

City University of New York (CUNY)

CUNY Academic Works

Publications and Research

City College of New York

2013

CREATE Cornerstone: Introduction to Scientific Thinking, a New Course for STEM-Interested Freshmen, Demystifies Scientific Thinking through Analysis of Scientific Literature

Alan J. Gottesman
CUNY City College

Sally G. Hoskins
CUNY City College

[How does access to this work benefit you? Let us know!](#)

More information about this work at: https://academicworks.cuny.edu/cc_pubs/421

Discover additional works at: <https://academicworks.cuny.edu>

This work is made publicly available by the City University of New York (CUNY).
Contact: AcademicWorks@cuny.edu

Article

CREATE Cornerstone: Introduction to Scientific Thinking, a New Course for STEM-Interested Freshmen, Demystifies Scientific Thinking through Analysis of Scientific Literature

Alan J. Gottesman and Sally G. Hoskins

City College and the Graduate Center, City University of New York, New York, NY 10031

Submitted November 21, 2012; Revised December 15, 2012; Accepted December 15, 2012

Monitoring Editor: Jose Vazquez

The Consider, Read, Elucidate hypotheses, Analyze and interpret data, Think of the next Experiment (CREATE) strategy for teaching and learning uses intensive analysis of primary literature to improve students' critical-thinking and content integration abilities, as well as their self-rated science attitudes, understanding, and confidence. CREATE also supports maturation of undergraduates' epistemological beliefs about science. This approach, originally tested with upper-level students, has been adapted in *Introduction to Scientific Thinking*, a new course for freshmen. Results from this course's initial semesters indicate that freshmen in a one-semester introductory course that uses a narrowly focused set of readings to promote development of analytical skills made significant gains in critical-thinking and experimental design abilities. Students also reported significant gains in their ability to think scientifically and understand primary literature. Their perceptions and understanding of science improved, and multiple aspects of their epistemological beliefs about science gained sophistication. The course has no laboratory component, is relatively inexpensive to run, and could be adapted to any area of scientific study.

INTRODUCTION

We think a significant number of students lose interest in studying science early in their college careers, because many science curricula do not promote open-ended discussion, critical analysis, and creative study design—activities that characterize science as it is practiced. We thought that one way

DOI: 10.1187/cbe.12-11-0201

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

This project addresses a new course developed by S.G.H. based upon work supported by the National Science Foundation (NSF) CCLI/TUES (Course, Curriculum and Laboratory Improvement/Transforming Undergraduate Education in Science) program under NSF grant 0942790. 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.

© 2013 A. J. Gottesman and S. G. Hoskins. *CBE—Life Sciences Education* © 2013 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.

to attract and retain students who might be considering science studies would be to give them an opportunity to develop their reading and analytical skills and gain a realistic sense of scientific thinking as soon as they started college. A Consider, Read, Elucidate hypotheses, Analyze and interpret data, Think of the next Experiment¹ (CREATE)-based course focused on scientific thinking, using a novel selection of readings whose analysis did not require years of content mastery, would, in principle, give freshmen a chance to engage deeply in activities characteristic of actual science practice. We hypothesized that such an experience could have a positive influence on students' scientific abilities, their attitudes toward science, and their understanding of the research process early in their academic careers. To test this idea, we developed a new elective, *Biology 10050: Introduction to Scientific Thinking*.

BACKGROUND

Biology 10050 was developed as an adaptation of an upper-level course, *Biology 35500: Analysis of Scientific Literature*

¹See Hoskins *et al.*, 2007.

with CREATE. That course, offered at City College of New York (CCNY) since 2004, aims to demystify and humanize science through intensive analysis of primary literature. In Biology 35500, “modules”—sets of journal articles published sequentially from single laboratories—are the focus for an intensive elective. Students learn a new set of pedagogical approaches, including concept mapping, cartooning of methodology, figure annotation, use of templates to parse experimental logic, and design of follow-up studies (Hoskins and Stevens, 2009; Hoskins, 2010b). These methods are applied first to an article from the popular press and then in the analysis of a series of primary literature papers that follow a particular scientific question (e.g., “How do axons find their targets in the embryo?” “How is axis polarity maintained during regeneration?”). By examining module articles in a stepwise manner, we develop a “lab meeting” atmosphere in the class, with experimental findings discussed as if they had been generated gradually by the students themselves. Within individual articles, every figure or table was analyzed with recognition that each specific question being addressed or question asked created a data subset that contributed to the major finding of the paper.

In CREATE class sessions, multiple aspects of study design are scrutinized closely as we work backward from data in each figure and table to reconstruct details of the particular experiment that generated those data before we analyze the findings. In the process of examining specific experiments and their outcomes, we repeatedly consider questions fundamental to much research, (e.g., “What is n ?” “How was the sample selected?” “What controls were done and what did each control for?” “How do the methods work?” “What is the basis of ‘specificity’ in staining, binding, or expression?” “How convincing are the data?”). In addressing such questions, students gain insight into the design and interpretation of research beyond the individual study under analysis. Because methods are examined in terms of fundamental biological and chemical properties (e.g., “What makes antibodies ‘specific’?” “Do antibody probes bind the same way that riboprobes do?” “How can you tell whether a particular stem cell undergoes division after injury to an organism?”), students review fundamental content from previous course work in a new context. By considering “evolution of methodology” (e.g., differential screening of cDNA libraries vs. gene chip analysis vs. RNAseq approaches; gene knockout vs. RNA interference) students become aware of the pace of technique development and how the range of tools available may influence the nature of questions asked. In this way, Biology 35500, the original CREATE course, involves both close analysis of papers presented in their original sequence as an individual “module” but also consideration of broader nature of science issues. For example, discussion centered on the fact that what used to be considered “junk” DNA is now recognized as having a key role in microRNA pathways illustrates the malleability of scientific knowledge.

After completing analysis of each paper, and before moving to the next paper in the series, students create their own follow-up experiments, thereby building experimental design skills, as well as awareness that a given study could, in principle, move forward in a variety of ways. Students’ proposed follow-ups are vetted in a grant panel exercise designed to mimic activities of bona fide panels (see Hoskins *et al.*, 2007). In turn, these sessions lead to discussion focused

on broader scientific issues, including interlaboratory competition, peer review, and the factors that might influence principal investigator (PI) decisions about what direction to take next.

Late in the semester, students, as a class, develop a list of 10–12 questions for paper authors. These are emailed as a single survey to each author (PIs, postdocs, graduate students). Many authors reply with thoughtful comments about their own paths to science, their motivations, and their lives beyond the laboratory. Discussion of authors’ varied responses complement the in-class data analysis with insight into the lives and motivations of “the people behind the papers.”

Our upper-level course led to gains in students’ content integration and critical-thinking ability, as well as in their self-assessed learning gains (Hoskins *et al.*, 2007). We also found that undergraduates’ self-assessed science abilities, attitudes, and epistemological beliefs changed during the CREATE semester (Hoskins *et al.*, 2011). Upper-level students’ postcourse interviews (see Tables 1 and S1 in Hoskins *et al.*, 2007), as well as conversations with alumni of Biology 35500 (“You have to do a version of this for freshmen—it changed how I read everything” and “If I had known sooner that research wasn’t boring, I might have joined an undergrad research program”) inspired us to consider adapting upper-level CREATE for freshmen.

A related motivation for developing the CREATE Cornerstone course was that the biology department at CCNY, like its counterparts elsewhere, loses many would-be majors during the early years of the biology curriculum. Some students who start with the intention of declaring a biology major do not follow through. Others who do choose biology later change majors and leave science altogether, with multiple factors likely playing a role. Students may be poorly prepared for college-level science, feel overwhelmed by the amount of new information covered in the introductory-level courses (Seymour and Hewitt, 1997), or be discouraged by textbooks’ depiction of biology as a largely descriptive science (Duncan *et al.*, 2011). Nationwide, some students get the impression from the laboratory components of introductory biology, chemistry, or physics classes that lab work is routine, predictable, and boring.

We felt that a CREATE Cornerstone course focused on scientific thinking could support and build students’ science interest at an early phase of their academic careers. In part, adapting upper-level CREATE for freshmen might benefit students by teaching them a variety of techniques (the CREATE toolkit; Hoskins and Stevens, 2009) that make complex material more accessible and understandable. At the same time, the course seeks to provide students with an inside look at the workings of real-world biology research labs and the diversity and creativity of the scientists who work in them. We hypothesized that students in such a course would become more adept at thinking critically about scientific material and at designing and interpreting experiments—key strategic foci of the CREATE approach. In addition, we hypothesized that students would gain in their abilities to critically analyze scientific writing, deepen their understanding of the nature of science, and develop more mature epistemological beliefs about scientific knowledge. We also suspected that some students who had not considered careers in research, or others who had but quickly rejected the idea, would consider research more positively as their college education progressed.

