Redesigning a General Education Science Course to Promote Critical Thinking

Matthew P. Rowe,[*] B. Marcus Gillespie,[§] Kevin R. Harris,[‖] Steven D. Koether,[¶] Li-Jen Y. Shannon,[#] and Lori A. Rose[*]

[*]Department of Biological Sciences, [§]Department of Geography & Geology, [¶]College of Sciences, and [#]Department of Computer Science, Sam Houston State University, Huntsville, TX 77340; [‖]Center for Assessment

graduation (Arum and Roksa, 2014). Science courses, with their focus on evidence and logic, should provide exemplary exposure to and training in critical thinking. Here, too, we appear to be failing, both at the level of individual science classes and programmatically in the science core, given the ineffectiveness of these courses to either improve students' scientific knowledge or mitigate their acceptance of pseudoscientific claims (Walker ., 2002; Johnson and Pigliucci, 2004; Impey ., 2011; Carmel and Yezierski, 2013).

The inadequacy of standard approaches to teaching science is demonstrated by the fact that 93% of American adults and 78% of those with college degrees are scientifically illiterate (Hazen, 2002); that is, they do not understand science as an empirically based method of inquiry, they lack knowledge of fundamental scientific facts, and they are unable to understand the science-related material published in a newspaper such as the (Miller, 1998, 2012). Such deficiencies extend to science majors as well. For example, a study of 170 undergraduates at the University of Tennessee found that, while science majors knew more science facts than non–science majors, there were no differences between the two groups in their conceptual understanding of science or their belief in pseudoscience (Johnson and Pigliucci, 2004). This poor understanding of science adversely affects the ability of individuals to make informed decisions about science-related issues, including well-established theories like the big bang, which is rejected by nearly two-thirds of Americans (National Science Foundation, 2014). The woeful lack of scientific literacy similarly provides insight into the public (though not scientific) controversies surrounding such issues as evolution (Miller ., 2006), global climate change (Morrison, 2011; Reardon, 2011), and the safety of childhood immunizations (Mnookin, 2011; Offit, 2011). In short, there appears to be a gap between a fundamental goal of science education, to produce scientifically literate citizens, and the results of the pedagogical approaches intended to meet this goal. Particularly troublesome is the ripple effect of inadequate science education at the university level, leading to poor teacher preparation and threatening the quality of science instruction in our public schools (Eve and Dunn, 1990; Rutledge and Warden, 2000).

Commonly identified causes of the impotency of science courses, especially the introductory courses taken by the majority of college students, are their tendency to focus on scientific "facts" rather than on the nature of science (Johnson and Pigliucci, 2004; Alberts, 2005), often reinforced by exams that reward memorization over higher-order thinking (Alberts, 2009; Momsen ., 2010); the reluctance to directly engage students' misconceptions (Alters and Nelson, 2002; Nelson, 2008; Alberts, 2005; Verhey, 2005); the failure to connect "science as a way of knowing" with decisions faced by students in their daily lives (Kuhn, 1993; Walker ., 2002); and the resistance of faculty trained in more innovative pedagogical approaches to actually employ them (Ebert-May ., 2011). The traditional approach to science education not only fosters scientific illiteracy, but also alienates many students from science (Seymour and Hewitt, 1997; Ede, 2000; Johnson, 2007) and, ultimately, jeopardizes America's global competitiveness (National Academy of Sciences, National Academy of Engineering, and Institute of Medicine, 2010). While methods emphasizing active learning demonstrate significant pedagogical improvements for students majoring in the sciences (Freeman ., 2014), ~85% of the 1.8 million students graduating from college annually in the United States are not science majors (Snyder and Dillow, 2013). Our goal, therefore, was to develop and test an intervention targeting this larger, frequently overlooked, yet extremely important audience. But what would scientific literacy comprise for students completing only one or two science courses during their college careers? What tools could we use to measure said literacy? And how might we best, in a single course or two, help our students achieve it?

Our answer to these questions was an integrative, general education (gen ed) science course titled Foundations of Science (FoS), selected as the centerpiece of the Quality Enhancement Plan for reaffirmation at Sam Houston State University (SHSU; Sam Houston State University, 2009). Per Sagan's (1996) admonition, the FoS course focuses as much on the nature of science as on its facts. We intentionally sought to demystify the process of science by selecting examples, such as the vaccine-autism controversy, that not only held the students' attention but also, and as importantly, helped demonstrate the utility of "evidentiary thinking" in their daily lives. A brief list of the central tenets of the course is provided below; more detail is available in the "Expanded Course Rationale and Structure" in our Supplemental Material.

