Evaluating Knowledge of the Nature of (Whole) Science

DOUGLAS ALLCHIN
Minnesota Center for the Philosophy of Science and Department of Curriculum and Instruction, University of Minnesota, Minneapolis, MN 55455, USA

Received 17 May 2010; revised 12 October 2010; accepted 2 November 2010

DOI 10.1002/sce.20432
Published online 9 March 2011 in Wiley Online Library (wileyonlinelibrary.com).

ABSTRACT: I profile here a prospective method for assessing nature of science (NOS) knowledge, as an alternative to VNOS and similar approaches. Questions about cases in contemporary news and from history probe scientific literacy in context. Scoring targets how "well informed" the analysis is, based on identifying relevant NOS information and interpreting its import appropriately. The assessment shifts focus from declarative statements to functional (or interpretive) analysis. It also entails reframing current NOS characterizations from selective lists of tenets to the multiple dimensions shaping reliability in scientific practice, from the experimental to the social—namely, to Whole Science. This approach underscores the role of reflective student inquiry and historical and contemporary cases in NOS instruction. © 2011 Wiley Periodicals, Inc. *Sci Ed* **95:**518–542, 2011

INTRODUCTION

> Doubting everything or believing everything are two equally accommodating solutions, either of which saves us from reflection.
>
> Henri Poincaré, *La Science et l'Hypothèse*

Climategate, revised mammogram recommendations, autism and the measles vaccine, and facilitated communication of coma patients: Science education purports to inform public and personal decision making in such cases recently in the news. To interpret such

Correspondence to: Douglas Allchin; e-mail: allch001@umn.edu

cases, one may need to understand some basic scientific concepts as background and/or be able to assess simple evidence. But for many cases, understanding the nature of science (NOS) is also essential, if not central: Whose expertise can be trusted, especially when experts seem to disagree? What forms of communicating scientific findings to the public are credible? How do scientists manage data? How do they communicate with each other? What kind of conditions warrant a change in scientific consensus? Where does verifiable information end and value judgment begin? Here are important benchmarks for effective K–12 science education. Yet science educators currently seem ill-equipped to assess in any standardized way (or on any widespread scale) how well informed an analysis a student can develop in such cases. How does one measure this critical dimension of scientific literacy? In this paper, I survey the task and sketch a few prototypes, toward promoting discussion on how to effectively assess practical, culturally functional knowledge of NOS.

The challenge is framed in part by the current politics of education, with its emphasis on accountability. Science education experts, as well as national and international science education reform programs, uniformly advocate NOS education. Yet realistically teachers have few alternatives: They inevitably teach to the test (Bill & Melinda Gates Foundation, 2010). What, then, can one consider a good test of NOS skills? If the goal is "placing the nature of science and its processes at the core rather than the margins of science education" (Osborne, Collins, Ratcliffe, Millar, & Duschl, 2003, p.716), then "efforts are urgently needed to help teachers and creators of high-stakes tests to accurately assess students' understanding of NOS" (Clough & Olson, 2008, p. 145; also see Hodson, 2008, p. 170).

A PROTOTYPE

Consider the question posed in Table 1 as a prospective prototype. It presents a case from current events, as a typical citizen might encounter it in the news media. It asks for a *well-informed analysis*. It does *not* ask for a statement or justification of a particular position, which would largely tend to reflect personal ideology and prompt unconscious rationalization. Rather, it endeavors to probe the level of sophistication in a student's *functional* understanding of scientific practice and its relevance to decision making. In particular, it asks for an analysis of NOS features, *independent of specific content knowledge* (here, of clinical research on mammogram effectiveness). It seeks *breadth and depth* of NOS knowledge. For example,

1. Can the student identify all the relevant NOS factors?
2. Can they articulate their relevance for interpreting the reliability of the claims?
3. Can they profile key information *not* provided and where it may likely be found?

Such questions invite free response, a format extensively developed by the Educational Testing Service for their Advanced Placement (AP) exams, whose scoring is routinely standardized. Other examples of this style of question (based on other contemporary cases that are cited in the opening above, and modeled roughly on the AP essay style) are presented in an Appendix (and available online at http://ships.umn.edu/knows/).