Introduction to Scientific Thinking is a three-credit, one-semester elective for first-year college students with a declared interest in science, technology, engineering, and math (STEM) disciplines at the CCNY, a minority-serving institution. The course meets twice-weekly for 75 min/session, and on our campus is taken before the introductory-level courses in any of the basic sciences. The goal is to develop the science-related reading and analytical skills of freshmen by using the CREATE strategy to critically evaluate a number of recent and ongoing research studies. Ideally, the experience should also encourage students to persist in STEM disciplines, participate in undergraduate research experiences (UREs) in later years, and consider research as a career choice.

At CCNY, first-year students cannot declare a biology major. The course is thus aimed at presumptive biology majors and in principle could be taken concomitantly with the standard introductory biology (or other science) course. On campuses where students can or must declare a major in the first year, this course would be appropriate for students who evince interest in biology studies. The data reported here address changes in Biology 10050 students' critical-thinking/experimental design abilities and in their attitudes and beliefs about science. The question of student persistence in STEM and participation in undergraduate research projects will be tracked in upcoming semesters.

METHODS AND ASSESSMENT TOOLS

Participants in this study were first-year students at CCNY who enrolled in the semester-long Biology 10050: Introduction to Scientific Thinking course during Fall 2011 and Spring 2012. In each semester, at the first class session, students were invited to participate anonymously in our study on a voluntary basis that had no bearing on class grade. Precourse data were collected during the first few classes and postcourse data in the final class session of the semester. All participating students were asked to devise a "secret code" number known only to them and to use this code on all surveys. Identifying surveys in this way allowed us to compare individual and group scores pre- and postcourse, while preserving student anonymity (Hoskins *et al.*, 2007).

Critical Thinking Assessment Test (CAT). Students in the Fall cohort of Biology 10050 completed the CAT (Stein *et al.*, 2012). In the CAT, which is a reliable and valid test of critical thinking, students spent 1 h reading a number of informational passages and writing responses to a variety of prompts asking them to evaluate the information and draw conclusions. The same test was taken again at the end of the semester. The CAT tests were graded and analyzed statistically (Student's *t* test) by a scoring team at Tennessee Tech University, where this survey was created.

Experimental Design Ability Test (EDAT). Students in both cohorts of Biology 10050 also completed the EDAT, the reliability and validity of which have been established by the EDAT developers (Sirum and Humburg, 2011). In the EDAT, students were presented with a claim and challenged to "provide details of an investigative design" and indicate the evidence that would help them decide whether to accept the claim. Students were given 15 min to respond to a written prompt that described the assertion. Precourse and post-

course versions of the EDAT present different scenarios. Precourse, students read a paragraph presenting the claim that the herb ginseng enhances endurance; postcourse, the selected text alleged that iron supplements boost memory. The EDAT survey was scored separately by two investigators following the scoring rubric created and explained in Sirum and Humburg (2011). After the individual scoring, any discrepancies were discussed and reconciled. Tests for statistical significance were performed using the Wilcoxon signed-rank test (<http://vassarstats.net/index.html>; Arora and Malhan, 2010). Effect sizes (Cohen, 1992; Coe, 2002) were also determined.

Survey of Student Self-Rated Abilities, Attitudes, and Beliefs (SAAB). To investigate students' reactions to the CREATE course, we asked them to complete the SAAB. In this Likert-style survey, students reported their degree of agreement on a 5-point scale (range: strongly disagree to strongly agree) with a series of statements concerning their attitudes, self-rated abilities, and beliefs about analyzing scientific literature; the research process; the nature of scientific knowledge; and scientists and their motivations. The surveys were identical precourse and post course, and used statements whose derivation and description is described in Hoskins *et al.* (2011). Students were given 20 min to complete the survey. For statistical analysis, all response scores were aggregated into their appropriate categories (see Supplemental Material for derivation of categories) and changes precourse to postcourse were analyzed for statistical significance using the Wilcoxon signed-rank test. Because these data and those of the EDAT are nonparametric (a score of "4" is not twice as good as a score of "2," for example) and noncontinuous, the signed-rank test was deemed an appropriate analytical tool (Arora and Malhan, 2010).

The SAAB data for the Biology 10050 class include pooled results from the Fall and Spring sections (18 and 13 participating students, respectively). Data collected using the same survey, administered in the same manner, were also obtained from one contemporaneous section of the upper-level CREATE course (Biology 35500, 21 students; two meetings per week for 100 min/session). Additionally, the SAAB survey was administered to volunteers in a course in Organismic Biology (general physiology, 23 students; one 100-min lecture and one 3.5-h lab session/wk), none of whom had taken a CREATE class. This group was not a matched-control population (students were not freshmen). Rather, data from this cohort of students provided insight into potential changes in attitudes, abilities, and epistemological beliefs that might happen naturally during the course of a semester in a non-CREATE science class. The CREATE classes were taught by the same instructor (S.G.H.); the Organismic Biology class was taught by a colleague not otherwise involved in this study. Both instructors were experienced at teaching their respective courses.

Student Comments on Author Emails. To gain insight into students' reactions to author email responses, we assigned students to read and annotate the responses as they were received. Students included the responses in their notebooks/portfolios, with marginal notes indicating which aspects of each response they found most surprising and/or interesting. In the Spring session of Biology10050, we included a question on a late-semester (in-class, open-book) exam,

asking students whether the emails changed their ideas about science research or scientists. We compiled responses and analyzed them for repeated themes.

Student Participation. The CREATE study was approved by CUNY Institutional Review Board (Exemption category 1 and 2). Of the students in Bio 10050, 69% were female and 59% were members of minority groups currently underrepresented in academic science. Students were invited, in week 1 of class, to anonymously participate in an education study with the goal of “improving undergraduate education in science.” Participation was optional and the instructor noted that student participation or nonparticipation had no bearing on course grade or any other relationships with CCNY. There were no points or extra credit awarded for participation. We think that students who participated were motivated by the chance to take part in a science education study and/or to be part of a scientific experiment.

CURRICULAR DESIGN

Adapting CREATE for Freshmen

In the original (upper-level) CREATE course, the class studied, sequentially, a series of papers published by a single lab that tracked the development of understanding in a particular field of scientific inquiry (e.g., how embryonic retinal axons find their targets in the brain; how planaria maintain positional information during regeneration). For the freshmen, we changed the types of articles studied, using popular press articles and a wider range of scientific literature, but applied the same overall CREATE teaching/learning strategies. The freshmen initially read and analyzed numerous popular press stories based on journal articles. We also read a variety of newspaper and magazine pieces describing scientific investigations or researchers. These warm-up exercises, used more extensively for the freshmen than in upper-level CREATE, started students toward developing the skills they would need for reading and analyzing primary literature later in the semester. All the readings (in all CREATE courses) are actual texts as originally published. In some cases, we read only parts of papers, but we did not rewrite or simplify any of the material. The freshmen ultimately read a pair of papers published in sequence that addressed a subject—the ability of infants to recognize and judge the social actions of others—related to a number of the shorter readings.

Toward the end of the semester, the freshmen, as a class, composed a list of 10–12 questions about the studies we had read, “research life,” and the researchers themselves. These questions were emailed as a single survey to each paper’s authors, with a cover letter explaining our approach and inviting a response. This key strategic component of CREATE courses seeks to shift students’ often-negative preconceptions about what research/researchers/research careers are like. Many of the scientist-authors responded with comprehensive answers related to their personal and professional lives, their contributions to the work that we studied, and their scientific experiences as their careers developed. The generosity of authors in preparing thoughtful responses is especially valuable and memorable, according to our students.

CREATE Cornerstone Objectives and Selected Exercises

Students learned to use CREATE tools, including concept mapping, paraphrasing, cartooning, annotating figures, applying templates to parse experimental logic, designing follow-up experiments, and participating in grant panels (Hoskins and Stevens, 2009). The CREATE techniques aim to sharpen students’ analytical skills and build critical-reading habits that can be used in new situations. These approaches also build students’ metacognition—the ability to track their own understanding (Tanner, 2012). To construct a concept map successfully, for example, students need to understand individual ideas and discern the relationships between them. To sketch a cartoon that shows what took place in the lab to generate the data presented in a particular figure, students must make sure they understand the relevant methodology. We applied concept mapping and cartooning along with other CREATE tools to a novel combination of readings. Articles selected for Biology 10050 were chosen because of their topicality, relatively simple methodology, and aspects of each that provoked controversy, exemplified the role of controls, and/or highlighted important distinctions between data and their interpretation. Goals for the Cornerstone students included learning: to read with skepticism, to critically analyze data and generate alternative interpretations, to recognize the malleability of scientific knowledge, and to develop and evaluate experiments with particular emphasis on controls and their roles. A final goal was for students to develop a more realistic view of research and researchers than the one often promoted in popular culture.