Critical Thinking

Our central hypothesis was that critical thinking—defined as the ability to draw reasonable conclusions based on evidence, logic, and intellectual honesty—is inherent to scientific reasoning (Facione, 1990, 2015; American Association for the Advancement of Science [AAAS], 1993; Bernstein ., 2006) and is therefore an essential aspect of scientific literacy. Scientific literacy, then, can best be achieved by offering an alternative type of integrated science course that focuses on these foundations rather than on the traditional "memorize the facts" approach to science education. A simple, operational approach to critical thinking is provided by Bernstein . (2006) via a set of questions one should ask when presented with a claim (e.g., vaccines cause autism, global warming is a hoax, there are no transitional fossils). 1) What am I being asked to accept? 2) What evidence supports the claim? 3) Are there alternative explanations/hypotheses? And, finally, 4) what evidence supports the alternatives? The most likely explanation is the one that is best supported. Evidence matters, but only when of the evidence for and against of the competing hypotheses has been examined—fully, thoughtfully, and honestly. Sounds like science, doesn't it? But how can we get science-phobic college students to use it? Perhaps by focusing on topics the non–science student finds interesting, including astrology, homeopathy, Bigfoot, and even intelligent design. But aren't these ideas just pseudoscientific nonsense? Of course, but students need to understand why they are pseudo rather than real science, and critical thinking/scientific literacy is the key. This is the approach adopted by Theodore Schick and Lewis Vaughn (2014) in . *A* , one of the two main texts we adopt in the course.

This text and the course also help students identify and analyze the validity and soundness of arguments. We include a discussion of common heuristics and several logical fallacies, some examples being correlation proves causation, appeal to the masses, and ad hominem attacks. An understanding and awareness of strong versus weak arguments, and the informal fallacies used to surreptitiously circumvent the former, are essential to critical thinking and to the evaluation of claims—whether scientific or pseudoscientific.

Integrating Content with Process

While there has been a clarion call for teachers to focus more on scientific process and less on scientific facts (Rutherford and Ahlgren, 1990; AAAS, 1993, 2010), content still matters. Therefore, in addition to the critical-thinking text by Schick and Vaughn, we also use an integrated science textbook (e.g., Hewitt ., 2013; Trefil and Hazen, 2013) as our second text, typically a custom printing that includes only those chapters whose content we cover in the course. We are fortunate that our course includes both “lecture” and “lab” components, providing multiple, weekly opportunities for active learning. We employ, as a cornerstone of our approach, case studies we have built specifically for the FoS course. Cases, we have found, permit us to teach content and process at the same time, in a manner that engages the non–science student. One of our cases, for example, examines the purported connection between vaccines and autism (Rowe, 2010). Working in small groups, students examine the data from Andrew Wakefield .’s (1998) paper, the proverbial match that lit the current firestorm of antivaccine hysteria (Mnookin, 2011; Offit, 2011). After dissecting Wakefield’s data and his conclusions, students are tasked with designing a better study. In so doing, they learn a great deal about sample size, replication, double-blind studies, and scientific honesty, that is, the procedural underpinnings of good science. But the students also learn about antibodies, antigens, herd immunity, and autism spectrum disorders, that is, the findings of science. Similarly, in a case in which students use the science of ecology to go “hunting” for the Loch Ness monster (Rowe, 2015), they must learn and then apply scientific “findings” ranging from the second law of thermodynamics to minimum viable population sizes to postglacial rebound. A large part of the success we witness in our experimental course is due, we believe, to this integration of scientific facts with scientific process.

Addressing Cognitive Barriers

An emphasis on evidentiary thinking combined with an integration of content and process will achieve little if students are unable or unwilling to objectively evaluate a claim, hypothesis, or theory. Cognitive barriers can stand in the way of rational decision making (Posner ., 1982; Sinatra ., 2008). We designed the FoS course to overcome two such barriers. One hurdle is peoples’ personal experiences, which, for many, trump critical thinking (Chabris and Simons, 2010). If something feels real, looks real, tastes real, if we saw it, experienced it, then it must be true. Zinc is not effective against the common cold? Why, then, did my headache disappear when I used zinc-infused cough drops? Vaccines do not cause autism? What else could explain why my son stopped walking two days after his MMR shot? To help students understand the limitations of anecdotal evidence, including their own personal experiences, we guide them through an exploration of the science of perception and memory. We use illusions to show how our brain unconsciously takes shortcuts that can lead to misperceptions. And we employ simple exercises to demonstrate the malleability and fallibility of memories. Critical thinking requires we recognize that perceptions and memories may be flawed.

