supplemental Material, SD 3, for quantitative analysis of the data set.

participants and the response of their students to the CREATE strategy.

Our participants were a self-selected group of faculty members who applied to the workshop because of their strong interest in science education. Although the participants were already carrying out a number of CREATE-type practices, our results indicate the workshop series was effective in shifting their attitudes, particularly with respect to the use of specific CREATE tools such as concept mapping, as well as in the use of precourse student assessment and making reference to the science education literature. We attribute this success in part to the dual function of the workshops in providing both a theoretical basis for the CREATE tool kit and associated classroom activities, as well as practical hands-on experience; for example, “student teaching” in workshops 4 and 5. Partici-

pants’ anonymous postworkshop SALG surveys (average: 4 = “much help” for each of the five workshops) support this view, as does the OE’s summary evaluation of the workshop series as “overall excellent” (Hurley, 2009).

Hypothesis 2: Workshop-trained faculty from diverse colleges and universities would teach CREATE effectively in their first attempt.

As described above, previous research suggests that effective workshop training does not guarantee successful implementation of new pedagogical strategies. Some projects have had strong starts with workshop training of faculty and additional plans to subsequently implement new teaching methods, but foundered when participants failed to follow through in the implementation process (Silverthorn *et al.*,

2006). Our faculty implementers agreed to file paperwork with their campus IRB board, to be observed twice by the OE, to return to the CUNY Graduate Center for midsemester discussion and troubleshooting, and to administer pre- and postimplementation surveys and tests to their students. We think it is likely that the workshop leaders' continued contact with the implementers contributed to the participants' persistence in carrying out these essential activities. In addition, as this project was supported by a grant from the NSF, we were able to provide stipends to the implementers in recognition of the time and effort they contributed to the project.

The OE noted that faculty participants differed substantially in teaching experience and their preworkshop understanding of pedagogy and that it is likely that all participants experienced traditional teaching during their own scientific training. Despite these variables, the OE determined that each of the seven implementations was "largely successful," and her assessment of individual CIs ranged from "moderately effective" to "extremely effective." Differences were attributed largely to individual differences in the implementers' levels of teaching experience and in their student cohorts, as well as in the particular approaches each implementer used when teaching CREATE. Some had larger classes or initially resistant students, which made introduction of a novel teaching strategy more challenging. Student resistance to change in teaching methods is not unusual (Dembo and Seli, 2004; Seidel and Tanner, 2013); change is perhaps particularly difficult for students who have achieved success in traditional classrooms. The duration of implementations as well as their timing within a given semester may have been an additional factor influencing CI effectiveness. We anticipate examining these issues in our ongoing expansion of CREATE implementations to 20 4-yr and community colleges in courses taught by a subset of the 96 faculty members we trained in intensive summer workshops in 2012–2013 (K. Kenyon and S. Hoskins, study in progress as part of DUE 1021443).

Although all student participants were aware that their class was participating in a science education study, the rationale behind the CREATE method was not explained by all of the CIs. This initially created anxiety for some students, as described in the following comments made in response to an anonymous survey administered by the OE (full data set not shown): 1) "At the beginning of the course the CREATE methodology was overwhelming, because I was accustomed to being expected to fully understand what I read very quickly. I was nervous about 'not knowing' until I realized that there would be time for me to explore the content and become familiar with it. This process of discovery has rarely been found in my other science courses . . . and is much more enjoyable and rewarding than being handed information and connections. It was empowering to realize that I can understand research in a field I was not familiar with, without the help of a textbook." 2) "While at first, I was hesitant and relatively unnerved by the CREATE method, I have found it to be very helpful. Not only in terms of understanding scientific articles (via mapping/cartooning), but also in terms of the social aspects of the field and an introduction to the processes (via small group work/mini grant panels)." We suggest that students will adapt more quickly if faculty explain the pedagogical bases of the novel CREATE tools throughout the implementation period.

Hypothesis 3: Students on diverse campuses, in courses taught by CREATE workshop-trained faculty, would make cognitive and/or affective gains paralleling those seen previously at an MSI.

We used three measures, a CTT and SAS that were both administered anonymously pre- and postcourse, and an anonymous postcourse SALG survey to gauge student gains in cognitive and affective areas.