Developing an Appropriately Skeptical Reading Style

The class sessions were typically run as discussions or debates about points that arose in the assigned readings. We rarely presented all the information at once, instead examining each reading in stages. For example, one unit early in the semester used an op-ed in the *New York Times* claiming that iPhone owners experienced “love” for their phones and outlining study outcomes that purported to support this conclusion (Lindstrom, 2011). We also read a published refutation of the op-ed signed by 44 neuroscientists (Poldrack, 2011a), and the original version of the refutation letter before it was edited by the *New York Times* (Poldrack, 2011b). We started with the op-ed and only later distributed the challenge from the neuroscience community, considering:

How, in principle, would one determine “the most appealing sounds in the world,” whether babies “automatically” swipe iPhones expecting a response, or whether “love” is experienced by phone owners (as claimed by Lindstrom, 2011)?

What evidence would you find convincing?

What studies would you do if you were interested in such issues?

How did Lindstrom make such determinations?

On what basis do the neuroscientists challenge the stated conclusions?

Do the *New York Times*’ edits shift the message of the original letter to the editor? If so, how?

Taking all of the readings and analyses together, what do you conclude about iPhone “love”? Why?

As they learned to use and apply CREATE tools, students accustomed to reading and passively accepting the information encountered in their textbooks, on the Internet, or in newspapers began to recognize that just because something is published does not mean it is beyond criticism (Hoskins, 2010a).

Data Analysis—Developing Alternative Interpretations

“Writing about Testing Worries Boosts Exam Performance in the Classroom” (Ramirez and Beilock, 2011) is a *Science* paper examining the degree to which stress may contribute to undergraduates’ “choking” on exams. We initially distributed only some of the paper’s narrative and a single figure illustrating the first study performed, holding back the title, abstract, and all other information. During class, students diagrammed the experiment, which compared test scores of two groups of students. Each group had been administered a baseline math test. Posttest, both groups were told a stress-inducing story about how outcomes on a later test covering the same material would be used. Before taking the second test, one group wrote for 10 min about their fears of poor test performance, while the other group sat for 10 min. The data revealing the test scores of the two groups show the nonwriting group performing worse on the second test than they did on the first, thus “choking,” while the writing group scored gains. We considered:

Can we conclude that writing about one’s test concerns leads to less choking on exams? How solid is that conclusion?
If we had generated these data ourselves, could we publish now? Why? Why not?
Are any alternative interpretations of the data plausible?

Through discussion, students proposed a third “write about anything” group as an additional control. We next provided the paper’s figure 2 and associated narrative. The authors had added a third group that was instructed to write about “an unrelated unemotional event.” Students saw that the investigators had added the same control group they had asked for, extending the study to resolve the “writing-only” issue. This bolstered students’ sense that they were “thinking like scientists.”

Using Sketching to Clarify Design—Developing Alternative Interpretations

One paper’s abstract alone served as the focus for a class. The abstract for “Empathy and Pro-Social Behavior in Rats” (Bartal *et al.*, 2011) outlines five individual experiments. As homework, students cartooned each experiment, all of which tested conditions under which one rat would open a transparent plastic container that restrained a second rat. Students defined the specific hypothesis being addressed in each study, the controls needed in each case (none are included in the abstract), the conclusions stated, and possible alternative interpretations.

After comparing cartoons and resolving discrepancies, the class considered whether the behaviors observed were necessarily signs of “empathy.” Might there be other explanations? Working in small groups, students proposed multiple alternatives that could in principle account for rats’ appar-

ently helpful behavior: inquisitiveness, a pheromone signal, an aversion to squeaky distress calls, and the like. The published paper provoked substantial interest and some controversy, as reported in *Nature* (Gewin, 2011). We reviewed the published critique, and students found that some of “our” alternative interpretations had also been raised by top scientists in the field, again recognizing that their own thinking was scientific. Students also noted that even peer-reviewed work published in *Science*, where the original article appeared, can evoke intelligent criticism, and that scientists do not always agree.

Established Knowledge Can Change

A provocative set of readings discuss the discovery that peptic ulcers have a bacterial origin (Associated Press, 2005; Centers for Disease Control and Prevention, 2005). It took the PI’s ingestion of *Helicobacter pylori*, the suspected pathogen, hardly a canonical step in “The Scientific Method,” to generate the conclusive data. This nature of science story illustrates how established scientific knowledge—that ulcers had psychological not bacteriological etiology—can be wrong. Reading the description of Dr. Barry Marshall being met with scorn at meetings where he initially presented his unconventional hypothesis, students saw that novel (and possibly revolutionary) ideas may not be instantly welcomed. This recent scientific development highlighted the personal factors and genuine passion that can underlie science, making the point that as scientific study continues, some established ideas of today will inevitably be supplanted. The ulcer readings also illustrated the value of a healthy skepticism even about “obvious” facts, such as that the stomach’s acidity would kill all bacteria within.

Introducing Experimental Design and Peer Review

At the conclusion of many of the discussion units, the freshmen proposed follow-up experiments. The challenge: If your research team had just performed the work we reviewed, what would you do next? Each student independently devised two distinct follow-ups as homework. Three or four times during the semester, students formed teams of four to act as grant panels charged with assessing the studies designed by their peers. The first time this was done, we challenged the panels to establish appropriate funding criteria before looking at the proposed studies. Discussions of criteria led to consideration of evolution, evolutionarily conserved mechanisms, and the meaning of model systems, as many groups only wanted to fund work that is “relevant to humans.” We also discussed realities of reputation and how it may affect funding success. Some groups sought to fund “established investigators who have already published in the field,” leading other students to question how anyone gets started in research. Such discussions build students’ understanding of the sociological context of science.

After criteria had been discussed, each student submitted one of his or her experiments, sans name or other identifier, into the grant pool. The instructor then presented each proposed follow-up study to the class without evaluative comments. When the panels subsequently conferred to rank the proposed experiments, students thought critically about the work of their peers, debating and defending their judgments in the sort of open-ended give-and-take that characterizes

science as it is practiced. There is no single correct answer to the question: “Which of the ≈ 25 proposed studies is the best?” Students were thus freed from the pressure to be right, or to divine, somehow, what the instructor’s opinion might have been.

Using Multiple Popular Press Articles to Build Toward a Mini-Module of Primary Literature

We developed students’ critical-reading skills through repeated practice with short articles. In the process, we pointed out multiple aspects of scientific thinking, and introduced the subject matter knowledge that would be needed in the later reading of primary research reports exploring infant cognition. Early in the semester, we read and analyzed “Babies Recognize Faces Better Than Adults, Study Says” (Mayell, 2005) and a popular press account of “Plasticity of Face Processing in Infancy” (Pascalis *et al.*, 2005), a study that tested the memories of 6- to 9-mo-old infants. Students discovered gaps in the popular press version (no information on “*n*” or gender distribution of infant subjects, and unclear methodology, for example). We added additional information from the *Proceedings of the National Academy of Science* paper as discussion required it (for details of teaching with this paper, see Hoskins, 2010b). Exercises of this sort challenge students to read actively and seek key missing information (e.g., “How many female vs. male babies were studied?” or “Exactly how was the visual training done?”) that is essential to their evaluations.

Two additional popular press stories (Talbot, 2006; Angier, 2012) and a study on babies’ perception of normal versus scrambled facial features (Maurer and Barrera, 1981) were critically analyzed in other class sessions. Discussions covered broader questions including: How can you tell whether a baby who is too young to talk notices something novel, and why might it matter? Because one of the studies was funded by the National Institutes of Health, we considered how a real-life grant panel might evaluate the work’s health relevance. Students raised the possibility of using methods from the infant studies for early detection of neurological abnormalities, such as autism, and discussed the degree to which environmental enrichment activities could be considered “health related.” These readings and discussions set the stage for the analysis of two full-length papers.

