Scientific Citizenship

Using a Concept Inventory to Assess the Reasoning Component of Citizen-Level Science Literacy: Results from a 17,000-Student Study †

Edward B. Nuhfer[1]*, Christopher B. Cogan[2], Carl Kloock[3], Gregory G. Wood[4], Anya Goodman[5], Natalie Zayas Delgado[6], and Christopher W. Wheeler[7]

[1]Retired Professor of Geology and Director of Faculty Development & Assessment, Niwot, CO 80503,
[2]Independent Consultant, Environmental Science, Geography, and Geographic Information Systems, Camarillo, CA 93012, [3]Department of Biology, California State University Bakersfield, Bakersfield, CA 93311, [4]Department of Physics, California State University Channel Islands, Camarillo CA 93012, [5]Department of Chemistry and Biochemistry, California Polytechnic State University, San Luis Obispo, CA 93407, [6]Department of Science & Environmental Policy, California State University Monterey Bay, Seaside, CA 93955, [7]Geology Program, California State University Channel Islands, Camarillo, CA 93012

After articulating 12 concepts for the reasoning component of citizen-level science literacy and restating these as assessable student learning outcomes (SLOs), we developed a valid and reliable assessment instrument for addressing the outcomes with a brief 25-item science literacy concept inventory (SLCI). In this paper, we report the results that we obtained from assessing the citizen-level science literacy of 17,382 undergraduate students, 149 graduate students, and 181 professors. We address only findings at or above the 99.9% confidence level. We found that general education (GE) science courses do not significantly advance understanding of science as a way of knowing. However, the understanding of science's way of knowing does increase through academic ranks, indicating that the extended overall academic experience better accounts for increasing such thinking capacity than do science courses alone. Higher mean institutional SLCI scores correlate closely with increased institutional selectivity, as measured by the institutions' higher mean SAT and ACT scores. Socioeconomic factors of a) first-generation student, b) English as a native language, and c) interest in commitment to a science major are unequally distributed across ethnic groups. These factors proved powerful in accounting for the variations in SLCI scores across ethnicities and genders.

INTRODUCTION

Nearly all colleges and universities require one or more science courses as part of general education (GE). Students can meet the GE science requirement by selecting from courses that provide different disciplinary content and skills. This implies that the disparate courses hold a construct in common. Meeting the GE requirement rests on achieving an understanding of that unifying construct rather than on simply learning the content of particular science disciplines. To articulate the unique contribution that science courses provide requires recognizing that science explains the physical world through testing (51). Other metadisciplines—mathematics, the arts, the humanities, and social sciences—that are part of GE also contribute to the development of critical thinking. However, no other way of knowing has proven as successful as science in producing the understanding of physical phenomena.

How much "understanding" suffices to meet a GE requirement? The level of operational ability that GE curricula seek to impart is "citizen literacy." Citizen literacy achieves competency sufficient for making decisions and taking informed actions in diverse personal and civic challenges encountered in everyday life (35). Further, a college education should be able to demonstrate that it increases adult competencies beyond those which the students acquired in K–12 experiences. Gaining the ability to engage across the breadth of life's challenges differs from gaining the competencies of specialization produced by academic majors, which prepare a graduate to enter a career within her or his major field of specialization. When instructors teach only disciplinary content and skills, the enacted

course goals focus on training majors and differ from the stated GE goals of educating for critical thinking or citizen literacy (6).

Most university catalogs convey that general education requirements exist to provide graduates with an improved capacity for thinking, often expressed as "critical thinking." Science's evidence-based reasoning constitutes one of the five major traditions of critical thinking (4). Adult intellectual development (22, 23, 38) and "critical thinking" are overlapping constructs that are not identical (27), and the ability to employ evidence occupies most of the overlap. By immersing students in a study of verifiable, evidence-based understanding of the physical world, general education science courses can provide a substantial contribution to developing students' elevated thinking skills.

Engaging students in evaluative problems that require the use of evidence for resolution seems to promote intellectual development (45). While research confirms that such problems presented across several courses produce measurable gains (37), it remains uncertain whether one or two GE science courses can do so, at least as prevalently taught. The ability to reflect metacognitively on science's way of knowing should be helpful to understanding the use of evidence. In a review of many GE science textbooks, the authors discovered that presentations of science's way of knowing were sketchy and usually confined to pages early in the text. None used the discipline's content throughout the textbook to reinforce an understanding of science's way of explaining the physical world.

Assessing gains in elevated thinking is more difficult than assessing gains in content knowledge. Research reveals that increases in college students' capacity for reasoning occur over several years (22, 38). Small changes produced at the scale of a term- or semester-long GE course can be imperceptible to students and instructors alike. Measuring small changes will require reliable, thoroughly vetted instruments.

In this paper, we address three questions:

1. Can we isolate and describe the major concepts that constitute GE citizen-level science literacy in understanding science's way of knowing?
2. Can we reliably assess the construct of such science literacy through addressing specific concepts on a simple concept inventory?
3. Can the resulting data yield information of value?

To address Question 1, we distilled a vast literature and the combined experiences of our team of nine (36) into 13 concepts. We express 12 of these concepts as equivalent student learning outcomes (SLOs) in Table 1. To address Question 2, we constructed a 25-item science literacy concept inventory (SLCI) to assess the 12 concepts and documented its validity and reliability. To address Question 3, we summarized the results of using this instrument to assess such literacy with 17,382 undergraduate students plus 149 graduate students and 181 professors.

METHODS

This research was reviewed, classified as exempt, and conducted with institutional review board (IRB) oversight and full disclosure to all participants through IRB-105122 from 2010–2013 at CSU Channel Islands and IRB-13-019 from 2013–2016 at Humboldt State University to comply with all relevant federal guidelines and policies.

Articulating concepts of reasoning for citizen science literacy

From an extensive literature review (including 1, 2, 3, 7, 12, 13, 15, 19, 20, 24, 28–30, 33, 40, 41, 43, 44, 46, 47, 51) and our combined multidisciplinary experiences, we developed a 25-item concept inventory, mapping to 12 learning outcomes (Table 1). The creators of the SLCI represent four California State University campuses and five science disciplines (biology, chemistry, environmental science, geoscience, and physics).

Table 1 serves as a resource for designing lessons that enable students to reflect metacognitively (8, 9) on the supporting concepts of science's way of knowing as they learn content throughout a course. For example, in each text reading assignment, an instructor might pick any figure in the assignment and ask students to reflect on the dominant method through which the knowledge was acquired to yield the illustration (concepts, 3, 4, 5 or a combination). Repeated reflections should eventually develop better mastery.

Relating SLCI concepts to prevalent definitions of science literacy

Developing global definitions of science and science literacy requires consideration of the concepts that support the general construct of science. Some authors of relevant definitions also developed instruments to assess understanding of the construct. In this section, we address the relevance of the 12 concepts of Table 1 to some influential definitions of science.