CTT. As the CREATE approach had been successful in promoting gains in critical thinking in CCNY students (Hoskins *et al.*, 2007), we wanted to assess the extent to which similar gains might be made by students in the CREATE implementations. Our CTT had no questions directly related to module content in any of the implementations and thus was designed to measure potential transferable thinking skills developed in CREATE courses. The implementers' students, taken as a whole, made significant gains on the numbers of logical statements they made in support of their arguments on all CTT questions (Table 4, A–D). The results from individual implementations are not as strong statistically, because of the low numbers, but most made critical-thinking gains. No major differences were noted when comparing full/partial-semester or private/public implementations, although the full-semester implementations may have been slightly more effective (Table 4, A–D). This might be expected, due to the iterative nature of the CREATE process. In ongoing work with a larger population of student participants, (K. Kenyon and S. Hoskins, DUE 1021443), we hope to address the relationship between implementation length and student outcomes with more statistical power.

In addition to the results reported here and in our original description of the CREATE method (Hoskins *et al.*, 2007), we have documented critical-thinking gains, assessed by the Critical thinking Ability Test (CAT) survey (Stein *et al.*, 2012), in first-year science, technology, engineering, and mathematics (STEM)-interested students during a one-semester CREATE course at CCNY (Gottesman and Hoskins, 2013). Others have shown that an innovative inquiry-based undergraduate microbiology laboratory course leads to critical-thinking gains, also assessed by CAT, during a single semester (Gasper *et al.*, 2012). We have not compared the gains in critical thinking made by CREATE students with those of students in a more traditional class that is also focused on the primary literature, as was done in a recent study that used a modified version of CREATE (Segura-Totten and Dalman, 2013). These authors reported comparable critical-thinking gains for both groups. We do not consider these findings relevant to the present study, however, as these authors omitted numerous CREATE activities that we consider essential to the power of the approach and also evaluated critical-thinking abilities through an entirely different method (Hoskins and Kenyon, 2014). Although this comparison may be of interest in future studies, our focus here is on the ability of CREATE to bring about multiple gains, in addition to critical thinking, in diverse student populations.

SAS. The SAS included statements addressing students' sense of their academic abilities and statements reflecting epistemological belief factors (Hoskins *et al.*, 2011). As described in *Results*, we saw significant pre/postimplementation changes with large or moderate

ESs on multiple factors (Table 5A; changes on individual campuses are shown in Supplemental Material, SD 3). Student self-assessment data must be interpreted carefully, as students tend to overestimate learning gains, for example (Cloud-Hansen *et al.*, 2008). However, when we administered a similar pre/postcourse survey to a comparison group of CCNY students in a semester-long physiology course with lab that was *not* taught using CREATE, those students did not self-assess as having made any significant gains either in the categories above or in any of the epistemological categories discussed below (Gottesman and Hoskins, 2013). These findings argue that the attitudinal and epistemological shifts that we documented in the CIs' students did not occur simply as a consequence of student maturation during a typical college semester. Rather, we suggest that the epistemological and attitudinal changes seen in implementers' students reflect their response to the cognitive challenges of the CREATE approach.

Of the seven epistemological categories defined in our previous study, four (1, science as creative; 2, sense of scientists as people; 3, sense of scientists' motives; 4, science as a collaborative activity) showed significant change in some of the pooled groups, with the largest shifts in categories 1 and 2 (Table 5B). As in our previous study at CCNY (Hoskins *et al.*, 2011), we saw more changes overall in the science attitudes/abilities categories than for epistemological beliefs; this finding holds both for the implementers' students and a CCNY 2009 cohort (data not shown) taught by an experienced CREATE faculty member (SGH; see Supplemental Material, SD 3). These results are consistent with the well-known stability of epistemological beliefs (Perry, 1970; Baxter Magolda, 1992), but it is encouraging that changes in some aspects of students' epistemological beliefs did occur in multiple CREATE implementations.