“Social Evaluation by Preverbal Infants” (Hamlin *et al.*, 2007), examines 6- and 10-mo-old babies’ abilities to discriminate between and react to helpful, neutral, and hindering “behaviors” by observed “characters.” The babies witnessed scenarios in which experimenter-manipulated blocks of wood bearing large googly eyes interacted on a hill. One sort of block (e.g., red circle) would move partway up the hill, but slide down before reaching the summit. Another block (e.g., yellow square) might, in a subsequent “episode,” seemingly help it move up. A third block (blue triangle) might hinder upward movement. A series of control experiments explored the need for eyes and other animating traits on the blocks. Other controls investigated whether babies preferred particular colors/shapes, or upward motion to downward, rather than seemingly helpful interactions (which moved the target block up) to hindering ones (which moved it down).

We started by providing the introduction, first figure, and associated text for initial analysis. As before, we did not

tell the students what additional experiments were done. Through class discussion, students developed their own questions and alternative interpretations (e.g., “maybe the babies aren’t judging behavior; they just like yellow better than blue”). As in the discussions of “Babies recognize faces” and “Writing about testing. . .,” only after the students raised particular issues did we provide sections of the paper with the relevant additional information and control experiments. After analyzing the full paper, students designed follow-up experiments, vetted them in a grant panel, and then read and analyzed the authors’ actual next paper.

“Three-Month-Olds Show a Negativity Bias in Their Social Evaluations” (Hamlin *et al.*, 2010) was concerned with younger babies’ reactions to similar social interactions. This second paper used many of the same methods as the first, facilitating students’ ability to read the material. Interestingly, the later work produced a different result, finding that younger babies were averse to hinderers but (unlike their “elders”) did not show any particular preference for helpers. As the authors discussed possible evolutionary implications of their work, we were able to return to a critical theme that had arisen earlier in the semester, in the “model systems” discussion.

Assessment in Biology 10050

The study presented here is based on tools (CAT, EDAT, SAAB) administered anonymously pre- and postcourse. To evaluate students’ understanding of course material as a basis for determining grades, we assess students in all CREATE classes using a combination of in-class activities; writing assignments; open-book, open-notes exams; and class participation. There is no assigned textbook, but students can consult during exams the notebooks/portfolios they compiled throughout the semester (see Hoskins *et al.*, 2007, for details). We find that open-book testing changes the classroom atmosphere and relieves students from the pressure to study primarily by memorizing, making it easier for them to focus on critically evaluating scientific writing and explaining their insights. With the exception of analysis of one exam question (see *Student Reactions to Emails*, below), the classroom assessments were not used as data for this study.

RESULTS

CAT Outcomes

Students in the Fall CREATE Cornerstone course took the CAT (Table 1; Stein *et al.*, 2012), and tests were scored by a trained team at Tennessee Tech University, where this test was created. Biology 10050 students’ overall CAT scores improved significantly postcourse versus precourse, with a large effect size (0.97). While there is overlap between categories, CAT questions address four main areas. Overall, the largest gains made by CREATE Cornerstone students were on CAT questions that tested “evaluating and interpreting information.” Students also made gains on questions involving problem solving, creative thinking, and/or effective communication (the other three subcategories addressed by the CAT). While these findings must be interpreted with caution due to the small sample size, they suggest that students in the pilot CREATE Cornerstone course made substantial gains in their

Table 1. CAT test results^a

Critical Thinking Ability Test (CAT)	Precourse	Postcourse	<i>n</i>	Significance	Effect size
Mean (SD)	9.6 (2.5)	13.0 (4.4)	15	$p < 0.05$	0.97

^aThe CAT (duration 1 h) was administered pre- and postcourse to the Fall 2011 Biology 10050 class and scored at Tennessee Tech University. We present the overall score for the test, precourse vs. postcourse. Fifteen students took both tests. Significance: Student's *t* test.

ability to read, understand, and critically analyze information, and that such gains are transferable to the content domain addressed by the CAT test, which was not related to the material covered in the course.

EDAT Outcomes

Students in both Fall and Spring CREATE Cornerstone classes completed a pre- and postcourse EDAT that was scored using a 10-point rubric (Sirum and Humburg, 2011). Results are summarized in Table 2. Scores suggest that the first-year students gained significantly in experimental design ability over the semester, citing more components of an "ideal" experimental design postcourse than precourse.

SAAB Outcomes

Results from the SAAB surveys for each class are displayed in two groupings in Table 3. The upper group reflects the items related to students' self-rated skills and understanding; the lower group shows results for items that reflect students' epistemological beliefs about science (see Hoskins *et al.*, 2011, for a discussion of the derivations of all categories).

SAAB results show significant gains made by CREATE Cornerstone students in all six skills and attitudes categories and in the majority (four out of seven) of epistemological categories. Students in the upper-level CREATE course (for which a year of introductory biology, a semester of genetics, and a semester of cell biology are prerequisites) shifted significantly on all skills and attitudes categories, and three of the seven epistemological categories. Students in the mid-level physiology course (for which a year of introductory biology and a semester of genetics are prerequisites), in contrast, did not shift significantly in any category.

Effect sizes help to determine whether statistically significant changes are likely to be meaningful. For skills and attitudes shifts, effect sizes for freshmen were large (Cohen, 1992) in five of the six categories and moderate for "interpreting data." Effect sizes for upper-level CREATE students in these categories were all large. In this regard, it may be relevant that upper-level students read literature that was substantially more complex and looked closely at more fig-

ures during the semester than did the first-year students. It is also interesting to note that the mid-level physiology course included a weekly laboratory, in which data were generated and analyzed, and one experimental design activity.

For epistemological beliefs categories, effect sizes in three of the four categories that shifted significantly in the freshman CREATE group (certainty of knowledge, innate ability, creativity of science) were moderate. The effect size of "sense of scientists as people" was large. Upper-level CREATE students also shifted significantly in this category, but with a smaller effect size, possibly reflecting the fact that many upper-level students were working in labs and had a better sense precourse of what research scientists were like. Upper-level CREATE students also showed significant changes in understanding of the uncertainty of scientific knowledge (large effect size), and of "sense of scientists' motivations" (moderate effect size).

Both the CREATE courses, but not the mid-level physiology course, sent email surveys to authors of papers and discussed author responses late in the semester. Different material was read and analyzed in each CREATE course; thus, different authors were queried and different responses were received by the two groups. We think it likely that this component of the CREATE courses played a large role in changing students' opinions about what scientists are like and (for upper-level CREATE students) why they do what they do.

Student Reactions to Emails

On the second exam in the Spring semester, we included a question asking students about their reactions to the author emails, focusing on their preconceptions about "scientists/research careers" and whether the author responses changed these views. We coded all the responses ($n = 15$), extracting key themes from each, and summarize below the themes mentioned by four or more students.

The most prevalent response to the emails was students' statements that, precourse, they had assumed today's researchers were "straight-A" students in college (14/15 responses; 93% of students). The same students (14/15) noted that they no longer believed this to be true, citing several authors who described academic struggles that preceded their eventual success. Thirteen out of 15 students (86%) said that the responses had changed their preconceptions about researchers, and 9/15 (60%) noted that respondents stressed the importance of passion (as opposed to good grades) as a key to research success. Seven out of 15 students (47%) expressed enthusiasm on learning that the responding scientists described a great deal of work-related travel, including international travel. Forty percent of students (6/15) described having held one or more of the preconceptions that 1) scientists were loners or nerds, 2) who lacked social lives, 3) because science consumed all their time. A similar

Table 2. EDAT results: mean and SD^a

EDAT test	Precourse	Postcourse	<i>n</i>	Significance	Effect size
Mean (SD)	4.3 (2.1)	5.9 (1.4)	28	$p < 0.01$	0.91

^aPool of two classes of Biology 10050: $n = 28$ total. Statistical significance tested with Wilcoxon signed-rank test. Scores can range from 0 to 10, per the EDAT rubric (see Sirum and Humburg, 2011).