The SLCI concepts of Table 1 touch on all three of Miller's 1983 three-component definition of science literacy (29): 1) understanding of the norms and methods of science (i.e., the nature of science), 2) understanding of key scientific terms and concepts, and 3) understanding of the impact of science and technology on society. Here, we consider "key terms and concepts" as metadisciplinary and not to be confused with disciplinary jargon. Miller (30) later developed one of the most widely cited assessment instruments to address the construct defined by these three components.

Strahler (43) captures our most essential concepts with: "Science is the acquisition of reliable but not infallible knowledge of the real world, including explanations of the phenomena." The National Academy of Sciences definition of science (31) as "the use of evidence to construct testable explanation and prediction of natural phenomena, as well as

TABLE 1.

Concepts and equivalent student learning outcomes important to understanding science as a way of knowing and explaining the physical world.

Concepts for Citizen Literacy in the Metadiscipline of Science	Equivalent Student Learning Outcomes for Science Literacy: "Students should be able to…"
1. Science explains physical phenomena based upon testable information about the physical world.	1. Define the domain of science and determine whether a statement constitutes a hypothesis that can be resolved within that domain.
2. A theory in science is a unifying explanation for observations that result from testing several hypotheses.	2. Explain how "theory" as used and understood in science differs from "theory" as used and understood by the general public.
3. Science can test certain kinds of hypotheses through controlled experiments.	3. Explain how science employs the method of reproducible experiments to understand and explain the physical world.
4. Scientists use evidence-based reasoning to select which among several competing working hypotheses best explains a physical phenomenon.	4. Explain how scientists select which among several competing working hypotheses best explains a physical phenomenon.
5. Science employs modeling as a method for understanding the physical world.	5. Explain and provide an example of how science employs modeling as a way of acquiring testable knowledge.
6. Scientific knowledge is discovered, and some discoveries require an important history.	6. Cite a single major theory from one of the science disciplines and explain its historical development.
7. Doubt plays a necessary role in advancing science.	7. Explain why the attribute of doubt has value in science.
8. All science rests on fundamental assumptions about the physical world.	8. Articulate how science's way of knowing rests on some assumptions.
9. Science differs from technology.	9. Distinguish between science and technology by examples of how these differ as frameworks of reasoning.
10. Peer review generally leads to better understanding of physical phenomena than can the unquestioned conclusions of involved investigators.	10. Explain why peer review generally improves our quality of knowing within science.
11. In modern life, science literacy is important to both personal and collective decisions that involve science content and reasoning.	11. Describe through examples how science literacy is important in everyday life to an educated person.
12. Scientific knowledge imparts power that must be used ethically.	12. Explain why ethical decision-making becomes increasingly important to a society becoming increasingly advanced in science.
13. A student can meet the minimal learning outcomes specified by the discipline for the GE course that address the content and skills of the science discipline.	

See 34 for a tabulation of technology's equivalents to the 12 concepts of science in this table. GE = general education.

the knowledge generated through this process" is similar to Strahler's. However, it does not stress the imperfect nature of science's body of knowledge.

Strahler's "real world" includes both living and non-living material and thus the realms of both the biological and physical sciences. This helps to distinguish science from what it is not by recognizing the realm of science as that which explains the physical world through testable information (concept 1). Implicit within this concept is science's meaning of "hypothesis" as a testable statement about the physical world. Accumulated evidence may eventually enable a theory (concept 2) that provides a unified explanation for several hypotheses. These concepts fall within Miller's first and second components.

Strahler's "acquisition of … knowledge" encompasses science's methods of knowing through testing of the physical world. This concept also falls within Miller's components 1 and 2. Prevalent methods of testing to discover new knowledge include a) controlled experiments (concept 3), b) evaluation of multiple working hypotheses (5) (concept 4), and c) modeling (concept 5).

Strahler's "… acquisition of reliable but not infallible knowledge" recognizes science's imperfect knowledge of the physical world at any given time. The phrase elegantly captures the better understanding that results from further discoveries (concept 6). For some phenomena, building understanding requires discoveries that take place over centuries. New hypotheses originate from curiosity or doubt, which lead to testing and improved explanations (concept 7).

Science does rest on some basic assumptions, two being that physical laws govern the physical world and that we can understand them (concept 8). We cannot test

these assumptions directly, but scientists accept them as unproblematic simply because they lack empirical evidence that contradicts them.

We observed that distinguishing science from technology (concept 9) posed a difficult learning challenge. Both address the physical world; both mutually support and advance one another. Conflating the two in the language "science and technology" or the STEM acronym (science, technology, engineering, and mathematics) leads laypersons to perceive the two as equivalent, if not identical. Strahler's phrase "including explanations of the phenomena" addresses the primary attribute for distinguishing science from technology.

The body of knowledge of science advances through an effort to discover and explain phenomena. Technology advances through discovery and invention, but technology does not strive to explain how phenomena operate. For example, much of effective medical practice employs technology. Roman physicians discovered that they could successfully prevent many infections in wounds by binding them with mixtures that contained honey. The Romans had neither knowledge of the role of microbial pathogens in infection nor how honey might affect pathogens, and so they could not explain why honey was effective. Nevertheless, they could benefit by applying the technology without the knowledge required for explaining its effectiveness. Today, explaining how the treatment works is a current research effort in science (39).

In our experience, citizens do not understand that scientists do not invent science's body of knowledge but rather discover it. Likewise, few citizens perceive how striving to explain physical phenomena differs from striving to achieve benefit from them. Eight of the nine designers of the SLCI believed it important that students be able to articulate how science differs from technology. This was the only SLO on which the authors of the SLCI items did not agree unanimously. Wolpert (51) provided some of the strongest arguments for distinguishing between science and technology. He noted that, until the early 20th century, technology's way of thinking was mainly responsible for human progress, with science's way of thinking becoming a dominant contributor only recently. Universities seldom teach distinctions between science and technology. This may account for both experts and novices scoring lowest on items that address this concept.

Concept 10 addresses the role of peer review and applies to many disciplines. In science, it conveys that peer-reviewed primary literature has more consistent reliability than other literature, including textbooks. Concept 11 asserts that if students cannot articulate the value of a GE requirement by the time they have completed the requirement, one of the most valuable learning outcomes remains unmet.

Development of higher-level thinking occurs on both an intellectual and an ethical level (38). Because the practice of science is replete with ethical decisions (concept 12), a GE science course should offer some case examples promoting ethical development.

Concepts 10, 11, and 12 are not specific to science. Strahler's definition does not address these, and many discussions of science literacy omit one or more of these. However, concepts 11 and 12 do fall under component 3 of Miller's definition of science (29). We included the three concepts as components of citizen-level science literacy because science courses contribute additional value to GE through facilitating integration with other metadisciplines.