The second barrier starts once perceptions and memories have solidified into an opinion. Opinions, once formed, resist change; the more important the belief, the more stubbornly we hang onto it, even in the light of contradictory evidence (Tavris and Aronson, 2007). An honest evaluation of competing explanations requires that students understand cognitive dissonance and its servant twins, expectation bias and confirmation bias. Facts do not matter to someone who does not want to hear them, and evidence is easily discounted when examined with prejudice. Indeed, simply throwing facts at biased conclusions may cause further retrenchment as, for example, was demonstrated in a recent study (Nyhan ., 2014) of the rebellion against childhood immunizations. Results of the study, which surveyed 1759 parents, are discouraging, in that an intervention presenting the overwhelming evidence that vaccines do not cause autism made parents likely to vaccinate, not more (Nyhan ., 2014).

Social judgment theory (SJT) offers an explanation of Nyhan .’s (2014) counterintuitive results. SJT postulates there is a range, a latitude, of ideas similar to a person’s current position he or she might be willing to consider as being true if presented with information that supports the idea. However, if the idea is too different from the person’s initial belief, if it lies outside his or her latitude of acceptance, it will be rejected (Erwin, 2014). Furthermore, the more involved a person is with a view, the wider the latitudes of rejection and the narrower the latitudes of acceptance (Benoit, n.d.). If we want students to understand and accept the big bang theory and the theory of evolution, ideas many find uncomfortable, we cannot simply present the overwhelming evidence in favor of these ideas, we must also accommodate and overcome the dissonance these explanations engender. SJT was, come Tw thew T*l.odes je.7e id / S4Td[(Socio.7BDCTj-3.35837 -1.16666 Td[(ther)1 and 1990s (Nathan and Snedeker, 2001). As with the earlier hunts, hundreds of people were accused, convicted, and sent to jail, even though there was little or no empirical evidence to support the allegations (Lanning, 1992). Here, too, the students, with little emotional investment and, thus, little dissonance, draw the reasonable conclusion that scientific literacy, evidence, and critical thinking are good things, because they

as international. The average age of the institution's undergraduates is 22 yr. Approximately half of the students are first-generation college students. Because the FoS course is an open-enrollment, gen ed core science course with no prerequisites, the demographic makeup of the course likely represents that of the university. We compared the effectiveness of the FoS course with several traditional introductory science courses for nonmajors taught at the university, courses which, as gen ed survey courses, should also reflect the demographics of the university as a whole.

Experimental Approach

We used a pretest versus posttest design to assess the effectiveness of the FoS. Our treatment group consisted of several sections of the experimental course taught over multiple semesters (Table 1). Our comparison group was composed of several different, traditional gen ed science courses, also sampled over multiple semesters, offered by the departments of chemistry, physics, biology, and geography / geology (Table 1). During the study period of Fall semester 2008 through Fall semester 2012, the average class size in each section of our experimental FoS course was 51.75 (±1.17 SE) students; the lab / discussion sections that accompanied the FoS course were capped at 30 students / section. Over the same period, average class size in the traditional courses that formed our comparison group was 51.00 (± 6.07 SE) students. All of the comparison courses also included a lab, similarly capped at 30 students.

Assessment Tools

To examine changes in student analytical skills, we used the Critical thinking Assessment Test (CAT) developed by the Center for Assessment & Improvement of Learning at Tennessee Tech University (TTU; Stein and Haynes, 2011; Stein ., 2007). The CAT exam assesses several aspects of critical thinking, including the evaluation and interpretation of information, problem solving, creative thinking, and communication. Student skills encompassed by the CAT include their ability to interpret graphs and equations, solve basic math problems, identify logical fallacies, recognize when additional information might be needed to evaluate a claim, understand the limitations of correlational data, and develop alternative explanations for a claim. These aspects of the CAT exam conform to accepted constructs that characterize critical thinking (Facione, 1990, 2015), and align well with those taught in the FOS course, which specifically emphasizes the ability to draw appropriate conclusions based on multiple working hypotheses, evidence, and reason. The CAT instrument consists of 15 questions, most of which are short-answer responses. More than 200 institutions of higher education are now using the CAT for assessing programmatic changes designed to improve critical thinking among college students, permitting us to compare our results not only with traditional gen ed science courses being taught at our own institution but also with national norms.