Designing NOS assessment as analyses of extended examples or cases reflects the informal consensus of a group of about two dozen philosophers of science, science policy scholars, high school science teachers, and scientists, who assembled to discuss this problem at a recent interdisciplinary conference on "the philosophy of science in practice."[1]

[1] Session on "Evaluating NOS Knowledge in Science Teaching" at the 2nd Biennial Meeting of the Society for the Philosophy of Science in Practice (Minneapolis, MN, June 2009).

TABLE 1
Sample NOS Evaluation Question

Revised Mammogram Recommendations, Nov. 2009

A female acquaintance of yours is just turning 40. Concerned about the possibility of breast cancer, she had planned to get a mammogram in the next few months, despite her fears about excessive radiation. She has heard that a major national task force now advises waiting until 50, yet finds reassurance in *Women's Health* magazine about still following the old guidelines. You both knew another woman who was diagnosed unexpectedly with breast cancer at age 43 and died last year. Your acquaintance is unsure how to interpret the apparently conflicting information and asks your help. What analysis of this reported change in scientific consensus would you provide to inform her decision?

Resource documents

- *Women's Health* magazine article (Feb. 8, 2010)
  www.womenshealthmag.com/health/medical-tests?cat=18753
- *New York Times* article (Nov. 17, 2009)
  www.nytimes.com/2009/11/17/health/17cancer.html
- U.S. Preventative Services Task Force report: recommendation & supporting stmt. (Nov. 2009)
  www.ahrq.gov/clinic/uspstf09/breastcancer/brcanrs.htm#clinical
  www.ahrq.gov/clinic/uspstf09/breastcancer/brcanart.htm
- Editorial published in *Annals of Internal Medicine* (Feb. 15, 2010)
  www.annals.org/content/early/2010/02/12/0003-4819-152-8-201004200-00210.full

See Table 2 for scoring rubric.

Participant comments underscored the need to allow students to articulate a multifaceted NOS understanding in the context of a case study that is sufficiently complex, on the one hand, yet also clearly delineated (concrete and detailed) on the other. This view paralleled another diffuse consensus: that NOS instruction likewise needs to engage students in problem solving and decision making in context-rich case studies. Discussants felt that students needed experience and guidance (possibly in judiciously simplified or streamlined contexts) for such real-life scenarios. The views of this group, expressing interest in practical cultural contexts of philosophy of science, clearly reflected a now standard educational ideal: that assessment be *authentic*. They also implicitly endorsed the notion that if scientific literacy is the educational goal, scientific-literacy-in-practice should also be integral to both the assessment and the instruction.

(RE)FRAMING THE NATURE OF SCIENCE: FROM LISTS TO WHOLE SCIENCE

The approach exemplified in the sample questions reflects important shared views among educational researchers, teachers, and science studies scholars. Yet one may articulate the context and structure of these prototypes in more detail.

First, what exactly are we assessing—and why? The why is familiar: To prepare citizens to participate in a society where science and technology are increasingly important in public policy and personal lives (Krajik & Sutherland, 2010; OECD, 2009, pp. 14, 126; Osborne, 2007; Rutherford & Ahlgren, 1990). Here, content knowledge is insufficient. The prototype cases from the news above, as samples, are concrete touchstones (Millar & Osborne, 1998; National Research Council [NRC], 1996, p. 22). They highlight the practical importance of also understanding NOS. Sidestepping the political posturing and inflated rhetoric surrounding the concept of scientific literacy, one may adopt the following as a foundational principle:

> Students should develop an understanding of how science works *with the goal of interpreting the reliability of scientific claims in personal and public decision making*.

This approach gives central significance to analysis of *reliability*, or *trustworthiness* (Ziman, 1978).[2] Indeed, former debates about what to teach about NOS—amplified to extremes during the so-called Science Wars—nearly always hinged on how to interpret the *authority* of scientific claims or scientists as spokespersons for those claims. Without prejudicing the resolution to such problems, this is what students foremost need to learn: what, or whom, to trust.