The 12 concepts/outcomes of Table 1 never appeared together in any single publication that we consulted. Each does appear separately in one or more papers that address science literacy. For producing science literacy, we deem all twelve as important, but we consider the first seven as essential to citizen-level literacy. Our results here reveal that experts outscore novices on all twelve concepts at extremely high levels of confidence.

A 13th concept (Table 1) recognizes the value of providing a direct study of the physical world as a way of achieving an understanding of science's way of knowing. Because disciplinary content is the usual vehicle for teaching the outcomes, mastery of relevant disciplinary content is a legitimate expectation for GE science courses. The SLCI does not address the thirteenth outcome. It is important for a GE course that the thirteenth outcome not be allowed to displace the others.

Other instruments for assessing science literacy

Wenning (48) offers an excellent summary of the historical development of science literacy. Wenning noted the relevance of assessment: "With scientific literacy being the "holy grail" of science education, it would seem reasonable that there should exist some means of assessing progress toward that goal. Unfortunately, such an instrument does not appear to exist." Wenning went on to construct the Nature-of-Science Literacy Test (NOSLiT), which addresses many concepts of the SLCI.

Mathematicians recently developed the Quantitative Literacy and Reasoning Assessment (QLRA) instrument, which is freely available (10, 11) and addresses citizen-level literacy in quantitative reasoning. The Test of Scientific Literacy Skills (TOSLS) is another recent test constructed by biologists as a two-category test of scientific inquiry and quantitative reasoning (14). The TOSLS also contains items that map to some concepts found within the SLCI.

We sought to promote intellectual development by emphasizing the understanding of science's evidence-based way of knowing. Assessing such understanding differs from assessing facts that educated citizens should know (30) or assessing the skills of citizens to solve disciplinary science problems (Classroom Scientific Reasoning Test—CSRT (25)) and TOSLS (14). Both Lawson's CSRT and Wenning's NOSLiT aimed to assess students' science literacy at the pre-college level, but the items they developed are also applicable to some outcomes at the lower-division college level.

The growing number of diverse instruments that attempt to define and assess science literacy show a currently recognized need for these within the academic community. The recent availability of several science literacy instruments is a boon to college instructors. An instructor may now consider the concepts addressed in several instruments, use a separate instrument for assessment, and thereby avoid any stigma of "teaching to the test."

Developing the SLCI

After finding no existing instrument that addressed all 12 concepts, we opted to construct our own instrument to do so. We studied the process of constructing concept inventories (18, 26) before beginning this work.

Writing items that tested reasoning without offering advantages to respondents with high content knowledge offered a special challenge. Our breadth of represented disciplines (biology, chemistry, environmental science, geoscience, and physics) and the fact that we were all largely laypersons in each other's disciplines aided us greatly in meeting that challenge. We used 10 screening criteria (36), and no single item written by any of us emerged unchanged after vetting through our team.

We drafted nearly 80 multiple-choice items that addressed the 12 concepts. Next, we employed a writing specialist to examine all the items and polish the wording and presentation of distracters for clarity. We initially tested the polished items in two 40-item banks on over 1,000 volunteer participants consisting of students from about 20 institutions and faculty from several metadisciplines. Both test banks yielded Cronbach alpha reliabilities of $r = 0.86$ and gave results that were highly consistent with one another. After tabulating the initial pilot test results, we employed an expert in the philosophy of science to review the comprehensiveness and value of our items and our concepts.

We employed the current versions of the SAS Institute's JMP software available between 2010 and 2015. From the initial results, we selected 25 items that addressed all 12 concepts, three of which mapped to concept 1, with two each mapping to the other 11 concepts. We examined our data using classical test theory (CTT) and item response theory (IRT). Since readers are more likely to assess in units of test scores rather than in units of IRT ability formulae, we present our results here as test scores expressed as percentages.

Because we sought to assess a citizen level of science literacy, we collected our information under conditions close to those in which citizens encounter science, which is not in a timed in-class test environment. Citizens encounter science while reading at home or in a local coffee shop, usually online, without blocked access to information. Participants took the SLCI under these same informal conditions. Most students completed the SLCI in about 30 minutes.

Our study engaged students from a wide geographic area of North America, mostly from 20 institutions with diverse missions and a large range of selectivity. By reported Carnegie classification, four were "Research very high," one was "Research high," eight were "Master's large," one was "Master's medium," one was "Master's small," three were "Bac/A&S," and two were "Public 2-year." Our database includes only participants who completed the entire inventory.

RESULTS

In this paper, we report only levels of significance at or above the 99.9% confidence level, in accord with the recommendations of Johnson (21). As one test of reproducibility of the SLCI, we randomized the 17,741 lines of data and then calculated Cronbach's standardized coefficient alpha for successive 50% splits. Research grade reliability was retained to small N values (Appendix 1, Table 1). We summarize the reliability and validity attributes of the current 25-item SLCI in Appendix 1, Table 2.

Low-stakes effort and guessing

Wise and DeMars (50) believed that low-stakes assessment tests induce low student motivation and lead to substantial underestimation of student proficiency, but Henderson (17) noted little difference in scores on the Force Concept Inventory (18), whether or not students took it as a high-stakes graded test. The structure of the SLCI, with 25 items with one correct answer and three plausible distracters, allowed our employing a Bernoulli test to model the distribution of scores in the extreme case of all students guessing randomly (zero science literacy) and compare this to the distribution of actual SLCI scores (Fig. 1).

By probability, random guessing can generate only about 1% of SLCI scores above 48%. If all students were

TABLE 2.
Correlations from 12 institutions by academic rank between their mean SLCI averages that we calculated and their institutionally-reported average SAT and ACT scores.

	SAT Verbal	SAT Math	ACT Composite
SAT Verbal	1.00		
SAT Math	0.95	1.00	
ACT Composite	0.98	0.97	1.00
Freshmen SLCI	0.83	0.77	0.80
Sophomore SLCI	0.90	0.84	0.85
Junior SLCI	0.80	0.70	0.73
Senior SLCI	0.78	0.69	0.79

SLCI = science literacy concept inventory. All correlations shown are significant. We chose the SAT Composite for further comparison with the SLCI.

random-guessing, 90% of the scores should be between 12% and 36%, but only 10.4% of undergraduates' actual scores registered in this low range. Even undergraduates who had never completed a college science course (N = 3,123) had a mean score of 65%, and 85% of undergraduate students achieved scores greater than 48% (Fig. 1). Figure 1 confirms that the vast majority of students made a sincere effort to respond to the SLCI when provided as a low-stakes assessment.

Removing scores that lie between 12% and 36% would eliminate 90% of respondents who may have guessed and had no interest in trying to answer the challenges. However, such culling eliminates legitimate contributions from students who sincerely answer the items and yet perform no better than people who randomly guess. The average for 17,362 undergraduates on the 25-item SLCI was 68.33%, and the culled data set for 15,580 undergraduates students produced an average of 72.98%. For reasons discussed later, the actual average of American undergraduates overall is likely close to the lower value. Unless otherwise noted, we used the full data set of actual student performance without any culling.