The positive shifts in students' sense of science as creative are particularly interesting. The full-semester and private school pooled groups made significant gains on this issue, as did two of the three individual full-semester implementations. We are not aware of other studies in which an intervention focused on primary literature documented students' self-reported changes in their sense of science as a creative process. In a recent review, DeHaan noted that "Evidence suggests that instruction to support the development of creativity requires inquiry-based teaching that includes explicit strategies to promote cognitive flexibility. Students need to be repeatedly reminded and shown how to be creative, to integrate material across subject areas, to question their own assumptions, and to imagine other viewpoints and possibilities" (DeHaan, 2009, p 172). In a discussion of programs aimed at developing creativity specifically, Scott *et al.* (2004, p. 363) state that divergent thinking, "the capacity to generate multiple alternative solutions as opposed to the one correct solution," is an important component of creativity. We speculate that CREATE contributes to students' increased sense of scientific creativity by providing repeated opportunities for experimental design and encouraging students' realization that it is possible to interpret the same data in multiple ways and that more than one of the "next experiments" designed by their classmates may be valid.

Previous work with students in introductory college science courses indicates that, following instruction, introductory biology students may shift to a more naive rather than a more mature epistemology (Semsar *et al.*, 2011). A sim-

ilar trend of science students showing more naive epistemological beliefs postinstruction has been seen in introductory level physics (Adams *et al.*, 2006; Gray *et al.*, 2008) and chemistry (Adams *et al.*, 2008) courses. In the Semsar study, students in upper-division biology courses did not revert to more naive epistemology during the semester, but neither did they show substantial maturation of beliefs or large gains in expert-like thinking. Rather, upper-level students largely retained their precourse beliefs throughout the semester (Semsar *et al.*, 2011). The finding that students in multiple CREATE implementations developed more mature epistemological beliefs about science during their first CREATE course suggests that this teaching strategy, even when taught by inexperienced faculty, has the ability to influence diverse students' core beliefs about science.

SALG. The SALG survey designed by the PIs covers a wide range of aspects of the CREATE classroom experience (Table 6). Overall, students rated "the way the material was approached," "the opportunities for in-class review," and "the relationship between classwork and assigned reading" as substantially helping their learning. Self-assessed gains in confidence and learning were seen in all implementations. Thus, a single CREATE experience, which takes place in a classroom rather than a laboratory setting, can shift students' views of research scientists while at the same time enhancing students' self-assessed confidence in their ability to think scientifically. This is a particularly important finding, as students' attitudes can positively influence their motivation and, in turn, facilitate both persistence (President's Council of Advisors on Science and Technology, 2012) and success in science (House, 1994, 1995; Oliver and Simpson, 1998; Pintrich, 2004). We are encouraged that, overall, students in the implementations reported positive views of CREATE pedagogy and its influence on their abilities.

In Fall 2013, we surveyed the seven 2008 implementers regarding whether they were still using CREATE and its associated pedagogical tools. All reported continuing to teach with CREATE methods, using many or all of the CREATE tools. Some of the original implementing faculty participants described having designed additional creative enhancements of the core strategy, for example, building on the experimental design activities with a grant-writing project. On three of the five campuses where CREATE was tested, the 2008 implementers reported that additional faculty members (not trained in our workshops) have adopted CREATE strategies for their own courses (CREATE implementers #1, 3, 7, personal communications).

CONCLUSIONS

In a provocative recent essay, Charlene D'Avanzo argues that while the Vision and Change biology teaching reform process produced a strong set of recommendations (AAAS, 2011), it is unclear whether we "know how to change," citing a paucity of evidence from models based in education research and noting that some influential models have achieved substantial traction in the absence of such evidence (D'Avanzo, 2013; see also Henderson and Dancy, 2011, and Henderson *et al.*, 2011, on this point). Others have speculated that scientists' sense of their professional identities (Brownell and Tanner,

2012) and/or resistance at the departmental or institutional level (Singer *et al.*, 2012) are potential additional barriers to change. Thus, a large gap exists between defining goals for change and implementing them. The recent *Discipline-based Education Research* report “Crosscutting Themes” section calls for more studies to examine teaching innovations across multiple disciplines, in diverse cohorts of students, and in upper-division as well as introductory-level students. The report also recommends collecting data on “a wider range of outcomes” (Singer *et al.*, 2012).