Table 3. SAAB survey outcomes in three student cohorts: freshman CREATE students ($n = 28$), upper-level CREATE students ($n = 19$), and mid-level non-CREATE students ($n = 23$)^a

Freshman-level CREATE class					
Category	Precourse mean (SD)	Postcourse mean (SD)	Significance ^b	Effect ^c	# Ss ^d
Decoding literature	17.3 (3.2)	21.9 (3.0)	<0.001	1.48	6
Interpreting data	14.1 (2.6)	15.4 (2.3)	0.008	0.53	4
Active reading	13.0 (2.4)	16.1 (2.3)	<0.001	1.32	4
Visualization	12.5 (2.8)	15.6 (2.1)	<0.001	1.27	4
Think like a scientist	11.9 (2.2)	15.5 (1.6)	<0.001	1.90	4
Research in context	10.9 (1.5)	13.8 (1.4)	<0.001	2.00	3
Certainty of knowledge	22.1 (2.9)	24.3 (3.8)	0.002	0.66	6
Ability is innate	6.9 (1.4)	8.0 (1.6)	0.005	0.73	2
Science is creative	3.9 (0.8)	4.3 (0.7)	0.005	0.53	1
Scientists as people	2.8 (0.9)	3.6 (1.0)	0.004	0.84	1
Scientists' motives	3.8 (1.1)	4.1 (0.8)	ns	0.35	1
Known outcomes	3.9 (1.0)	4.2 (0.9)	ns	0.32	1
Collaboration	4.2 (0.6)	4.3 (0.8)	ns	0.14	1
Upper-level CREATE class					
Category	Precourse mean	Postcourse mean	Significance ^b	Effect ^c	# Ss ^d
Decoding literature	15.5 (2.8)	20.3 (2.7)	<0.001	1.75	6
Interpreting data	13.7 (2.2)	16.4 (1.7)	<0.001	1.39	4
Active reading	13.9 (2.1)	16.9 (1.9)	<0.001	1.50	4
Visualization	13.3 (2.1)	16.6 (1.7)	<0.001	1.74	4
Think like a scientist	13.3 (2.5)	16.5 (2.1)	<0.001	1.39	4
Research in context	13.5 (1.1)	14.3 (0.9)	0.037	0.80	3
Certainty of knowledge	23.0 (2.7)	26.1 (2.9)	0.021	0.82	6
Ability is innate	7.3 (1.9)	8.4 (1.5)	ns	0.65	2
Science is creative	4.1 (0.7)	4.6 (0.9)	ns	0.63	1
Scientists as people	2.5 (1.0)	3.9 (0.8)	0.007	0.44	1
Scientists' motives	3.8 (1.0)	4.3 (0.6)	0.014	0.63	1
Known outcomes	3.7 (1.0)	4.3 (1.1)	ns	0.57	1
collaboration	4.4 (0.6)	4.5 (0.6)	ns	0.17	1
Mid-level non-CREATE class					
Category	Precourse mean	Postcourse mean	Significance ^b	Effect ^c	# Ss ^d
Decoding literature	19.6 (4.3)	20.4 (3.3)	ns	0.21	6
Interpreting data	15.3 (2.3)	15.7 (2.6)	ns	0.16	4
Active reading	14.7 (2.3)	14.7 (2.4)	ns	0.01	4
Visualization	14.0 (2.3)	14.7 (1.8)	ns	0.34	4
Think like a scientist	13.8 (2.7)	13.8 (2.5)	ns	0.00	4
Research in context	13.5 (1.1)	13.5 (1.4)	ns	0.00	3
Certainty of knowledge	23.7 (3.3)	23.6 (3.2)	ns	-0.03	6
Ability is innate	7.3 (1.5)	7.6 (1.5)	ns	0.20	2
Science is creative	3.8 (1.2)	4.2 (0.7)	ns	0.42	1
Scientists as people	3.1 (0.9)	3.3 (1.1)	ns	0.03	1
Scientists' motives	3.9 (0.9)	4.0 (1.1)	ns	0.10	1
Known outcomes	3.9 (1.0)	4.0 (0.8)	ns	0.11	1
Collaboration	4.4 (0.5)	4.2 (0.6)	ns	-0.36	1

^aResponses were tabulated using a 1–5 scale. (1 = “I strongly disagree”; 2 = “I disagree”; 3 = “I am neutral”; 4 = “I agree”; 5 = “I strongly agree”). Some propositions were worded so that an answer reflecting a more mature understanding would get a lower score (“I accept the information about science presented in newspaper articles without challenging it,” for example). These were reverse-scored for analysis.

The Wilcoxon signed-rank test for statistical significance was performed on precourse/postcourse raw data totals for all categories. Category 1–6: self-rated skills and attitude factors; categories 7–13: epistemological factors.

The survey was developed in a previous study of upper-level CREATE students (Hoskins *et al.*, 2011). Different categories are probed by different numbers of statements (#Ss).

^b p values for statistical significance (Wilcoxon signed-rank test). ns = not significant.

^cMean difference/average SD.

^d# Ss = number of statements in category.

percentage noted that precourse they had assumed all scientists had lofty goals of “helping people,” but they had come to realize that many had more personal goals of satisfying their own curiosity. Five out of 15 students (33%) stated that precourse they had assumed most scientists did *not* enjoy their jobs, that research was not fun, and that lab life was boring, but they no longer held these views. Five out of 15 (33%) said they were surprised to learn scientists had flexible work schedules, and a similar percentage stated that they had learned from the emails that motivation was very important. Finally, 4/15 (27%) noted their surprise that the authors answered at all.

DISCUSSION

Genesis of the CREATE Strategy

The CREATE strategy originated as a response to the observation that many upper-level undergraduate biology majors—despite the years spent studying a wide range of scientific topics—were not well-prepared to read and understand primary literature; did not readily “think like scientists,” with an appropriately critical eye; did not see science research as an attractive career choice; and had little or no practical experience mustering their content knowledge to attack novel scientific problems. Discussions with students in other courses over the years, and with other faculty on our campus and elsewhere, revealed that many students believed: research is dull, and lab exercises formulaic and boring (Luckie *et al.*, 2004); there is a single and eternal right answer to every scientific question (Liu and Tsai, 2008); primary literature is written in a nearly unbreakable code; and scientists themselves are stereotypic nerds or “machinery kind of people” (Hoskins *et al.*, 2007). Our findings in the pilot CREATE Cornerstone course suggest that these viewpoints can be changed over a single semester through intensive analysis of scientific literature.

Themes Highlighted in Readings

The curriculum examples outlined above illustrate how fundamental features of scientific thinking can be studied in a realistic domain-specific context, which appears to be a key element in developing critical-thinking skills (Willingham, 2007). Students repeatedly thought carefully about control groups—what they “control” for, how they are interpreted, and why they are needed. Multiple studies underscored the importance of careful attention to sample size and selection. In the experiments on infants, for example, students raised issues of possible gender-related behavioral differences, whether postnatal age is comparable between full-term and premature infants, and the like. Students practiced developing alternative interpretations of data and noted that not all conclusions are equally strong. Several studies highlighted the potential for introducing unanticipated bias (see discussion of a possible “Clever Hans” effect in “Babies Recognize Faces Better Than Adults, Study Says” in Hoskins, 2010b). Students saw that original, interesting, and important investigations are currently ongoing (many readings were published in 2011–2012). Students also recognized that even very early in their academic careers they are capable of reading, understanding, and intelligently criticizing scientific litera-

ture, and that research science is neither routine, predictable, or boring, nor something found only in textbooks.

Grant Panels Promote Open-Ended Thinking and Insight into the Nature of Science. CREATE Cornerstone students made significant gains on the EDAT, which presents a scenario distinct from that of any of the Cornerstone readings. Students’ gains on this test suggest that their general experimental design skills have improved during the semester.

Experimental design skills are honed in class through grant panel activities that focus on follow-up experiments to the studies we analyzed that are designed by the students as homework. These are repeated several times during the semester. Although panels focus specifically on experimental systems under study in class, they likely help students develop a more generalized skill in experimental design and creative thinking. In each panel all students’ experiments are reviewed, and the panels (groups of four students) discuss the merits of each. Early in the semester, some experiments must be culled based on absence of a hypothesis, absence of a cartoon, or general lack of clarity (approximately five of 20 in early panels). In end-of-semester exercises, virtually every experiment meets the basic criteria and can be considered seriously. Statements of hypotheses become clearer, controls stronger, designs and procedures better illustrated, and potential outcomes well anticipated.

Besides likely contributing to the development of students’ experimental design skills, the grant panels provide insights into the nature of science. It becomes evident, as the activity is repeated during the semester that, among the top experiments (typically four or five stand out), the study perceived by a particular panel to be “best” is to some degree a matter of taste. Some students prefer a reductionist approach, others an expansion of the study to encompass additional sensory modalities (e.g., an experiment investigating whether babies learn to recognize faces faster if each face is associated with a different musical tune). Some students focus mainly on experiments aimed at developing treatments for humans (e.g., take genes involved in planarian regeneration and immediately seek their counterparts in mammals). Many of our students are accustomed to “textbook” science where, typically, only the (successful) end points of studies are described, and very little current-day work is featured. The grant panel activity introduces the idea that working scientists likely select their follow-up experiment from a variety of valid possibilities, and that personal styles and preferences could influence such decisions.