The data participants provided in this study supported the validity of the SLCI (Appendix 1, Table 2). The SLCI given as a low-stakes homework assignment seems sufficient for obtaining valid data. Full credit for doing the inventory with incremental extra credit for scoring above a certain percentage on the assignment appeared to be an acceptable motivation for students to make a sincere effort in addressing the assignment.

For a person with a citizen-level science literacy, the inventory items should not prove very difficult. We confirmed that condition with 24 out of our 25 items (see Appendix 1, Fig. 1 and Table 10, disclosing difficulty and discrimination of the 25 SLCI items). The most troublesome (item 15) sought to register participants' ability to distinguish "why and how" explanations of physical world phenomenon (science) from a statement that only described utilizing the phenomenon
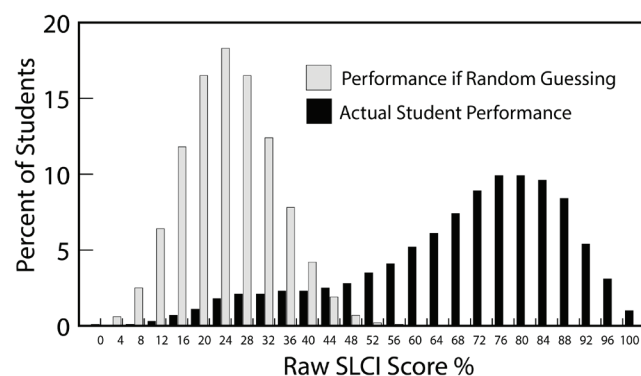
for a purpose (technology). We initially assumed an inappropriately worded item accounted for this difficulty. Yet scores did not improve after modifying the item or testing the concept with other items. Experts (faculty) did much better than did novices (undergraduates) on this item and at a highly significant level (99.9% confidence) on all 12 concepts. Nevertheless, correctly distinguishing the thinking of science from the thinking of technology remained the most difficult of the 12 concepts even for experts. Possibly, American education so rarely addresses the distinction between technology and science that both experts and novices are less prepared to engage successfully with this concept than with the others. Item 15 even detracts slightly from the overall SLCI (Cronbach alpha = 0.841 with item 15 and 0.847 without the item). Because we feel that this dilemma of distinguishing science from technology remains unresolved, we report results of both the full 25-item SLCI and the 24-item SLCI (minus item #15), where appropriate, in Appendix 1.

One-dimensional character of the SLCI

The SLCI addresses 12 concepts/outcomes, but factor analysis showed that the inventory is one-dimensional (Appendix 1, Fig. 2). The multiple-choice items draw from our 12 concepts, but do not factor into separate dimensions that might reveal useful information about relationships between understandings of separate concepts.

The QLRA also addresses about a dozen concepts/outcomes through a brief, 20-item instrument. It too proves to be one-dimensional (11). We believe that both instruments exhibit one-dimensionality because of their necessary brevity. The concepts represented in the items of the SLCI and QLRA collectively meet the goal to measure the constructs of science or quantitative literacy respectively (35). Instruments capable of reliably measuring separate comprehension of each of the concepts that the instruments address would require many more items. The length of such an instrument seems to make it impractical for routinely assessing the overall goals of citizen-level quantitative literacy or science literacy.

Effectiveness of general education in teaching science's way of knowing

Our results indicate that one or two GE science courses have little impact on increasing undergraduates' abilities to recognize and use science's way of explaining the physical world. No clearly significant mean differences in understanding seemed apparent between those who had never completed a GE science course and those who had completed two science courses (Fig. 2). Three courses produced a significant difference over none or two courses, but not over one course. This somewhat supports White et al. (49), who reported, "science curricula failed to develop essential critical thinking skills in many students." A few

FIGURE 1. Actual undergraduate student performance (N = 17,362) compared with performance through random guessing. Mean overall score = 68.33%. Distribution reveals little total ignorance of science's way of knowing among respondents, but room exists for significant improvement. SLCI = science literacy concept inventory.

institutions that we studied achieved significant increases at the completion of four science courses, and a clearly significant increase in every institution that we examined occurred after having more than four science courses (Fig. 2; Appendix 1, Table 4).

Some instructors employed the SLCI in pre-/post-course assessments. This provided 1,310 SLCI course-scale assessments employing longitudinal data (pre/post data from the same students) obtained from over 40 course sections. The correlation between pre- and post-course scores is $r = 0.71$. The same students who scored low or high on the pre-course inventory continue to do so on the post-course measure (see Appendix 1, Fig. 3). This result indicates that it is difficult for single courses, at least as currently taught, to produce convincing changes in students' understanding of science as a way of knowing.

The overall post-course mean (74.65%) was 2.95% higher than pre-course means (71.70%), and the small difference was nevertheless significant at the 99.9% confidence level. This was a greater gain than that shown overall by the none-through-two science courses shown in Fig. 2. The improved gain may have resulted from the fact that the particular instructors who tracked pre- and post-course assessment were striving to place a greater focus on teaching science as a way of knowing than did instructors who only gave the SLCI once. In even the course sections that measured pre/post SLCIs, the means revealed that zero gains, small losses, or small but insignificant gains were common. In looking at the differences between post-course and pre-course assessments for 1,310 students, 43% of students fell between 4% gains and 4% losses.

In our study, gains in understanding science's way of knowing seem more the result of prolonged education across all disciplines rather than of GE science courses in particular. Our database of 17,712 that included 17,382 undergraduates, 149 graduate students, and 181 professors revealed significant jumps in average SLCI scores with academic rank (Fig. 3). Freshmen SLCI averages were lowest. Sophomores-juniors' scores were significantly higher than those of freshmen but were not significantly different from one another. After that, seniors, graduate students, and professors each scored successively significantly higher than the ranks below them.