When a pedagogical innovation produces no impact on students, it is difficult to know whether the innovation is itself ineffective or whether the faculty members implementing it are ineffective, perhaps because they were not well trained. We have addressed these issues in the current study by assessing at each level (workshop, faculty implementation, student outcomes) and by coordinating assessments by the PIs with those of an independent OE. Based on our expansion of CREATE to seven courses taught by workshop-trained faculty on campuses in the New York/New Jersey/Pennsylvania area, we report that the CREATE strategy, aimed at demystifying and humanizing science through intensive analysis of primary literature, is broadly applicable beyond the MSI where it was originally developed and piloted. Further, our data indicate that CREATE is effective in topic areas (psychology and biochemistry) beyond those in which it was originally tested. Finally, our data indicate that both full-semester and partial-semester CREATE implementations produce substantial gains, but that epistemological gains are very rare in the shorter implementations. Thus, the CREATE strategy is effective in the hands of faculty at a broad range of institutions, producing gains in diverse cohorts of students. Faculty members can learn to teach CREATE by participating in interactive workshops and can effectively adapt CREATE strategies to their existing courses. The persistence of CREATE courses taught by the original CIs 5 yr after the initial implementation and the spread of CREATE courses on several implementers’ campuses argue for the strategy’s adaptability and perceived value.

We posed three key hypotheses at the outset of our study: 1) Our workshops would be an effective way to convey CREATE pedagogies and change participants’ views about science education and their classroom approaches. 2) Workshop-trained faculty from diverse colleges and universities would teach CREATE effectively in their first attempt. 3) Students on diverse campuses, in courses taught by CREATE workshop-trained faculty, would make cognitive and/or affective gains paralleling those seen previously at an MSI. Each hypothesis was supported by data gathered in the study. We found that our workshops successfully prepared the CIs to teach effective CREATE courses and that students taught by CREATE faculty on a variety of campuses made critical-thinking and self-rated attitude/ability/epistemological belief gains while also self-reporting enhanced learning in multiple categories, deeper understanding of scientists, and more confidence in their scientific thinking ability. These changes occurred at all colleges/universities where CREATE was implemented, ranging from public universities to private liberal arts colleges and one private R1 university. Thus, CREATE is an effective strategy for diverse student cohorts. To our knowledge, CREATE is unusual in being easily learned, low cost, adaptable to multiple content areas, and effective in diverse

student cohorts, as well as able to convey transferable analytical skills, provoke positive changes in student self-rated attitudes/beliefs, and provide novel insights into the motivations of working researchers. As such, we believe that this teaching strategy could contribute to solving long-standing problems of student STEM learning and persistence by providing faculty members a framework for change that builds on their existing research skills (see Hoskins and Stevens, 2009, for discussion of this point). CREATE strategies play to faculty strengths and are built on analytical skills that faculty members already possess. They have the potential to transform undergraduate science education.

ACKNOWLEDGMENTS

We thank Prof. Laurel Eckhardt for support of the faculty development workshops at the CUNY Graduate Center; Profs. Shubha Govind and Anu Janakiraman for comments on a draft of the manuscript; and Dara Duncan, Alan Gottesman, and Jamila Hoque for assistance in data processing and discussions of the data. We also thank all of the CREATE implementers and their students for participating in the CREATE courses and associated surveys. This project was approved by CCNY IRB (protocol H-0633). We are very grateful to the NSF for support of the CREATE project. We also thank two anonymous reviewers for their comments on a previous version of the manuscript. This paper is based on work supported by the NSF Course, Curriculum and Laboratory Improvement/Transforming Undergraduate Education in Science program under DUE 0618536. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

REFERENCES

- Adams WK, Perkins KK, Podolefsky NS, Dubson M, Finkelstein ND, Wieman CE (2006). New instrument for measuring student beliefs about physics and learning physics: the Colorado Learning Attitudes about Science Survey. *Phys Rev ST Phys Ed Res* 2, 010101.
- Adams WK, Wieman CE, Perkins KK (2008). Modifying and validating the Colorado Learning Attitudes about Science Survey for use in chemistry. *J Chem Educ* 85, 1435.
- American Association for the Advancement of Science (2011). *Vision and Change in Undergraduate Biology Education: A Call to Action*, Washington, DC.
- American Association of Physics Teachers (2013). *The Role of Scientific Societies in STEM Faculty Workshops: Report of the May 2012 meeting*. www.aapt.org/Conferences/newfaculty/upload/STEM_REPORT-2.pdf (accessed 6 December 2013).
- 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.
- Baxter Magolda MB (1992). *Knowing and Reasoning in College: Gender-Related Patterns in Students’ Intellectual Development*, San Francisco, CA: Jossey-Bass.
- Brownell SE, Tanner KD (2012). Barriers to faculty pedagogical change: lack of training, time, incentives, and . . . tensions with professional identity? *CBE Life Sci Educ* 11, 339–346.
- Cloud-Hansen K, Kuehner J, Tong L, Miller S, Handelsman J (2008). Money, sex, and drugs: a case study to teach the genetics of antibiotic resistance. *CBE Life Sci Educ* 7, 302–309.
- Coe R (2002). It’s the Effect Size, stupid. What Effect Size is and why it is important. Presentation to British Educational Research Association, Annual Conference, held September 12–14, 2002, in Exeter, UK,