To examine changes in the attitudes of students about science in general, and controversial scientific theories in particular, we used the Measure of Acceptance of the Theory of Evolution (MATE), a 20-question, Likert-scale survey (Rutledge and Warden, 1999; Rutledge and Sadler, 2007) that has been widely used for assessing the acceptance of evolutionary theory among high school teachers and college

Table 1. CAT scores in traditional versus experimental gen ed science courses, by semester

	Course	Treatment[a]	Term		Design[b]	Incentive[c]	CAT pre score	CAT post score	$_{\text{actual}}$ ( )	Pre–post value	Effect size
1	Introductory geography[d]	T	Fall 2008	36	Post only	None		15.00			
2	Introductory geology[e]	T	Fall 2008	40	Post only	None		15.05			
3	Introductory biology[f]	T	Spring 2009	37	Post only	None		14.66			
4	Introductory geography[d]	T	Spring 2009	39	Post only	None		14.91			
5	Introductory environmental studies[g]	T	Fall 2010	10	Pre and post	EC	17.07	16.90	(9) = 0.232	ns	
6	Introductory physics[h]	T	Fall 2011	16	Pre and post	EC	13.94	14.63	(15) = −0.696	ns	
7	Introductory chemistry[i]	T	Fall 2011	25	Pre and post	EC	13.16	13.68	(24) = −0.586	ns	
8	FoS[j]	E	Fall 2009	53	Pre and post	PoC	16.03	19.77	(52) = −5.385	<0.001	+0.71
9	FoS[j]	E	Spring 2010	53	Pre and post	PoC	17.95	22.43	(52) = −5.872	<0.001	+0.76
10	FoS[j]	E	Fall 2010	47	Pre and post	PoC	15.52	19.98	(46) = −4.848	<0.001	+0.36
11	FoS[j]	E	Spring 2011	69	Pre and post	PoC	14.95	19.60	(68) = −8.999	<0.001	+0.84
12	FoS[j]	E	Fall 2011	25	Pre and post	EC	13.41	17.75	(24) = −3.984	<0.001	+0.85
13	FoS[j]	E	Fall 2012	25	Pre and post	EC	12.25	16.16	(24) = −3.310	<0.01	+0.83

[a] T =

students (Moore and Cotner, 2009; Nadelson and Southerland, 2010; Peker ., 2010; Kim and Nehm, 2011; Abraham ., 2012).

Beginning in the Fall of 2010, approximately half the students in each of the experimental and comparison courses were assessed pre- and postcourse using the CAT, the other half with the MATE. The pretests were administered during the second week of the term, while the posttests were given in the penultimate week of classes. Instructors teaching both the FoS and the traditional courses agreed on identical incentives each semester, with the exception of Fall 2010: as no credit (baseline data before creation of the FoS) or as extra credit / part of the course grade thereafter (Table 1). Details regarding how the incentive was applied are provided in the example course syllabus in our Supplemental Materials.

All CAT exams were graded using a modified rubric that enabled the exams to be graded quickly. These scores were used to assign performance points to the students. A subset of all the CAT exams from each course was randomly selected for formal grading using the rubric developed by the Center for Assessment & Improvement of Learning at TTU. Based on the grading procedures established by the center, graders were blind to the identity of the student, whether an exam was a pretest or posttest, and the treatment group. Results of the formal grading are reported herein.

The MATE was coupled with a locally developed assessment not presented in this publication. Because the responses on the MATE assessment represent personal opinions and attitudes, no incentives were provided to students for their responses on the MATE, and they were informed that their answers would not be graded. However, students were still able to earn rewards equivalent to those of students taking the CAT based on their performance on the locally developed assessment tool.

Assessment Reliability and Validity

Arguments regarding the effectiveness of the FoS course demand both reliability and validity. While these concepts are frequently ignored (Campbell and Nehm, 2013), researchers who address the issues of reliability and validity often mistake them as required properties of one's assessment tools rather than, correctly, as characteristics of the interpretations we make from the tools' results (Cronbach and Meehl, 1955; Messick, 1995; Brown, 2005; Campbell and Nehm, 2013). The reliability and validity of interpretations based on the CAT have strong evidentiary support (Tennessee Technological University, 2010; Stein and Haynes, 2011; Stein ., 2007, 2010).