One strategy is to equip students to evaluate evidence on their own: to prepare everyone to make the same judgments scientists do. Such skills certainly seem appropriate where problems and evidence are simple (Prototype Question 5). Few will dispute the goal of developing skills in recognizing relevant empirical findings, interpreting graphs and statistical measures, thinking about controls, considering alternative explanations, etc.

However, there are limits. In the mammogram case (Prototype Question 1), a typical citizen, no matter how well informed, is simply unable to collect and evaluate all the evidence on the benefits, costs, and risks of the procedure at different ages. This was the rationale for a government task force, and one relies on their expertise. Even scientists inevitably rely on other scientists (Goldman, 1999; Shapin, 1996). By comparison, mistaken impressions of one's abilities to evaluate evidence open the way to mischief. For example, Web sites critical of global warming present selective counterevidence, relying on readers' intuitions that their personal "commonsense" judgments can trump the Intergovernmental Panel on Climate Change. Such persuasive tactics apparently have been effective. Pretending that we might train each person to always evaluate the evidence on their own or to "participate in science" in every instance discounts the important lessons on expertise from recent sociology of knowledge and social epistemology (Goldman, 2002; Hardwig, 1991; Selinger & Crease, 2006). Teaching an understanding of the nature of expertise and systems of credibility seems essential in a modern society where technical knowledge is widely distributed among specialized experts (Gaon & Norris, 2001; Norris, 1995).[3] A prospective test of scientific literacy must surely also address skills in analyzing credibility.

Yet understanding the role of expertise, while important, still falls short. Credibility may be challenged. Here, one needs to understand, more deeply, just how scientific practices contribute to credibility. For example, in the case of Climategate (Prototype Question 2), using "tricks" with graphs or trying to limit publication by critics (as discussed in the hacked e-mail messages)—while sounding suspect on the surface—do not reflect fraud. In other cases, knowledge of how science works may help keep claims of credibility in check. For instance, someone aware of Andrew Wakefield's sources of funding may well

[2] The emphasis on reliability may not provide enough attention to other questions, such as the funding of science and the choice of research programs and problems.

[3] These views pose an implicit challenge for those in the William Perry tradition of epistemological development to expand their schemes (King & Kitchener, 1994; Kuhn, Cheney, & Weinstock, 2001; and others). Sophisticated epistemological understanding in today's world must incorporate an additional layer or level, beyond "intellectual independence," that addresses the distribution of expert knowledge, the limits of individual knowledge, epistemic dependence, and systems of credible testimony or knowledge transfer, such as the certification of professional experts (Allchin, 1999).

have questioned his original claim about autism and vaccines, well before it was formally retracted (Prototype Question 4).

Furthermore, experts may be mistaken, as in the case of communication with coma patients by noted Belgian neurologist Steven Laureys (Prototype Question 3). One needs to also understand the nature of uncertainty and possible sources of error. In particular, cultural biases tend to go unnoticed. In addition, experts sometimes disagree. Credible claims may conflict. One needs additional resources to assess the nature of the disagreement and the relative status of alternative claims. Even credible claims may come with qualifications and caveats, whose meaning becomes clear only when one understands the various methods for ensuring reliability, as well as their limits. Trust should not be blind. Credibility merely signals responsible communication; it does not wholly substitute for it.

As science on a particular topic matures, problems of debate and uncertainty tend to be resolved. In most contemporary decision-making cases, however, the science is young, still in-the-making (Latour, 1987). In such circumstances, uncertainty is high. Neither credible voices nor evidence can fully resolve the uncertain possibilities. At such times—those most typical of the challenge of scientific literacy—assessments of the nature and limits of reliable knowledge are especially important for guiding decisions and helping to plan for contingencies. Teaching an understanding the uneasy status of scientific uncertainty, between ignorance and well-founded claims, seems just as important as understanding (the more familiar) "tentativeness" (Friedman, Dunwoody, & Rogers, 1999; Kolstø, 2001).