- www.psych.yorku.ca/mnelles/documents/ESguide.pdf (accessed 23 September 2012).
- D'Avanzo C (2013). Post *Vision and Change*: do we know how to change? *CBE Life Sci Educ* 12, 373–382.
- DeHaan RL (2009). Teaching creativity and inventive problem solving in science. *CBE Life Sci Educ* 8, 172–181.
- Dembo MH, Seli HP (2004). Students' resistance to change in learning strategies courses. *J Dev Educ* 27, 2–11.
- Duncan D, Lubman A, Hoskins SG (2011). Introductory biology textbooks under-represent scientific process. *J Microbiol Biol Educ* 12, 143–151.
- Ebert-May D, Derting T, Hodder J, Momsen J, Long T, Jardleza S (2011). What we say is not what we do: effective evaluation of faculty professional development programs. *BioScience* 61, 550–558.
- Emerson JD, Mosteller F (2000). Development programs for college faculty: preparing for the twenty-first century. In: *Educational Media and Technology Yearbook*, vol. 25, ed. RM Branch and MA Fitzgerald, Englewood, CO: Libraries Unlimited, 25, 26–42.
- Flanders N (1963). Intent, action, feedback: a preparation for teaching. *J Teach Educ* 3, 251–260.
- Freeman S, O'Connor E, Parks JW, Cunningham M, Hurley D, Haak D, Dirks C, Wenderoth MP (2007). Prescribed active learning increases performance in introductory biology. *CBE Life Sci Educ* 6, 132–139.
- Gasper BJ, Minchella DJ, Weaver GC, Csonka LN, Gardner SM (2012). Adapting to osmotic stress and the process of science. *Science* 335, 1590–1591.
- Gottesman AJ, Hoskins SG (2013). CREATE cornerstone: Introduction to Scientific Thinking, a new course for stem-interested freshmen, demystifies scientific thinking through analysis of scientific literature. *CBE Life Sci Educ* 12, 159–172.
- Gray KE, Adams WK, Wieman CE, Perkins KK (2008). Students know what physicists believe but they don't agree: a study using the CLASS survey. *Phys Rev ST Phys Ed Res* 4, 020106.
- Gregg C, Ales J, Pomaric S, Wischusen E, Siebenaller J (2013). Scientific teaching targeting faculty from diverse institutions. *CBE Life Sci Educ* 12, 382–393.
- 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.
- Henderson C, Dancy MH (2007). Barriers to the use of research-based instructional strategies: the influence of both individual and situational characteristics. *Phys Rev ST Phys Ed Res* 3, 020102.
- Henderson C, Dancy MH (2011). Increasing the impact and diffusion of STEM education innovations.. A white paper commissioned for the Characterizing the Impact and Diffusion of Engineering Education Innovations forum, held February 7–8, 2011, in New Orleans, LA. www.nae.edu/File.aspx?id=36304 (accessed 20 October 2013).
- Henderson C, Finkelstein N, Beach A (2010). Beyond dissemination in college science teaching: an introduction to four core change strategies. *J Coll Sci Teach* 39, 18–25.
- Hoskins SG, Kenyon KL (2014). Letter to the editor. *J Microbiol Biol Educ*. <http://jmbe.asm.org/index.php/jmbe/article/view/725/html>.
- Hoskins SG, Lopatto D, Stevens LM (2011). The CREATE approach to primary literature shifts undergraduates' self-assessed ability to read and analyze journal articles, attitudes about science, and epistemological beliefs. *CBE Life Sci Educ* 10, 368–378.
- Hoskins SG, Stevens LM (2009). Learning our L.I.M.I.T.S.: less is more in teaching science. *Adv Physiol Educ* 33, 17–20.
- Hoskins SG, Stevens LM, Nehm RH (2007). Selective use of the primary literature transforms the classroom into a virtual laboratory. *Genetics* 176, 1381–1389.
- House JD (1994). Student motivation and achievement in college chemistry. *Int J Instr Media* 21, 1–12.
- House JD (1995). Student motivation, previous instructional experience, and prior achievement as predictors of performance in college mathematics. *Int J Instr Media* 22, 157–168.