Professors scored about 90% on the SLCI, and many professors who took the inventory were not professors of science. The others came from the social sciences, arts, humanities, and mathematics and engineering. Mean scores between professors of science, social science, and humanities were not significantly different from one another but were significantly higher than those of professors of arts and engineering, which did not differ significantly from one another. Interestingly, the first three groups emphasize evidence-based evaluative reasoning and logic, whereas the latter two emphasize creative design and innovation. The factual knowledge and skills provided by
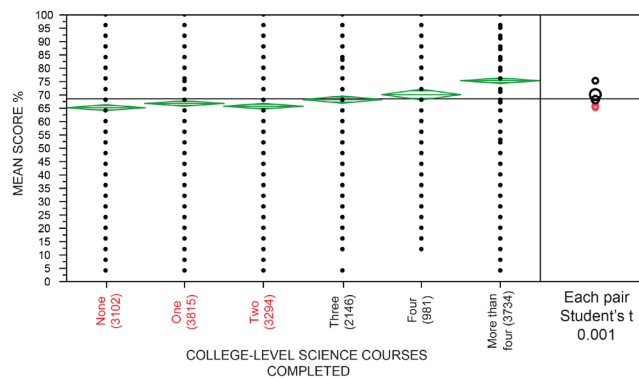

FIGURE 2. SLCI mean scores for 17,072 undergraduates based on number of college science courses completed. Little mean difference exists between none, one, and two courses (red font). The height of the diamonds denotes the 99.9% confidence interval, which is the same as the diameters of the circles in the Student's t-test box to the right. Space separating the circles denotes significant differences between course categories. SLCI = science literacy concept inventory.
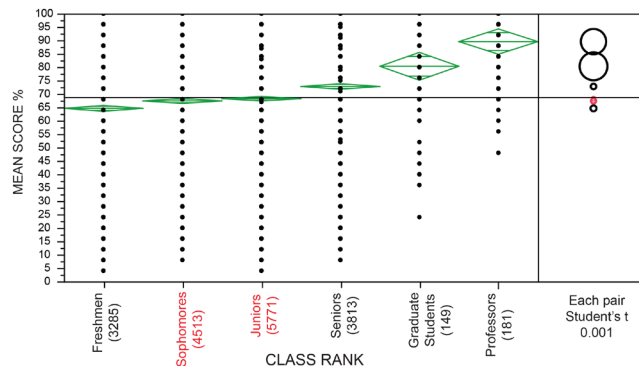

FIGURE 3. SLCI mean scores arranged in increasing order of self-reported ranks for undergraduates, graduate students, and professors. Highly significant differences exist between successively higher academic ranks except for an absence of significant difference between sophomores and juniors (red font). SLCI = science literacy concept inventory.

diverse science courses in the curricula of engineering faculty did not advantage their performance on the SLCI over the performances rendered by professors of arts, humanities, sciences, social sciences. The groups of professors from every metadiscipline significantly outscored the average undergraduate students, the average senior, and the average graduate student.

Based on the consensus afforded by adult models of intellectual development (22), successively higher academic ranks should be associated with higher-stage thinking and reasoning. The SLCI results (Fig. 3) reflect this. We remain uncertain how much of the pattern of gains displayed in Figure 3 is the result of successful intellectual development and how much selectivity and attrition influence this pattern. Many of the lowest-scoring freshmen represent those who are later missing from students tested as sophomores. Fewer of the lowest-scoring seniors may be able to enter graduate schools, and fewer of the lowest-scoring

graduate students may complete their degrees and become professors. We caution individual institutions that employ the SLCI for assessment to interpret their results in light of their attrition and graduation rates.

Our database includes undergraduates' scores from 21 institutions. Institutional SLCI mean scores calculated from sufficient institutional student participants ranged from 63.2% to 87.4% (Appendix 1, Table 9), which showed that the overall difficulty of the SLCI is appropriate for distinguishing in aggregate the varied competencies of undergraduates. A dozen institutions in our study had sufficient data to allow comparisons of the mean scores of the institution's average SAT Verbal and Math, and ACT Composite scores with the SLCI scores by undergraduate class rank. The SLCI correlations with these well-established proprietary tests were sufficiently high (Table 2), to allow meaningful line-fits to determine what each rank's approximate mean SLCI scores should be, given their institution's average SAT and ACT scores (Fig. 4).

The resulting graph (Fig. 4) revealed strong effects of selectivity of the institutions on the mean SLCI scores, an effect also noted in the assessment of quantitative reasoning (11). Higher selectivity also produces a ceiling effect that reduces the gains that highly selective institutions can produce over time in their students.

One of the most important insights gleaned from Figure 4 is that "institutional success" should be determined by the gains imparted to the students within the institution and not by comparing gains or scores across institutions of varied selectivity.
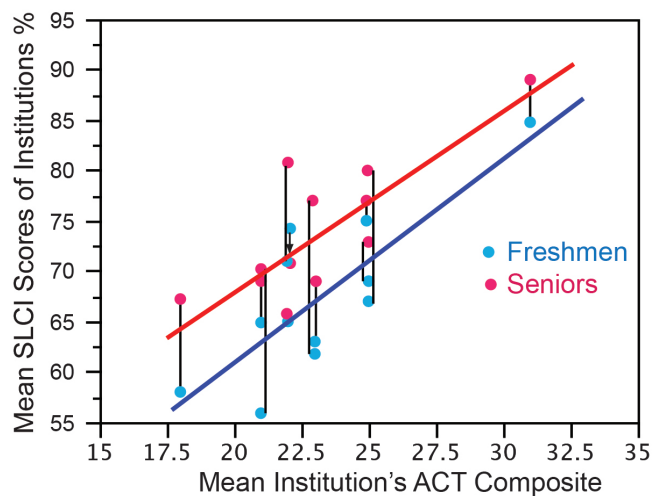

FIGURE 4. Gains in SLCI mean scores of freshmen and senior ranks for 12 institutions of varied selectivity. One school with a mean ACT of about 22 showed a decrease, but all seniors sampled at that school came from introductory courses and were not likely representative of the institution's actual seniors. To do meaningful plots requires a school to obtain a good representative sampling of upper- and lower-division courses. As shown by the slopes of the line fits and their convergence, a ceiling effect makes it more difficult for selective schools to achieve large gains. SLCI = science literacy concept inventory.

Demographic effects

Table 3 and Appendix 1, Figure 5 and Table 7, summarize the mean SLCI scores on 1) gender, 2) status as a first-generation college student, 3) status of English as a first language, and 4) status of science interest, indicated by registration as a science major or aspiring to become a science major combined. The influence of the last three of these four demographic factors produced highly significant differences in the means of undergraduates' SLCI scores.

The percentage of women (62%) in our study populace (Table 3) is somewhat higher than the 2013 national average, where undergraduate enrollments consisted of 56% women (National Center for Education Statistics, http://nces.ed.gov/programs/coe/indicator_cha.asp). Reports vary on the percentage of undergraduates consisting of first-generation students, but sources such as Diverse Education reported about 50% (http://diverseeducation.com/article/50898/), which is about 10% greater than the percentage of first-generation students in our study populace. Less than 8% of American Bachelor's degrees awarded in 2011–2012 were in the sciences (32), but science majors and aspiring majors comprised about 41% of our study. This much higher percentage occurred because participants in our study came mostly from science classes. Status of commitment to science confers a significant mean point advantage, indicating that the actual overall national SLCI mean of nonscience majors (66.2%) better represents the actual national average than does the overall mean of 68.3% we obtained from the 17,362 undergraduates in this study.

Differences between men's and women's SLCI scores were not statistically significant at the 99.9% confidence level (Table 3; Appendix 1, Fig. 5). Estimating instrument gender bias by comparing only the scores of two genders in the absence of demographic variations can be misleading if the populations of the two genders differ substantially in socioeconomic makeup of English as a first language, first generation, or science major/interest.