Interpretations based on the MATE also have demonstrated reliability and validity, at least for certain populations (Rutledge and Warden, 1999; Rutledge and Sadler, 2007). A recent study (Wagler and Wagler, 2013), however, found the MATE lacked construct validity for Hispanic elementary education majors and questioned the utility of the tool for assessing student acceptance of evolutionary theory. Our results do not support this criticism, an argument we present more fully in our .

FoS course. Similarly, we have posttest MATE scores from 1250 undergraduates, with 417 representing the three traditional courses and 833 from the five semesters of the FoS course.

RESULTS

Critical Thinking

FoS Experiment versus Traditional Gen Ed Science Courses. Our results are robust and consistent; quite simply, students who complete the experimental FoS course show significant improvement in their critical-thinking skills, as measured by the CAT, while students who complete a traditional gen ed science course do not. In no semester, for example, did students completing a traditional course show improvement in their critical-thinking scores (all values > 0.49; Table 1), while students completing the experimental course showed highly significant improvement each semester (all values < 0.01, Cohen's typically > 0.70; Table 1). An analysis of pooled end-of-course (posttest only) CAT scores for all six semesters of the FoS course (Table 1, rows 8–13) versus the pooled posttest CAT scores for all six traditional gen ed science courses (Table 1, rows 1–7) reinforce this finding; students completing the FoS course scored significantly higher (19.76 ± 0.35) than did students completing a traditional (14.83 ± 0.37) introductory science course for nonmajors ($(473) = 4.93$, < 0.001, Cohen's $= 0.89$; Figure 1A). A comparison of our pooled pre- versus posttest CAT scores for all six semesters of the FoS course (Table 1, rows 8–13) versus the pooled CAT scores for the three different gen ed science courses (introductory environmental studies, introductory physics, and introductory chemistry) for which we had pre- and postcourse CAT test scores (Table 1, rows 5–7) show similar results. Students who completed the FoS course showed highly significant improvement in critical thinking (pretest = 15.45 ± 0.34, posttest = 19.76 ± 35; $(271) = 13.43$, < 0.001, Cohen's $= 0.76$), while there was no change in the critical thinking scores for students completing a traditional course (pretest = 14.17 ± 0.64, posttest = 14.61 ± 0.72; $(50) = 0.80$, $= 0.43$; Figure 1B).

The slightly higher pretest CAT scores for students in the experimental course relative to students taking a traditional course (15.45 vs. 14.61, respectively, Figure 1B) might suggest the significant pre versus post improvement in the former represents a cohort rather than a treatment effect; that is, stu

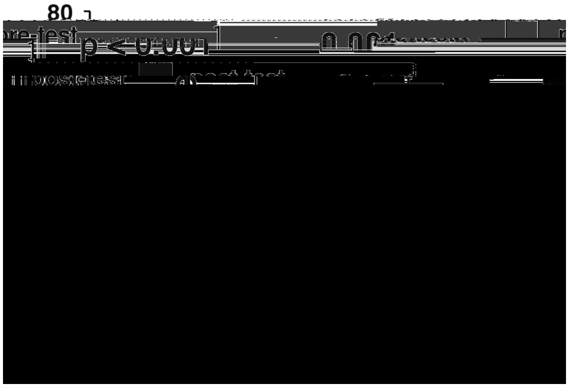

Figure 3. Students who complete the experimental FoS course show a significant increase in their acceptance of evolution, as measured by the MATE, while students who complete a traditional gen ed science course do not. Pooled pre- vs. posttest MATE scores for five semesters of the FoS course (Table 2, rows 4–8) vs. the pooled MATE scores for the three different gen ed science courses (introductory environmental studies, introductory physics, and introductory chemistry) for which we had pre- and postcourse MATE scores (Table 2, rows 1–3). Histograms show means + 1 SE.

of evolution (pretest = 66.17 ± 0.45, posttest = 75.45 ± 0.49; (1686.15) = 13.93, < 0.001, Cohen's = 0.67), while there was no change in the acceptance of evolution for students completing a traditional course (pretest = 65.27 ± 0.56, posttest = 64.91 ± 0.71; (976) = 0.40, = 0.69; Figure 3).