- Hurley MH (2009). CREATE Grant 2: Year 3. Unpublished final program evaluation report.
- Knight JK, Wood WB (2005). Teaching more by lecturing less. *Cell Biol Educ* 4, 298–310.
- Lovelace M, Brickman P (2013). Best practices for measuring students' attitudes toward learning science. *CBE Life Sci Educ* 12, 606–617.
- Maher JM, Markey JC, Ebert-May D (2013). The other half of the story: ES analysis in quantitative research. *CBE Life Sci Educ* 12, 345–351.
- Momsen JL, Long TM, Wyse SA, Ebert-May D (2010). Just the facts? Introductory undergraduate biology courses focus on low-level cognitive skills. *CBE Life Sci Educ* 9, 435–440.
- National Research Council (NRC) (2003). *BIO2010: Transforming Undergraduate Education for Future Research Biologists*, Washington, DC: National Academies Press.
- NRC (2009). *A New Biology for the 21st Century*, Washington DC: National Academies Press.
- Oliver JS, Simpson RD (1998). Influences of attitude toward science, achievement motivation, and science self-concept on achievement in science: a longitudinal study. *Sci Educ* 72, 143–155.
- Perry W (1970). *Forms of Intellectual Development in the College Years: A Scheme*, San Francisco, CA: Jossey-Bass.
- Pfund C *et al.* (2009). Professional development. Summer institute to improve university science teaching. *Science* 324, 470–471.
- Pintrich PR (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educ Psychol Rev* 16, 385–407.
- President's Council of Advisors on Science and Technology (2012). *Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics*, Washington, DC.
- Rybarczyk B (2011). Visual literacy in biology: a comparison of visual representations in textbooks and journal articles. *J Coll Sci Teach* 41, 106–114.
- Scott G, Levritz L, Mumford M (2004). The effectiveness of creativity training: a quantitative review. *Creat Res J* 16, 361–388.
- Segura-Totten M, Dalman NE (2013). The CREATE method does not result in greater gains in critical thinking than a more traditional method of analyzing the primary literature. *J Microbiol Biol Educ* 14, 166–175.
- Seidel SB, Tanner KD (2013). "What if students revolt?" Considering student resistance: origins, options, and opportunities for investigation. *CBE Life Sci Educ* 12, 586–595.
- Semsar K, Knight JK, Birol G, Smith MK (2011). The Colorado Learning Attitudes about Science Survey (CLASS) for use in biology. *CBE Life Sci Educ* 10, 268–278.
- Shi J, Power JM, Klymkowsky MW (2011). Revealing student thinking about experimental design and the roles of control experiments. *Int J Scholarsh Teach Learn* 5, 1–16.
- Silverthorn DU, Thorn PM, Svinicki MD (2006). It's difficult to change the way we teach: lessons from the Integrative Themes in Physiology curriculum module project. *Adv Physiol Educ* 30, 204–214.

Singer SR, Nielsen NR, Schweingruber H (eds.) (2012). *Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering*, Washington, DC: National Academies Press. www.nap.edu/catalog.php?record_id=13362 (accessed 19 July 2013).

Stanger-Hall KF (2012). Multiple-choice exams: an obstacle for higher-level thinking in introductory science classes. *CBE Life Sci Educ* 11, 294–306.

Stein B, Haynes A, Redding M (2012). The Critical thinking Ability

Test (CAT), Cookeville: Center for Assessment and Improvement of Learning, Tennessee Tech University.

Trujillo C, Cooper MM, Klymkowsky MW (2012). Using graph-based assessments within Socratic tutorials to reveal and refine students' analytical thinking about molecular networks. *Biochem Mol Biol Educ* 40, 100–107.

Weiss IR, Ridgeway CJ, Bond SL (1998). *Local Systemic Change through Teacher Enhancement: Year Three Cross-Site Report*, Chapel Hill, NC: Horizon Research.