Critical-Thinking and Experimental Design Skills—Tools of Science. A significant number of students show interest in science in high school or before (often significantly before [Gopnik, 2012]), but do not pursue STEM studies at the tertiary level. Either they never consider studying science in college, or they switch out of the field for a variety of reasons in their first or second year (Seymour and Hewitt, 1997; Committee on Science and Technology, 2006). At the same time, for students who persist in STEM majors, some of the most creatively challenging and thought-provoking courses—capstone experiences—are reserved for seniors (Goyette and DeLuca, 2007; Usher *et al.*, 2011; Wiegant *et al.*, 2011). We hoped to convey some of the analytical and creative aspects of science at the outset of students’ college careers with a CREATE course designed for

freshmen. Providing this training early in students' academic experience might help students gain skills and develop attitudes that would support their persistence in STEM (Harrison *et al.*, 2011).

We used the CAT and EDAT assessments to probe the development of students' abilities as they practiced the literature analysis process. The CAT test focuses on a content domain distinct from that of the CREATE class but challenges students in some parallel ways. Students must determine what data mean, decide which data are relevant, draw conclusions based on their understanding, and explain themselves in writing. Many campuses are using the CAT test for programmatic assessment, comparing scores of freshman with those of seniors, for example. We are aware of only one published study using CAT in a pre/post, single-course situation. First-year students in a semester-long inquiry-based microbiology module at Purdue University, performing hands-on research in an introductory class, make significant CAT gains during the semester (Gasper *et al.*, 2012). The finding that CREATE Cornerstone students at CCNY similarly made significant gains on this test in a single semester suggests that transferable critical-thinking skills, such as those measured by the CAT, can also be built through classroom activities that do not involve hands-on inquiry labs.

While the small sample size in this pilot study precludes broad conclusions, it is interesting that our students made the largest gains on CAT questions whose solution required "evaluation and interpretation." Introduction to Scientific Thinking emphasizes looking closely at data, reconstructing the experiment or study that gave rise to the data, and reasoning carefully about the logic of interpretations and the significance of the findings. Students carry out this process in a variety of content domains, engaging in friendly arguments about whether rats are empathic or just noise-averse, whether writing about fears really prevents choking on tests, and what it is that babies might prefer about a yellow square with googly eyes (the color? the shape? the eyes? the "helpful" behavior?). As noted by Stanger-Hall (2012), close to 80% of U.S. high school seniors performed below the science proficiency level on a recent national standardized test (National Center for Education Statistics, 2009). Among undergraduates, barely more than half the students sampled at 24 institutions made gains in critical thinking during their first 2 yr of college, as measured by the Collegiate Learning Assessment (Arum and Roksa, 2011). These data suggest that current course work in high school and during early college years (when standard introductory science courses are taken by STEM majors) is not promoting substantial development of higher-order thinking and analytical reasoning skills. We find CREATE Cornerstone students' outcomes on the CAT assessment encouraging in this regard. At the same time, some researchers suggest results of low-stakes tests like the Collegiate Assessment of Academic Proficiency may be influenced by low performance motivation among test takers, because participation in such exercises has no bearing on class grade (Wise and DeMars, 2005). This issue could potentially influence our students' performance on anonymous assessments. While we have no independent measure of students' motivation for participating in our study, we believe it likely that, as participants in a novel course, they find the opportunity to be part of a scientific study to be intriguing and a motive to perform well.

The EDAT assessment called on students to think like scientists: analyze a problem, determine evidence required to solve it, and design a properly controlled experiment that could generate the relevant data. Students made statistically significant gains in their experimental design ability, with their postcourse responses mentioning more of the points that experts see as essential to good experimental design (Sirum and Humburg, 2011). In the Cornerstone classroom, students repeatedly proposed and evaluated experimental designs as they participated in multiple grant panels and worked with different student-colleagues. We suspect that these exercises served as a form of practice during the CREATE semester, helping students build competence in their ability to formulate, express, and defend ideas about particular proposed studies (Ambrose *et al.*, 2010). At the same time, the challenge of producing an experiment that would be singled out by a grant panel for "funding" may have stimulated some students' efforts to be particularly creative in their experimental designs.

The CAT and EDAT findings also support our sense that skills deemed important by many science faculty (e.g., problem solving/critical thinking, data interpretation, written and oral communication; Coil *et al.*, 2010), including ourselves, can be taught in a course that emphasizes the process of science, including close reading and critical analysis of primary literature, creative experimental design, and a look behind the scenes into the lives and dispositions of paper authors. While we teach or review relevant content in the context of particular reading assignments, we do not seek to achieve the broad coverage of a typical introductory course. Students need not know the details of the electron transport chain in order to analyze "Babies Recognize Faces Better Than Adults, Study Says," although they do need to know the fundamental logic of study design, use of controls, and the danger of being unwilling to think beyond your preferred hypothesis. To analyze the rat studies, students must understand the terms "empathy" and "prosocial behavior," and know how to think about variables, controls, and multiple aspects of animal behavior. In each case, they also need metacognitive awareness—the ability to determine what they do and do not understand, as well as "how we know what we know" (Tanner, 2012), another skill developed through practice during the semester.

Student Attitudes and Beliefs—Influences on Learning and Career Options. On the SAAB survey, freshmen reported significant gains in their self-rated ability to: "decode" primary literature; interpret data; read actively (annotating, concept mapping and/or cartooning the material they were reading); visualize scientific procedures; feel like they were thinking like scientists; and see experiments in a broader context (Table 3). Effect sizes (Cohen, 1992; Coe, 2002) were large for five of the SAAB measures and moderate for "interpreting data." With regard to students' epistemological beliefs, previous researchers (Perry, 1970; Baxter Magolda, 1992) have noted that students' naive epistemological beliefs about science resist change, even after a 4-yr undergraduate program. In some cases, such beliefs appear to regress after students take introductory biology courses (Smith and Wenk, 2006; Samsar *et al.*, 2011). After one semester, the freshmen in the CREATE Cornerstone course reported significant increases in four of the seven epistemological categories we surveyed:

the uncertain nature of scientific knowledge; the question of whether one needs to have a special innate ability to do science; whether science is creative; and their sense of scientists as “real people.” A concurrent upper-level CREATE class also made gains in several epistemological categories, while students in a non-CREATE comparison course that included a weekly laboratory session did not change significantly in any category (Table 3). These findings argue that the shifts we see relate to the CREATE experience, rather than to intellectual maturation that might occur naturally in college biology students over the course of a semester.

While student epistemology is rarely emphasized in college teaching handbooks, students’ attitudes in this area can strongly influence their learning. For example, students who feel that intelligence is a fixed quantity in which they are lacking may decrease their efforts to learn and study ineffectively as a result (Henderson and Dweck, 1990). The high attrition rate of students from the biology major has been attributed in large part to students’ failure to connect intellectually with the subject, and the traditional mode of teaching introductory courses itself can slow students’ development of higher-order thinking skills (e.g., analysis, synthesis, evaluation; Bloom *et al.*, 1956). While the majority of faculty members who teach introductory biology courses want students to learn higher-order skills, exams in such courses tend to focus at lower levels (Momsen *et al.*, 2010). Multiple-choice testing (often considered a practical requirement for a large lecture course) shapes students’ study habits in unproductive ways and interferes with critical thinking (Stanger-Hall, 2012). Noting that epistemological change is typically slow, Smith and Wenk point out that “. . . one cannot ignore the potential retarding effect of an entrenched instructional system of lecture, textbook readings, and recitation on the students’ epistemological development” (Smith and Wenk, 2006, p. 777). This phenomenon may be reflected in the differences in responses of CREATE and non-CREATE students on the SAAB survey.