The informed citizen, then—the mature, well-educated student—will be able (at least) to interact with experts on topics they may know next to nothing about; recognize relevant evidence as well as presentations of bogus evidence; appreciate the limits as well as the foundations of emerging scientific claims; and negotiate through scientific uncertainty. She/he will be a competent interpreter, or "critic," of science, even if not a practitioner of science (in the same way that film or music critics can effectively assess art without necessarily producing art themselves; L. Thomas, 1981). As reflected in the sample cases from the news, interpreting the reliability of scientific claims requires a broad understanding of scientific practice, from experiment to science journalism (see Table 2 and below).

One may compare this perspective to recent consensus among educators on NOS. An analysis of NOS tenets in eight curriculum documents, including the influential American Association for the Advancement of Science (AAAS)'s Project 2061 and the U.S. National Research Council's *National Science Education Standards* and other international counterparts, yielded a short list of shared tenets (McComas & Olson, 1998). These formed the basis for the currently most widely used NOS assessment instrument, VNOS (Lederman, Abd-El-Khalick, Bell, & Schwartz, 2002). The consensus was a list of 8–10 explicit, declarative tenets: science is empirical; observations are theory laden; science is affected by its social and cultural milieu; science is tentative; etc. The strategy of developing a list of basic principles, or "ideas about science," also guided Osborne et al. (2003) in perhaps the most well-developed and authoritative study of this genre. Their analysis benefitted, first, from some intellectual distance from the so-called Science Wars over postmodernism, social constructivism, etc. It was thus securely balanced—neither radical nor reactionary. Second, it was informed by contemporary scholarship in history, philosophy, and sociology of science. At the same time, it was filtered through the perspective of educators, insulated from the bias of both scientists (all too inclined to safeguard their authority) and science studies scholars (sometimes too academic or rhetorically hyperbolic). Third, the study emphasized the process of developing consensus, using structured responses and iterated rounds of discourse. This was a *well-reasoned* consensus. The outcome, as fine a characterization of

NOS for the classroom as one is likely to find in this style, largely paralleled the earlier analysis.[4]

Nevertheless, these lists were not fully contextualized in the aim of "personal and social decision making" involving science. There is yet no evidence that mere recall or comprehension of such NOS tenets is adequate for applying them effectively in context (as in the sample cases). NOS understanding needs to be *functional*, not declarative (Ford, 2008; Rudolph, 2000). The consensus lists are thus deficient or their focus misplaced. For example, consider a presentation at a recent national conference of biology teachers in the United States. An educational researcher and a large team of teachers reported on their effort, clearly quite intensive, to teach students the difference between a hypothesis and a prediction. The presenters implicitly implied that this distinction was not only essential to teaching NOS, but that mastering it constituted a key triumph for students. Definitions and clear use of language were central. No one articulated how this might lead either to better scientific practice or to deeper understanding of scientific claims. Contrast this view with how scientists actually talk with one another—using language rather loosely (e.g., see Wong & Hodson, 2009) and clarifying ideas in context or through continued discourse. Many of them seem to violate the apparent NOS "rule" by referring to "hypothesized results"; yet other researchers neither misunderstand nor care. Was there ever a public controversy in science—about cloning or genetically modified organisms or the safety of chemicals leaching into the soil from decades-old disposal sites—that hinged on proper use of this terminology or distinction? As illustrated in this episode, NOS understanding is appropriately oriented to the reliability of claims in real-life contexts.

Characterizing NOS in terms of interpreting reliability entails reframing current views on NOS, while reflecting other familiar traditions in science education. For example, VNOS-C asks, "What is an experiment?" In the context of science in personal and social decision making, this question is rather metaphysical and irrelevant. What matters instead is whether the evidence, derived through experiment or observation—or any other means—is trustworthy. The classic concept of control (parallel observations differing by a single variable), by contrast, is fundamental, whether applied to a laboratory experiment or field study, a natural experiment, or statistical analysis of a large data set. Philosophizing about the abstract nature of experiment can be left to ... well, philosophers. As noted in Project 2061's presentation of its revised benchmarks, NOS is not philosophy of science (AAAS, 2009, chap. 1). To become well-informed adults and responsible citizens, students need to understand how evidence works—and where it can fail. Labels and formal definitions must yield to a practical and functional understanding.