DISCUSSION

Critical Thinking

Our results demonstrate that an introductory, gen ed science course for nonmajors, a course focusing on the nature of science rather than just its facts, can lead to highly significant improvements, with large effect sizes, in the ability of college students to think critically. Most college courses do not significantly improve CAT performance in a pre/post design; substantive gains are typically observed only at the program/institutional level (Center for Assessment & Improvement of Learning, TTU, unpublished data). Moreover, results from more than 200 institutions using the CAT show the average improvement in critical thinking observed over 4 yr of a typical undergraduate curriculum is 26% (Harris ., 2014); students who successfully completed the FoS course improved their CAT scores by almost 28% (15.45 vs. 19.76; Figure 1B). In short, students who complete a single-semester FoS course demonstrate levels of improvement in their critical-thinking skills typically requiring multiple years of college experience, demonstrating that it is possible to teach higher-order thinking skills to nonmajors in a single science course they are required to take, many begrudgingly.

A finer-grained analysis of our results further illustrates the need to rethink how we are teaching our gen ed science courses. The pretest CAT score for our lower-division students, pooled over all six semesters, was significantly higher than the national average for this age group (Figure 2A). By the end of the semester, our lower-division students' critical-thinking scores moved well beyond the national norm for freshmen/sophomores and were comparable to the CAT scores achieved by juniors and seniors nationwide (Figure 2A). This is the good news.

The pattern for our upper-division students, however, is more worrisome, as their pretest CAT average is significantly lower than the national mean for juniors and seniors (Figure 2B). Given that our lower-division students start with significantly better CAT scores than their peers nationally, results showing that our juniors and seniors are significantly worse (before taking the FoS course) than their countrywide counterparts might suggest our institutional curriculum degrades rather than improves a student's critical-thinking skills. An alternative interpretation is that the non–science comparable studies suggest we have much to learn about the factors influencing student acceptance of evolutionary theory. To contribute, we plan additional analyses, mining our database to examine the effects of gender, ethnicity, high school grade point average, and student attitudes on the MATE and on the CAT.

Instructors (who are also colleagues and friends) in the traditional gen ed science courses that served as our comparison group were disappointed their students showed no improvement in critical thinking after a semester of science. But, they argued reasonably, why should we expect student acceptance of evolutionary theory to improve in introductory gen ed chemistry or physics classes, given that biological evolution is not discussed in such courses? Four points are relevant, the last being most important. First, we suggest that all college graduates, science majors or not, should appreciate how the term "theory," used scientifically, differs from its conversational definition. Second, evolutionary theory was covered in the environmental studies course (Table 2) in which we used the MATE, yet students still failed to demonstrate improvement in their acceptance of the theory in this traditionally taught gen ed science course. Third, even though evolution is a topic we address explicitly in the FoS course, it is covered during the last week of the semester, the week following the posttest administration of the MATE.

The most important issue, however, relates to what the MATE may be measuring. Several authors have argued that the MATE more likely measures an individual's knowledge about evolution rather than his or her acceptance of the theory (Smith, 2010a; Wagler and Wagler, 2013). And while it is generally presumed that some content knowledge is required for a student to accept evolution as the best explanation of biological diversity, evidence also suggests that dispositional change may be required before a student is willing to entertain the theory (Sinatra ., 2003; Smith, 2010a,b). Whether the MATE measures an individual's content knowledge about evolution or his or her disposition toward the theory is beyond the scope of this analysis. Our results, however, are robust; a course focusing on the nature of science and applying SJT leads to significantly improved engagement of the non–science college student with evolution (see also Pigliucci, 2007; Lombrozo ., 2008).

Assessment Validity, Revisited

Wagler and Wagler (2013) criticized the construct validity and, thus, the generalizability of the MATE for populations other than the high school teachers used to originally test the tool's validity (Rutledge and Warden, 1999). The Waglers found, for example, that the MATE lacked construct validity for their sample of Hispanic college students majoring in elementary education. Construct validity is the degree to which a test actually measures the mental attribute it claims to measure (Brown, 2000); for the MATE, the attribute is thought to be an individual's acceptance of the theory of evolution (Rutledge and Warden, 1999). One technique for assessing construct validity uses factor analyses with structural equation modeling to identify the number of dimensions of the construct; if a significant unifying dimension or dimensions cannot be identified, the tool may be suspect; this was the approach used to demonstrate that