The change in first-year students’ attitudes about scientists as people is large. We saw previously that upper-level students harbored negative opinions about scientists and the research life (Hoskins *et al.*, 2007), but we did not know whether these ideas developed during college or before. Finding that first-year students also assumed, precourse, that scientists were antisocial and that research careers were dull suggests that students finish high school and enter college with negative preconceptions about research/researchers. Multiple years of college science education apparently do little to change these ideas. The shift we saw in students’ views of scientists and research careers is likely attributable to one of the more unconventional techniques used in CREATE classes, the email survey of paper authors (see analysis of email reactions in *Results*, above). This student response to a Cornerstone exam prompt regarding author replies is typical:

I had this preconception [pre-course] that. . . you had to be like Einstein to penetrate that field. I thought you always had to have straight A’s and be highly versatile, but after reading the e-mails from the authors I know that’s definitely not the case. From what they said I know that you don’t have to be perfect or like Einstein. It’s the passion and motivation to learn and make discoveries. You have to have a drive that leads you on. It was inspiring to hear that “science has many paths” by S—[the quote in the author’s response was

“there are many paths to science”]. To me this means that there’s no one path or just one requirement like reading a textbook, but many. I can conduct research with as little a space as a backyard or in one of the biggest labs and each one could lead to success and greatness. (Exam response, freshman in Biology 10050)

Students’ self-reported reactions to author emails suggest that students starting college, at least at CCNY, harbor serious misconceptions about research/researchers that could likely interfere with their potential development as scientists. Nearly all students in the class noted that before they read the authors’ responses they had assumed that “only straight A students can become scientists.” This supposition changed when responding scientists recounted particular academic travails (e.g., rejection from some graduate schools) that preceded their success. Other student comments regarding their precourse suppositions that research is boring; that researchers are both overworked and unhappy with their jobs; and that such jobs allow no time for hobbies, families, or personal life, suggest that students’ precollege science experience has not presented research careers in an accurate light. Notably these views defy logic, suggesting that some adopted the stereotype without giving it much thought. Why would people who are so smart (“like Einstein”) and who achieved “straight As” in college choose dull, boring careers? Why would someone engaged in a boring career that he or she did not enjoy, nevertheless work so intensely that he or she had time for nothing else? We have speculated elsewhere that popular culture’s depictions of scientists may influence students negatively, starting in high school or before (Hoskins and Stevens, 2009). Changing students’ negative views of researchers/research careers is a likely required first step, if such students are to be inspired to undertake undergraduate research experiences that can lead to research careers (Harrison *et al.*, 2011). Given that the no-cost email survey of authors can have a strong positive impact on students’ views, we encourage other STEM faculty, particularly those working with high school students or first-year undergraduates, to consider this activity.

Early Interventions. Traditionally, STEM-inclined students spend their early college years in conventional core courses. Electives, including capstone courses, are reserved for upper-level students. Recently, however, a number of colleges and universities have begun developing nontraditional courses for entering students. A 5-d presemester “boot camp” for biology students at Louisiana State University aims to teach students about the different expectations at college versus high school, focusing on study strategies and introductory biology material. This brief presemester experience resulted in gains for boot-camp veterans, as compared with a matched control group, in classroom performance and in persistence in the major (Wischusen and Wischusen, 2007). In a new course focused on freshmen’s ability to reason scientifically, students studied a variety of topics that a faculty group had deemed valuable for introductory STEM courses. Students made significant gains in understanding of control of variables and proportional thinking, and also showed greater persistence in STEM (Koenig *et al.*, 2012). Freshmen at Cabrini College participated in Phage Genomics rather than a standard Introductory Biology laboratory course. The novel course involved participation in a two-semester, hands-on research project

that significantly increased students' interest in postgraduate education, their understanding of scientific research, and their persistence in the biology major (Harrison *et al.*, 2011).

Beyond Textbooks. Visual representations in journal articles are both more frequent and more complex than those seen in textbooks (Rybarczyk, 2011). When visual representations do appear in textbooks, they rarely illustrate the process of science (Duncan *et al.*, 2011). Controversy, certainly a part of science, is virtually absent from textbooks (Seethaler, 2005). Some faculty members feel that encounters with primary literature, as well as capstone courses, and the majority of undergraduate research experiences, should be reserved for upper-level students, who have built a broad foundation of content knowledge in textbook-based courses. We agree that understanding the nuts and bolts of a paper is a prerequisite for full understanding. Further, analysis and comprehension skills are better taught in the context of a particular content domain (Willingham, 2007). At the same time, particularly for biology, the explosion of fundamental content makes it impossible for faculty to cover, let alone teach, "basic" material even to the same depth it was covered in the introductory courses of their own undergraduate years (Hoskins and Stevens, 2009). In addition, despite having encountered them in multiple courses, students may fail to retain key concepts (e.g., function of control experiments; see Shi *et al.*, 2011). Our compromise was to base a freshman course on particular examples of scientific literature, choosing topics in a limited range of content areas and focusing in-depth on scientific thinking and data analysis. While we designed Introduction to Scientific Thinking for future biology majors, the approach could be easily adapted to other STEM domains. Interestingly, a recent study argues that long-term cognitive advantages can arise from studying individual topics in depth. First-year undergraduates' grades in science courses were highest for students who had studied a single topic in depth for a month or more at any time during high school. Grades showed no correlation with economic status, region, class size, parents' academic level, or other factors (Schwartz *et al.*, 2009).

Taken together, our findings support the hypothesis that a CREATE Cornerstone course designed for first-year students can bring about gains in multiple areas, including critical-thinking and experimental design ability, self-rated attitudes, abilities and epistemological beliefs, and understanding of scientists as people. Our freshman students did not have a base of content knowledge in biology beyond what they retained from high school or had absorbed from popular media. By choosing articles from top journals (e.g., *Science*, *Nature*) but focusing on topics that did not require deep understanding of, for example, gene knockout techniques or electrophoresis, we were able to give students a taste of the sorts of design logic, interpretational challenges and controversies, and creativity that are hallmarks of real-world scientific investigation. At the same time that our students gained understanding of how authentic scientific studies are carried out and interpreted, their email interviews of authors provided a personalized glimpse behind the scenes into the lives, attitudes, and motivations of the researchers themselves. Ideally, such insights will help to dispel misconceptions that can drive students away from science. To the extent that students in the one-semester Introduction to Scientific Thinking course make significant gains in scientific thinking ability, they become bet-

ter prepared to master the material in any STEM major they choose, as gains in critical-thinking and reading/analytical skills should help them manage the information load in the more content-heavy science courses to come.

CONCLUSIONS

Introduction to Scientific Thinking, the CREATE Cornerstone course, improved critical-thinking and experimental design skills of freshmen at the same time that it positively shifted their attitudes about their reading/analytical abilities, their understanding of scientists as people, and multiple aspects of their epistemological beliefs. There are few reported approaches to changing the critical thinking of first-year science students, and it appears that epistemological beliefs among college students at all undergraduate levels are quite stable. We find that a one-semester course positively affects both. The course has no laboratory component, so it is relatively inexpensive to offer. Because the topic area of the articles that can be analyzed ranges broadly, readings can be selected for their utility in a variety of introductory science courses. Finally, the email survey responses from paper authors have a strong effect on students' sense of scientists as people, helping them to overcome misconceptions of the sort that can dissuade students from seeking research opportunities and, by extension, research careers. We are encouraged by the results of this pilot study and conclude that important gains—both practical and attitudinal—with potential to help students make progress in STEM, can be achieved in a one-semester course that meets 2.5 h/wk and could, in principle, be added to many curricula.

If, as exhorted by many science-education policy reformers, we are to do a better job at encouraging students to consider research careers seriously (National Research Council, 2003; American Association for the Advancement of Science, 2011), we need to move beyond standard first-year courses and reveal scientific research as a creative and exciting career choice undertaken by interesting and diverse individuals, not unlike the first-year students themselves. While it would be gratifying to see more students enter STEM research fields, the enhancement of skills, attitudes, and epistemological beliefs concerning science engendered by CREATE Cornerstone is aligned with societal and civic goals, even for students who go in other directions.

ACKNOWLEDGMENTS

Many thanks to CCNY students for participating in the CREATE courses and/or associated assessments. We thank Dr. Millie Roth and NKem Stanley-Mbamelu, Constance Harper, and Maria Harvey of City College Academy for Professional Preparation for assistance with first-year students, and Drs. Shubha Govind, Anu Janakiraman, Kristy Kenyon, Jonathan Levitt, and Leslie Stevens for assistance, comments on the manuscript, and ongoing discussions of CREATE teaching/learning issues. Many thanks to Dr. Barry Stein, Elizabeth Lisic, and the CAT team at Tennessee Tech University for use and grading of the CAT instrument. We also thank the anonymous reviewers, whose insightful comments strengthened the manuscript. We are very grateful to the NSF for support.

REFERENCES

Ambrose SA, Bridges M, DiPietro M, Lovett M, Norman M, Mayer R (2010). *How Learning Works: Seven Research-Based Principles for Smart Teaching*. San Francisco, CA: Wiley.