At the 99.9% confidence level, women overall were insignificantly different from men in: having English as a first language, being first generation students, or declaring an interest in majoring in science. The 25-item SLCI reveals a slightly higher men's score and the 24-item SLCI registers a slightly higher women's score (Table 3), neither of which is significant at the 99.9% confidence level. These results indicate a gender-neutral instrument. However, significant gender differences appeared in some single-campus data. In such cases, the three demographic factors were distributed differently between the genders on those campuses, which explained the significant gender differences in SLCI scores. These campuses were previously unaware that such demographic differences existed by gender in their students.

The fact that the inventory produced gender-neutral results does not mean that the 25 individual SLCI items are each gender-neutral. Our large database allowed us to

TABLE 3.
Summary of results for undergraduate students of the 25- and 24-item SLCI.

Subcategory	N and % of Subcategory		Mean 25-Item SLCI Score	Mean 24-Item SLCI Score
Gender				
Women	10,747	62.0%	68.199%	70.114%
Men	6,585	38.0%	68.496%	70.024%
First Generation Student				
First generation	6,846	39.5%	65.582%	67.375%
Not first generation	10,474	60.5%	70.101%	71.858%
Science Commitment				
Science major or interest in becoming a major	7,080	40.9%	71.454%	73.279%
Nonscience major	10,224	59.1%	66.196%	67.929%
Native Language				
English as first language	13,961	80.7%	69.722%	71.513%
English as non-native language	3,334	19.3%	62.454%	64.130%

All categories except gender revealed statistically significant differences at the 99.9% confidence level. A correlation of the 25-item SLCI with the 24-item SLCI, $r = 0.996$, (Appendix 1, Table 5) shows that the two versions provide consistent results. SLCI = science literacy concept inventory.

discern significant differences that were not evident from our initial pilot test data yielded by fewer students. Fourteen items proved biased for women; bias for men exists for the other eleven (Appendix 1, Fig. 4). The SLCI is gender-neutral because of its overall balance of biased items. To ensure a gender-neutral assessment, users should employ the entire inventory and not a subgroup of selected items. Explaining *why* certain items elicit different gender responses is beyond the scope of this study.

Mean differences in SLCI scores between several ethnic groups (Table 3) were statistically significant. As established above, we have very high confidence that status as a first generation student, as a speaker of English as a first language, and having an interest in majoring in science produce significant differences in the means. These three socioeconomic factors are distributed very differently across ethnicities (Table 4; Appendix 1, Table 3). The order of influence on SLCI scores from greatest to least are: 1) English as the native language (positive: +7.2% mean advantage), 2) interest in majoring in science (positive: +5.26% mean advantage), and 3) status as first-generation student (negative: –4.52% disadvantage).

In our efforts to relate the three socioeconomic factors to SLCI scores across ethnicities, we were only partially successful. Our first effort to combine the three factors through JMP's Fit Model produced the following regression equation: Adjusted Score = 0.6262 – 0.0316 (first generation) + 0.0611 (English as native language) + 0.0504 (science interest). The model retained the relative importance of the three factors and registered all three as significant at $p <$

0.0001, but together these three accounted for only 4.4% of the variance in individuals' scores. However, aggregating individuals' data by ethnicity (Table 4; Appendix 1, Table 3) yielded an adjusted score that greatly reduced the ethnicities' mean differences in scores.

In a second effort, we used Figure 1 to remove the effects of guessing by culling our data set of all scores less than or equal to 40% (Appendix 1, Table 3). We realize that such culling removed legitimate low scores as well as participants who engaged the survey through low-effort random guessing. Culling left a participants' N of 15,003 with which to generate a second regression equation: Adjusted Score = 0.6700 – 0.0233 (first generation) + 0.0390 (English as native language) + 0.0390 (science interest). Unequal degrees of possible guessing existed across ethnicities (Appendix 1, Table 3). In the culled data set, the order of importance from greatest to least were: 1) English as the native language (positive: +4.85% mean advantage), 2) interest in majoring in science (positive: +4.02% mean advantage), and 3) status as first-generation students (negative: –3.19% disadvantage). The greatest difference between this model and the first was the markedly lower effect of English as a native language. This second model also registered all three factors as highly significant at $p < 0.0001$ but only accounted for 4.6% of the variance in individuals' scores. Aggregating this data by ethnicity reduced the differences in different ethnicities' scores even further (Table 4; Appendix 1, Table 3).

The ethnic groups initially exhibited great differences across these three factors (Table 4), but the unequal distri-

TABLE 4.
Tabulated averages by ethnic distribution of undergraduates who took the 25-item SLCI.

Ethnicity	Mean 25-Item Raw SLCI Score	% that Were First Generation Students	% that Reported English as First Language	% that Were Science Majors or Interested in Becoming a Science Major	Adjusted Mean Score Estimated by Factoring in Effects of Prior Three Columns
Overall (N=17,306)	68.29%	39.52%	80.71%	40.94%	68.36%
Asian (N=1,548)	64.75%	41.98%	57.14%	47.17%	67.16%
Black (N=855)	59.88%	50.59%	88.53%	43.59%	68.63%
Caucasian (N=9,272)	71.27%	24.46%	97.09%	42.44%	69.92%
Hispanic (N=3,905)	66.49%	66.15%	51.98%	33.74%	65.41%
Middle Eastern (N=391)	52.81%	46.8%	37.60%	37.28%	65.32%
Native American (N=131)	68.00%	53.85%	96.15%	57.69%	69.70%
Pacific Islander (N=261)	66.04%	38.85%	90.77%	44.79%	69.20%
Other (N=944)	67.02%	39.64%	83.03%	40.86%	68.50%

Each group's makeup of three factors—first generation student, English as native language, and commitment to/interest in a science major—strongly influences each group's mean SLCI score. Employing a regression equation derived from all the undergraduate participants to adjust for these significant influences greatly reduced the score differences between ethnic groups. SLCI = science literacy concept inventory.

bution of these factors proved powerful in both models in accounting for the large mean differences in raw scores seen between ethnic groups. The significant differences in *mean raw* SLCI scores between ethnic groups (Table 4) appear to be largely the products of the socioeconomic conditions represented differentially by each ethnic group. This result supports the contention that socioeconomic conditions determine the kinds of support that students are likely to have in their homes and in the local schools that they attend, which offers a significant advantage to students (16).