the MATE lacked construct validity for preservice teachers (Wagler and Wagler, 2013). We applied the same technique to our MATE results and similarly found that no model, either uni- or multidimensional, could be fitted to the data (unpublished data). But researchers should never rely on a single method for assessing the validity of their interpretations (Cronbach and Meehl, 1955; Messick, 1995; Brown, 2000, 2005; Campbell and Nehm, 2013). Two related experimental approaches for assessing the construct validity of a test are intervention studies and differential-groups studies (Cronbach and Meehl, 1955; Messick, 1995; Brown, 2000, 2005). In the former, a group is tested before and following their exposure to the construct; significant improvement demonstrates the construct validity of the intervention. Differential-groups studies employ two groups, one presented with the construct, the other not; significantly better scores by the informed group similarly demonstrate the validity of the training. We used both approaches in this study; the "construct" was a novel gen ed science course (the FoS) focusing on the nature of science rather than just its facts (for more details please see "Expanded Course Rationale and Structure" in our Supplemental Materials). Students who completed the training demonstrated, over multiple sections of the course spanning multiple years, highly significant improvement both in their critical-thinking skills (as measured by the CAT; Table 1 and associated figures) and in their willingness to engage the theory of evolution (assessed with the MATE; Table 2 and associated figures).
(aS - /T1_2as Tfigripp, forrhavt7 T6aug(tutevatal, oct 0s see TD-p1Es;-ional76).)7 (

Cawthorne N (2004). Witch Hunt: History of a Persecution, London: Chartwell.

Chabris C, Simons D (2010). The Invisible Gorilla, New York: Crown.

Cronbach LJ, Meehl PE (1955). Construct validity in psychological tests. Psych Bull *52*, 281–302.

Ebert-May D, Derting TL, Hodder J, Momsen JL, Long TM, Jardeleza SE (2011). What we say is not what we do: effective evaluation of

Lombrozo T, Thanukos A, Weisberg M (2008). The importance of understanding the nature of science for accepting evolution. Evol Educ Outreach *1*, 290–298.

Maher JM, Markey JC, Ebert-May D (2013). The other half of the story: effect size analysis in quantitative research. CBE Life Sci Educ *12*, 345–351.

Messick S (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. Am Psychol *50*, 741–749.

Miller JD (1998). The measurement of civic scientific literacy. Public Underst Sci *7*, 203–223.

Miller JD (2012). What colleges and universities need to do to advance civic scientific literacy and preserve American democracy. Liberal Education . www.aacu.org/publications-research/periodicals/what-colleges-and-universities-need-do-advance-civic-scientific (accessed 7 February 2015).

Miller JD, Scott EC, Okamoto S (2006). Public acceptance of evolution. Science *313*, 765–766.

Mnookin S (2011). The Panic Virus: A True Story of Medicine, Science, and Fear, New York: Simon & Schuster.

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 , 435–440.

Moore R, Cotner S (2009). Educational malpractice: the impact of including creationism in high school biology courses. Evol Educ Outreach *2*, 95–100.

Morrison D (2011). Science denialism: evolution and climate change. Reports Natl Center Sci Educ *31*, 1–10.

Nadelson LS, Southerland SA (2010). Examining the interaction of acceptance and understanding: how does the relationship change with a focus on macroevolution? Evol Educ Outreach *3*, 82–88.

Nathan D, Snedeker M (2001). Satan's Silence: Ritual Abuse and the Making of a Modern American Witch Hunt, Lincoln, NE: Author's Choice Press.

National Academy of Sciences, National Academy of Engineering, and Institute of Medicine (2010). Rising above the Gathering Storm, Revisited: Rapidly Approaching Category 5, Washington, DC: National Academies Press.

National Science Foundation (2014). Science and Engineering Indicators 2014, Arlington, VA: National Science Board.

Nelson CE (2008). Teaching evolution (and all of biology) more effectively: strategies for engagement, critical reasoning, and confronting misconceptions. Integr Comp Biol *4* , 213–225.

Nyhan B, Reifler J, Richey S, Freed GL (2014). Effective messages in vaccine promotion: a randomized trial. Pediatrics *133*, E835–E842.

Offit PA (2011). Deadly Choices: How the Anti-Vaccine Movement Threatens Us All, New York: Basic.

Pascarella ET, Blaich C, Martin GL, Hanson JM (2011). How robust are the findings of academically adrift? Change *43*, 20–24.

Peker D, Comert G, Kence A (2010). Three decades of anti-evolution campaign and its results: Turkish undergraduates' acceptance and understanding of the biological evolution theory. Sci Educ *1* , 739–755.

Pigliucci M (2007). The evolution-creation wars: why teaching more science just is not enough. McGill J Educ *42*, 285–306.

Posner GJ, Strike KA, Hewson PW, Gertzog WA (1982). Accommodation of a scientific conception: toward a theory of conceptual change. Sci Educ *66*, 211–227.

Reardon S (2011). Climate change sparks battles in classroom. Science *333*, 688–689.