- American Association for the Advancement of Science (2011). *Vision and Change in Undergraduate Biology Education*, Washington, DC. <http://visionandchange.org/finalreport> (accessed 27 June 2012).
- Angier N (2012). Insights from the minds of babes. *New York Times* 1 May, D1.
- Arora PN, Malhan PK (2010). *Biostatistics*, Mumbai, India: Global Media. (eBook accessed 25 September 2012 through CCNY library).
- Arum M, Roksa J (2011). *Academically Adrift—Limited Learning on College Campuses*, Chicago, IL: University of Chicago Press.
- Associated Press (2005). Two Australians win Nobel Prize in medicine. www.msnbc.msn.com/id/9576387/ns/health-health_care/t/two-australians-win-nobel-prize-medicine/#.UE_1RCLlaOx (accessed 11 September 2012).
- Bartal IB-A, Decety J, Mason P (2011). Empathy and pro-social behavior in rats. *Science* 334, 1427–1430.
- Baxter Magolda MB (1992). *Knowing and Reasoning in College: Gender-Related Patterns in Students' Intellectual Development*, San Francisco, CA: Jossey-Bass.
- Bloom BS, Englehart MD, Furst EJ, Hill WH, Krathwohl D (1956). *Taxonomy of Educational Objectives: Cognitive Domain*, New York: McKay.
- Centers for Disease Control and Prevention (2005). *Helicobacter pylori and Peptic Ulcer Disease*. www.cdc.gov/ulcer/history.htm (accessed 11 September 2012).
- Coe R (2002). It's the effect size, stupid. what effect size is and why it is important. Paper presented at the British Educational Research Association Annual Conference, 12–14 September 2002, Exeter, UK. www.psych.yorku.ca/mcnelles/documents/ESguide.pdf (accessed 23 September 2012).
- Cohen J (1992). Statistical power analysis. *Curr Dir Psychol Sci* 92, 98–101.
- Coil D, Wenderoth MP, Cunningham M, Dirks C (2010). Teaching the process of science: faculty perceptions and effective methodology. *CBE Life Sci Educ* 9, 524–535.
- Committee on Science and Technology (2006). *Undergraduate science, math, and engineering education: what's working?* U.S. House of Representatives Committee on Science and Technology Subcommittee on Research and Science Education. Second Session. 15 March. http://commdocs.house.gov/committees/science/hsy26481.000/hsy26481_0f.htm (accessed 1 October 2012).
- Duncan D, Lubman A, Hoskins SG (2011). Introductory biology textbooks under-represent scientific process. *J Microbiol Biol Educ* 12, 143–151.
- Gasper BJ, Minchella DJ, Weaver GC, Csonka LN, Gardner SM (2012). Adapting to osmotic stress and the process of science. *Science* 335, 1590–1591.
- Gewin V (2011). Rats free each other from cages. *Nature* 480, doi:10.1038/nature.2011.9603.
- Gopnik A (2012). Scientific thinking in young children: theoretical advances, empirical research, and policy implications. *Science* 337, 1623–1627.
- Goyette SR, DeLuca J (2007). A semester-long student-directed research project involving enzyme immunoassay: appropriate for immunology, endocrinology, or neuroscience courses. *CBE Life Sci Educ* 6, 332–342.
- Hamlin JK, Wynn K, Bloom P (2007). Social evaluation by preverbal infants. *Nature* 450, 557–559.
- Hamlin JK, Wynn K, Bloom P (2010). Three-month-olds show a negativity bias in their social evaluations. *Dev Sci* 13, 923–929.
- Harrison M, Dunbar D, Ratmanský L, Boyd K, Lopatto D (2011). Classroom-based science research at the introductory level: changes in career choices and attitude. *CBE Life Sci Educ* 10, 279–286.
- Henderson V, Dweck CS (1990). Achievement and motivation in adolescence: a new model and data. In: *At the Threshold: The Developing Adolescent*, ed. S Feldman and G Elliott, Harvard University Press, Cambridge, MA, pp 308–329.
- Hoskins SG (2010a). “But if it's in the newspaper, doesn't that mean it's true?”: developing critical reading & analysis skills by evaluating newspaper science with CREATE. *Am Biol Teach* 72, 415–420.
- Hoskins SG (2010b). Teaching science for understanding: focusing on who, what and why. In: *Science and the Educated American: A Core Component of Liberal Education*, ed. J Meinwald and JG Hildebrand, Cambridge MA, American Academy of Sciences, 151–179.
- 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.
- Koenig K, Schen S, Edwards M, Bai L (2012). Addressing STEM retention through a scientific thought and methods course. *J Coll Sci Teach* 41, 23–29.
- Lindstrom M (2011). You love your iPhone. Literally. *New York Times* 1 October, A21.
- Liu S-Y, Tsai C-C (2008). Differences in the scientific epistemological views of undergraduate students. *Int J Sci Educ* 30, 1055–1073.
- Luckie DB, Maleszewski JJ, Loznak SD, Krha M (2004). Infusion of collaborative inquiry throughout a biology curriculum increases student learning: a four-year study of “Teams and Streams.” *Adv Physiol Educ* 28, 199–209.
- Maurer D, Barrera M (1981). Infants' perception of natural and distorted arrangements of a schematic face. *Child Development* 52, 196–202.
- Mayell H (2005). Babies recognize faces better than adults, study says. *National Geographic News*. http://news.nationalgeographic.com/news/2005/03/0321_050321_babies.html (accessed 12 September 2012).
- 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 Center for Education Statistics (2009). *The Nation's Report Card 2009*, Washington, DC: Institute of Education Sciences of the U.S. Department of Education. www.nationsreportcard.gov/science_2009/science_2009_report (accessed 3 November 2012).
- National Research Council (2003). *BIO2010: Transforming Undergraduate Education for Future Research Biologists*, Washington, DC: National Academies Press.
- Pascal O, Scott LS, Kelly DJ, Shannon RW, Nicholson E, Coleman M, Nelson PA (2005). Plasticity of face processing in infancy. *Proc Natl Acad Sci USA* 102, 5297–5300.
- Perry W (1970). *Forms of Intellectual Development in the College Years: A Scheme*, San Francisco, CA: Jossey-Bass.
- Poldrack R (2011a). The iPhone and the brain. *New York Times* 5 October, A26.
- Poldrack R (2011b). NYT letter to the editor: The uncut version. www.russpoldrack.org/search?updated-min=2011-01-01T00:00:00-08:00&updated-max=2012-01-01T00:00:00-08:00&max-results=8 (accessed 11 September 2012).
- Ramirez G, Beilock SL (2011). Writing about testing worries boosts exam performance in the classroom. *Science* 331, 211–213.

- Rybarczyk (2011). Visual literacy in biology: a comparison of visual representations in textbooks and journal articles. *J Coll Sci Teach* 41, 106–114.
- Samsar K, Knight JK, Gülnur B, Smith MK (2011). The Colorado Learning About Science Attitude Survey (CLASS) for use in biology. *CBE Life Sci Educ* 10, 268–278.
- Schwartz MS, Sadler PM, Sonnet G, Tai RH (2009). Depth versus breadth: how content coverage in high school science courses relates to later success in college science coursework. *Sci Educ* 93, 798–826.
- Seethaler S (2005). Helping students make links through science controversy. *Am Biol Teach* 67, 265–274.
- Seymour E, Hewitt N (1997). *Talking about Leaving: Why Undergraduates Leave the Sciences*, Boulder, CO: Westview.
- Shi J, Power JM, Klymkowsky MW (2011). Revealing student thinking about experimental design and the roles of control experiments. *Int J Sch Teach Learn* 5, 1–16.
- Sirum K, Humburg J (2011). The Experimental Design Ability Test (EDAT). *Bioscene* 37, 8–16.
- Smith C, Wenk L (2006). The relation among three aspects of college freshmen's epistemology of science. *J Res Sci Teach* 43, 747–785.
- 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). *Critical Thinking Assessment Test, Version 5*, Cookeville, TN: Center for Assessment & Improvement of Learning, Tennessee Tech University.
- Talbot M (2006). The baby lab. *The New Yorker* 82, 90–101.
- Tanner D (2012). Promoting student metacognition. *CBE Life Sci Educ* 11, 113–120.
- Usher DC, Driscoll TA, Dhurjati P, Pelesko JA, Rossi LF, Shleining G, Pusecker K, White HB (2011). A transformative model for undergraduate quantitative biology education. *CBE Life Sci Educ* 9, 181–188.
- Wiegant F, Scager K, Boonstra J (2011). An undergraduate course to bridge the gap between textbooks and scientific research. *CBE Life Sci Educ* 10, 83–94.
- Willingham DT (2007). Critical thinking: why is it so hard to teach? *Am Educ* 31, 8–19.
- Wischusen SM, Wischusen EW (2007). Biology intensive orientation for students (BIOS), a biology “boot camp.” *CBE Life Sci Educ* 6, 172–178.
- Wise SL, DeMars CE (2005). Low examinee effort in low-stakes assessment: problems and potential solutions. *Educ Assess* 10, 1–17.