Least of all does one need to distinguish between laws and theories (see Wong & Hodson, 2009, pp. 122–123, for how scientists talk). What matters, again, is how (irrespective of labels) one ascertains the degree of confidence in a particular claim. Indeed, the best way to disarm criticism of evolution as "merely a theory" may *not* be by clarifying the meaning of

[4] The appropriateness of the Osborne et al. consensus was informally surveyed recently at an interdisciplinary meeting of philosophers of science, science policy academics, and scientists (2nd Biennial Meeting of the Society for the Philosophy of Science in Practice, Minneapolis, MN, June 2009). The weight of opinion among the academic contingent was much narrower than Osborne et al. Ironically, perhaps, discussion comments underscored a prevalent desire to orient science education foremost to motivating an interest in or appreciation of science, not to NOS learning. By contrast, a group of 13 science teachers committed to the role of history and philosophy in science education (assembled at the same time to write classroom case studies) found the consensus far too conservative and added a near universal consensus on several features that the Osborne group apparently found problematic. Nonetheless, they generally felt (with equal irony) that such a list was not very helpful in informing NOS education or designing NOS curriculum.

the term "theory," but rather by rendering the whole discussion moot by redirecting focus to the robustness of the evidence.

As much as the NOS consensus list includes many items irrelevant to functional scientific literacy, it also omits many relevant others. The significant role of credibility, for example (even among scientists), rarely appears. Indeed, all items about the social interaction of scientists—especially the system of checks and balances through mutual criticism—are typically absent (Allchin, 2004a). The lists also disregard the role of funding, motivations, peer review, cognitive biases, fraud, and the validation of new methods. To gauge what is important, one may simply survey science in the news and catalog the NOS elements critical to interpreting the claims that appear there. One such list appears in Table 2 (with items linked to the five cases in the prototype questions) (compare to Millar, 2000, Table 3; Kolstø, 2001).

Such an inventory spans a wide range, from experimental protocols and controls through statistical analyses and theoretical reasoning to considerations of credibility, conflicts of interest, gender bias, and the economic contexts of science journalism. A scientifically literate individual will thus have a broad understanding of scientific practice (Rudolph, 2000). Namely, how are scientific claims generated and also transmitted? What ensures reliability at each step, as each may prove important in different cases? All are addressed in the science studies literature. The inventory may seem long and unwieldy, but (unlike the consensus list) it is unified by the theme of reliability. Items may also be easily organized: by following claims as they unfold in successively broadening contexts, from observational settings to public forums: from lab bench to judicial bench (Latour, 1987, pp. 195–257). For convenience, one may sort them into functional epistemic categories.[5] Table 2 presents one prospective taxonomy.

Ironically, such a profile of reliability in scientific practice parallels potential sources of error, or error types, in science (Allchin, 2001a). We may need to inform students about all the ways scientific claims may fail, so that they understand how we prevent, mitigate, or accommodate potential error (Allchin, 2004b, 2004c; Guinta, 2001). Complete understanding of NOS, in this view, has both breadth and depth (completeness and proficiency). Ideally, an individual will exhibit understanding *in each category or domain of potential error* and be able to articulate each *with a certain level of detail, concreteness,and context*. Each may be assessed.

One may call this reframing of NOS, sensitive to all the dimensions of reliability in scientific practice, *Whole Science*. Whole Science, like whole food, does not exclude essential ingredients. It supports healthier understanding. Metaphorically, educators must discourage a diet of highly processed, refined "school science." Short lists of NOS features should be recognized as inherently incomplete and insufficient for functional scientific literacy.