In the first regression model, the raw scores showed the Caucasian majority as scoring the highest on the SLCI (Table 4). Having the lowest percentage of first-generation students and the highest percentage with English as their native language advantaged Caucasian participants. Hispanics, the largest minority group, had the highest proportion of first-generation students of any ethnic group. Only about half of Hispanics had English as a first language, and Hispanics had the lowest percentage of participants expressing an interest in majoring in science. Middle Eastern students registered an aberrantly low SLCI mean score of about 16 percentage points below the overall means (Table 4). We considered stereotype threat (42) as possibly a significant factor, given that the Middle Eastern ethnicity was a highly maligned stereotype in America during this study period. However, the effect of the Middle Eastern ethnic group's high percentage of students with English as a non-native language largely explained the immense gap. In this model, the maximum difference between mean raw scores between ethnicities of 18.5% was reduced to a maximum difference

of 4.6% in adjusted scores.

In the second regression model, the maximum difference between mean raw scores of different ethnicities (8.28%) was reduced to a maximum difference of 3.04% (Appendix 1, Table 3). After adjustment, the ethnicities with the lowest mean SLCI scores were those with the highest percentages of English as a non-native language.

DISCUSSION

In this paper, we affirmatively answered three questions:

1. Can we isolate and describe the major concepts that constitute GE citizen-level science literacy?
2. Can we reliably assess the construct of citizen level science literacy through addressing the concepts on a simple concept inventory?
3. Can the resulting data yield information of value?

The SLCI provides reliable, valid, and relevant data on science literacy.

Most students enter college with a measurable ability to understand science's way of knowing that is far beyond zero literacy. Our results show that college GE science courses in themselves make unconvincing contributions to extending the understanding gained in K–12 grades. This result may occur because the courses focus on conveying disciplinary knowledge and skills without teaching science as a way of knowing. Alternatively, measurable gains in capacity to reason may require more time than a semester or

year to develop, regardless of the instructional emphasis. Our results show that the overall college experience does advance such understanding, and mean scores increase with academic rank in accord with the trend expected from established models of adult intellectual and ethical development. Helping students to comprehend science as a way of knowing may contribute to their achieving higher-level stages of adult intellectual development.

Most institutions that acquired a large database from the SLCI confirmed marked gains between freshmen and senior years. An institution's understanding of the nature of such gains requires tempering interpretations with institutional data on the magnitudes of attrition between students' ranks. Institutional selectivity correlates strongly with mean SLCI scores at all academic ranks. Demographic data show that the proportions of a populace with English as a first language, an interest in majoring in science, and having first generation status as college students significantly affect the mean SLCI score of that populace. The populace affected can represent an ethnic group, a gender, or an institution. Given the reliability of the SLCI (R = 0.84), the small populace representing some ethnic groups, and the imperfect nature of our regression models, we are unable to attribute the small differences that remain after adjustment to other factors. The three factors may exert synergistic influences with one another in ways that we do not yet understand. We interpret the results displayed in Table 4 and Appendix 1, Table 3, as indicating that the three socioeconomic factors are important influences.

Our analysis confirmed that women and men are equally adept at understanding science as a way of knowing. Likewise, every ethnic group seems equally capable of achieving higher-level reasoning afforded by understanding science's evidence-based way of knowing. Where merited, universities may be able to improve conditions in their communities by providing tutoring in college-level English and by increasing awareness of both the desirable opportunities that exist in the sciences for careers and the importance of citizen awareness of science.

We have shown that the current version of the inventory is sufficient for gaining valuable information (see Appendix 1, Figs. 3 and 5, Tables 3, 4, 6, 7, 8, and 9). Additional studies based on this rich and growing database are ongoing for future publication. Opportunity exists to improve any test through revision, and we continue to test one or two unscored items in addition to the 25 shown here, as a way to improve the inventory.

The selection of distracters by participants (Appendix 1, Table 10) offers information about the relative importance of misconceptions, and Appendix 1, Figure 6, confirms that participants understand the questions and collectively can even ascertain the relative difficulty of the 25 items, as confirmed by the percentage of participants who correctly answer the items. The explanation of self-assessment as informed by the SLCI is the topic of another paper in preparation.

To improve students' thinking, we recommend that instructors of general education science courses employ lesson designs that engage students in several metacognitive reflections on the reasoning component of citizen-level science literacy. These reflective exercises should become an integral part of learning, along with disciplinary content and skills. We recommend using the SLCI as pre-/post- assessments at the course and program levels to assess the impact of these reflections. Readers can view the SLCI at http://tinyurl.com/jmbeslciaccess. Faculty interested in using the SLCI should contact the corresponding author.

SUPPLEMENTAL MATERIALS

Appendix 1: Supplementary figures and tables

ACKNOWLEDGMENTS

REFERENCES

1. **American Association for the Advancement of Science.** 1989. Science for all Americans. Oxford University Press, New York, NY.

2. **American Association for the Advancement of Science.** 1993. Benchmarks for science literacy. Oxford University Press, New York, NY.

3. **Antonellis, J., S. Buxner, C. Impey, and H. Sugarman.** 2012. Surveying science literacy among undergraduates: insights from open-ended responses. J. Coll. Sci. Teach. **413**:82–90.

4. **Brookfield, S.** 2012. Teaching for critical thinking. John Wiley and Sons, San Francisco, CA.

5. **Chamberlin, T. C.** 1897. The method of multiple working hypotheses. J. Geol. **5**:837–848. Reprinted: 1965. Science

148:754–759.

6. **Coil, D., M. P. Wenderoth, M. Cunningham, and C. Dirks.** 2010. Teaching the process of science: faculty perceptions and an effective methodology. CBE Life Sci. Educ. **9**:524–535.

7. **Duit, R.** 2009. Students' and teachers' conceptions and science education: full version bibliography. [Online.] http://archiv.ipn.uni-kiel.de/stcse/download_stcse.html. Accessed 22 August 2015.

8. **Flavell, J. H.** 1976. Metacognitive aspects of problem solving, p 231–236. *In* Resnick, L. B. (ed.), The nature of intelligence. Erlbaum, Hillsdale, NJ.

9. **Flavell, J. H.** 1979. Metacognition and cognitive monitoring: a new area of cognitive-developmental inquiry. Amer. Psychol. **34**(10):906–911.

10. **Gaze, E.** 2014. Teaching quantitative reasoning: a better context for algebra. Numeracy **7**(1):Article 1. [Online.] http://scholarcommons.usf.edu/numeracy/vol7/iss1/art1/. Accessed 22 August 2015.

11. **Gaze, E. C., A. Montgomery, S. Kilic-Bahi, D. Leoni, L. Misener, and C. Taylor.** 2014. Towards developing a quantitative literacy/reasoning assessment instrument. Numeracy **7**(2):Article 4. [Online.] http://scholarcommons.usf.edu/numeracy/vol7/iss2/art4/. Accessed 22 August 2015.

12. **Gill, R. A., and I. C. Burke.** 1999. Using an environmental science course to promote scientific literacy. J. Coll. Sci. Teach. **25**:105–110.

13. **Glazer, F. S.** 2000. Journal clubs—a successful vehicle to science literacy. J. Coll. Sci. Teach. **29**(5):320–324.