Rowe MP (2010). Tragic choices: autism, measles, and the MMR vaccine. In: Science Stories: Using Case Studies to Teach Critical

Thinking, ed. CF Herreid, NA Schiller, and KF Herreid, Arlington, VA: NSTA Press. http://sciencecases.lib.buffalo.edu/cs/collection/detail.asp?case_id = 576&id = 576 (accessed 7 February 2015).

Rowe MP (2015). Crazy about cryptids! An ecological hunt for Nessie and other legendary creatures. National Center for Case Study Teaching in Science. http://sciencecases.lib.buffalo.edu/cs/collection/detail.asp?case_id=779&id=779 (accessed 26 June 2015).

Rutherford FJ, Ahlgren A (1990). Science for All Americans, Oxford, UK: Oxford University Press.

Rutledge ML, Sadler KC (2007). Reliability of the measure of acceptance of the theory of evolution (MATE) instrument with university students. Am Biol Teach *6* , 332–335.

Rutledge ML, Warden MA (1999). The development and validation of the measure of acceptance of the theory of evolution instrument. Sch Sci Math , 13–18.

Rutledge ML, Warden MA (2000). Evolutionary theory, the nature of science and high school biology teachers: critical relationships. Am Biol Teach *62*, 23–31.

Sagan C (1996). The Demon-haunted World: Science as a Candle in the Dark, New York: Ballantine.

Sam Houston State University (2009). QEP: SHSU: Foundations of Science. www.shsu.edu/qep/documents/QualityEnhancementPlanCombined.pdf (accessed 7 February 2015).

Schick T, Vaughn L (2014). How to Think about Weird Things: Critical Thinking for a New Age, New York: McGraw-Hill.

Seymour E, Hewitt NM (1997). Talking about Leaving: Why Undergraduates Leave the Sciences, Boulder, CO: Westview.

Sinatra G, Brem S, Evans EM (2008). Changing minds? Implications of conceptual change for teaching and learning about biological evolution. Evol Educ Outreach *1*, 189–195.

Sinatra GM, Southerland SA, McConaughy F, Demastes JW (2003). Intentions and beliefs in students' understanding and acceptance of biological evolution. J Res Sci Teach *40*, 510–528.

Smith MU (2010a). Current status of research in teaching and learning evolution: I. Philosophical/epistemological issues. Sci Educ *1* , 523–538.

Smith MU (2010b). Current status of research in teaching and learning evolution: II. Pedagogical issues. Sci Educ *1* , 539–571.

Snyder TD, Dillow SA (2013). Digest of Education Statistics, 2012, Washington, DC: National Center for Education Statistics.

Stein B, Haynes A (2011). Engaging faculty in the assessment and improvement of students' critical thinking using the critical thinking assessment test. Change *43*, 44–49.

Stein B, Haynes A, Redding M, Ennis T, Cecil M (2007). Assessing critical thinking in STEM and beyond. In: Innovations in E-Learning, Instruction Technology, Assessment, and Engineering Education, ed. M Iskander, Dordrecht, Netherlands: Springer.

Stein B, Haynes A, Redding M, Harris K, Tylka M, Lisic E (2010). Faculty driven assessment of critical thinking: national dissemination of the CAT instrument. In: Technological Developments in Networking, Education and Automation, ed. K Elleithy, T Sobh, M Iskander, V Kapila, MA Karim, and A Mahmood, Dordrecht, Netherlands: Springer.

Tavris C, Aronson E (2007). Mistakes Were Made (but Not by Me), Orlando, FL: Harcourt.

Tennessee Technological University (2010). CAT Instrument Technical Information. www.tntech.edu/files/cat/reports/CAT_Technical_Information_V7.pdf (accessed 7 February 2015).

Trefil J, Hazen RM (2013). The Sciences: An Integrated Approach, 7th ed., Hoboken, NJ: Wiley.

Van Gelder T (2005). Teaching critical thinking. Coll Teach *45*, 41–46.

Verhey SD (2005). The effect of engaging prior learning on student attitudes toward creationism and evolution. BioScience *55*, 996–1003.

Wagler A, Wagler R (2013). Addressing the lack of measurement invariance for the measure of acceptance of the theory of evolution. Int J Sci Educ *35*, 2278–2298.

Wakefield A, Murch SH, Anthony A, Linnell J, Casson DM (1998). Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. Lancet *351*, 637–641.

Walker WR, Hoekstra SJ, Vogl RJ (2002). Science education is no guarantee of skepticism. Skeptic , 24–27.