14. **Gormally, C., P. Brickman, and M. Lutz.** 2012. Developing a test of scientific literacy skills TOSLS: measuring undergraduates' evaluation of scientific information and arguments. CBE Life Sci. Educ. **11**:364–377.

15. **Hallowell, C., and M. J. Holland.** 1998. Journalism as a path to scientific literacy. J. Coll. Sci. Teach. **28**(1):29–32.

16. **Hattie, J.** 2009. Visible Learning: A synthesis of over 800 meta-analyses relating to achievement. Routledge, New York, NY, p 61–63; 68–71.

17. **Henderson, C.** 2002. Common concerns about the Force Concept Inventory. Phys. Teach. **40**:542–547.

18. **Hestenes, D., M. Wells, and G. Swackhamer.** 1992. Force Concept Inventory. Phys. Teach. **30**:141–158.

19. **Holton, G.** 1999. 1948: The new imperative for science literacy. J. Coll. Sci. Teach. **28**(3):181–185.

20. **Impey, C., S. Buxner, J. Antonellis, E. Johnson, E., and C. King.** 2011. A twenty-year survey of science literacy among college undergraduates. J. Coll. Sci. Teach. **404**:31–37.

21. **Johnson, V.** 2013 Revised standards for statistical evidence. Proc. Natl. Acad. Sci. U.S.A. **110**:19313–19317.

22. **Journal of Adult Development.** 2004. Special volume of nine papers on the Perry legacy of cognitive development. J. Adult Dev. **11**(2):59–161. Periodicals Service Co., Germantown, NY.

23. **King, P. M., and K. S. Kitchener.** 1994. Developing reflective judgment: understanding and promoting intellectual growth and critical thinking in adolescents and adults. Jossey Bass, San Francisco, CA.

24. **Laugksch, R. C., and P. E. Spargo.** 1996. Construction of a paper-and-pencil test of basic scientific literacy based on selected literacy goals recommended by the American Association for the Advancement of Science. Public Underst. Sci. **5**:331–359.

25. **Lawson, A. E.** 1978. Development and validation of the classroom test of formal reasoning. J. Res. Sci. Teach. **15**(1):11–24.

26. **Libarkin, J.** 2008. Geoscience concept inventory: suggestions for question review and development: Notes Provided from Geological Society of America Annual Meeting Short Course # 523. Writing and Evaluating Geoscience Concept Inventory Questions.

27. **Maskey, C. L.** 2011. An evaluation of the relationship between reflective judgment and critical thinking in senior associate-degree nursing students. Doctoral dissertation. Indiana University, Indianapolis, IN.

28. **McComas, W. F.** 1998. The principal elements of the nature of science: dispelling the myths, p 53–70. *In* McComas, W. F. (ed.), The Nature of Science in Science Education. Kluwer Academic Publishers, Netherlands.

29. **Miller, J. D.** 1983. Scientific literacy: a conceptual and empirical review. Daedalus **1122**:29–48.

30. **Miller, J. D.** 1998. The measurement of civic scientific literacy. Public Underst. Sci. **7**(3):203–223.

31. **National Academy of Sciences and Institute of Medicine.** 2008. Science, evolution, and creationism. The National Academies Press, Washington, DC.

32. **National Center for Education Statistics.** 2012. Institute of Education Sciences. Digest of Education Statistics 2013, Table 318.20. Bachelor's, master's, and doctor's degrees conferred by postsecondary institutions, by field of study: Selected years, 1970–71 through 2011–12. [Online.] http://nces.ed.gov/programs/digest/d13/tables/dt13_318.20.asp. Accessed 23 August 2015.

33. **NGSS Lead States.** 2013. Next generation science standards: for states, by states. The National Academies Press, Washington, DC.

34. **Nuhfer, E. B.** 2013. Understanding Cinderella—expanding an essential conversation on the metadiscipline of technology: educating in fractal patterns XXXVII. Natl. Teach. Learn. Forum **22**(2):7–10.

35. **Nuhfer, E. B., and E. Gaze.** 2014. Benefits of assessing citizen literacy in science and quantitative reasoning: educating in fractal patterns XLII. Natl. Teach. Learn. Forum **24**(1):8–11.

36. **Nuhfer, E. B., et al.** 2010. Multi-campus project: promoting and assessing science literacy in general education science courses. California State University, Institute for Teaching and Learning Connections **3**(4). [Online.] https://www.calstate.edu/itl/newsletter/10-summer.shtml. Accessed 18 August 2015.

37. **Pavelich, M. J., and W. S. Moore.** 1996. Measuring the effect of experiential education using the Perry model. J. Engineer. Educ. **85**(4):287–292.

38. **Perry, W. G., Jr.** 1999. Forms of intellectual and ethical development in the college years. Reprint of the original 1968 1st edition with introduction by L. Knefelkamp. Jossey-Bass,

San Francisco, CA.

39. **Seckam, A., and R. Cooper.** 2013. Understanding how honey impacts on wounds: an update on recent research findings. Wounds Internatl. **4**(1):20–24.

40. **Selby, C.** 2006. What makes it science?: A modern look at scientific inquiry. J. Coll. Sci. Teach. **35**(7):8–11.

41. **Specter, M.** 2010. Denialism: how irrational thinking harms the planet and threatens our lives. Duckworth Overlook, London, UK.

42. **Steele, C. M., and J. Aronson.** 1995. Stereotype threat and the intellectual test performance of African Americans. J. Personality Social Psych. **69**(5):797–811.

43. **Strahler, A. N.** 1992. Understanding science: an introduction to concepts and issues. Prometheus Books, Buffalo, NY.

44. **Strimaitis, A. M., J. Schellinger, A. Jones, J. Grooms, and V. Sampson.** 2014. Development of an instrument to assess student knowledge necessary to critically evaluate scientific claims in the popular media. J. Coll. Sci. Teach. **43**:55–68.

45. **Terry, D. R.** 2012. Assessing critical thinking skills using articles from the popular press. J. Coll. Sci. Teach. **42**:66–70.

46. **Trefil, J.** 1992. Science matters: achieving scientific literacy. Anchor Books. New York, NY.

47. **Trefil, J.** 2008. Why science? Teachers College Press, National Science Teachers Association, Arlington, VA.

48. **Wenning, C. J.** 2006. Assessing nature-of-science literacy as one component of scientific literacy. J. Phys. Teach. Educ. Online **34**:3–14.

49. **White, B., et al.** 2011. A novel instrument for assessing students' critical thinking abilities. J. Coll. Sci. Teach. **405**:102–107.

50. **Wise, S. L., and C. E. DeMars.** 2005. Examinee motivation in low-stakes assessment: problems and potential solutions. Educ. Assess. **10**(1):1–17.

51. **Wolpert, L.** 1992. The unnatural nature of science. Harvard University Press, Cambridge, MA.