The nature of science, then, cannot be fully or adequately expressed by a list of explicit tenets. Rather, one frames it as a set of dimensions about how reliability is achieved as knowledge develops and how it is preserved as it moves from one place to another. The dimensions profile how science works—and at the same time how it sometimes does not work (and why). The concept of dimensions helps unify some puzzling contradictions in widespread views of NOS. For example, how does the student reconcile "investigation is

[5] The term "epistemic" here contrasts with "epistemological," more familiar perhaps in educational literature. *Epistemics* addresses the pragmatic, naturalized process of developing knowledge and characterizing its uncertainties, in contrast to *epistemology*, which concerns the more abstract (sometimes metaphysical) justifications and limits of knowledge in an ultimate form. Epistemic perspectives are inherently empirically based, relying on historical, sociological, and cognitive studies. In addition, epistemology typically focuses on the "context of justification," whereas epistemics addresses both "discovery" and "justification."

TABLE 2
Dimensions of Reliability in Science

Bulletpoints identify possible scorable items on a free response question about NOS. This list is illustrative, not exhaustive. [Items relevant to questions in Table 1 and Appendix are noted as follows: M = mammogram recommendations; C = Climategate; A = autism; F = facilitated communication; B = beriberi.]

1. Observations and reasoning
   - evidential relevance
   - role of systematic study or observation (versus anecdote) [M,A]
   - completeness of evidence [F,A]
   - robustness (agreement among different types of data) [M,C]
   - role of probability in inference [M]
   - alternative explanations [F,B]
   - verifiable information versus values [M]
2. Methods of investigation
   - controlled experiment (one variable) [B]
   - blind and double-blind studies [F]
   - statistical analysis of error [A]
   - replication and sample size [M,A,B]
   - correlation versus causation [A,B]
3. History and creativity
   - consilience with established evidence [F]
   - role of analogy, interdisciplinary thinking
   - conceptual change [M,A]
   - error and uncertainty
   - role of imagination and creative synthesis
4. The human context
   - spectrum of motivations for doing science
   - spectrum of human personalities in science [C]
5. Culture
   - role of cultural beliefs (ideology, religion, nationality, etc.)
   - role of gender bias [M]
   - role of racial or class bias
6. Social interactions among scientists
   - collaboration or competition among scientists
   - forms of persuasion
   - credibility [M,C,A,F]
   - peer review [M,A,F]
   - limits of alternative theoretical perspectives and criticism [F]
   - resolving disagreement
   - academic freedom [A]
7. Cognitive processes
   - confirmation bias/role of prior beliefs [M,F]
   - emotional versus evidence-based perceptions of risk [M,A,F]
8. Economics / funding
   - sources of funding [M,A]
   - personal conflict of interest [A]
9. Instrumentation & experimental practices
   - new instruments and their validation [F]
   - models and model organisms [B]
   - ethics of human subject experimentation [B]
10. Communication and transmission of knowledge
    - norms of handling scientific data [C]
    - nature of graphs [C]
    - credibility of various scientific journals and news media [M,C,A,F]
    - fraud or other forms of misconduct [C,A]
    - social responsibility of scientists

theory laden" with "scientists are creative"? Or: "scientific knowledge is durable" with "scientific knowledge is tentative"? Or: "science is empirical" with the influence of cultural and social ideas and values? In the dimensions framework (Table 2), items about, say, conceptual change and continuity—treated separately in other lists—are brought together. The introduction of new ideas is the converse of the persistence of old ones. Such groupings are possible because the ultimate goal is to inform students' interpretive skills, not to dictate whether science is (absolutely) either tentative or durable, conservative or creative (also see Clough, 2007). Students should be free, but also responsible, in interpreting the degree to which science may be empirical or culture-laden—at least for any *particular* scientific claim of concern.

In recent years, treatment of NOS in some places has yielded to more particular discussion of "science as a way of knowing" (or "how scientific knowledge is constructed," "scientific inquiry," or "the scientific worldview"), "scientific practices" or the "scientific enterprise," and "how science works" (AAAS, 2009; Board on Science Education [BOSE], 2010; Duschl, Schweingruber, & Shouse, 2007; Ford, 2008). For teachers and curriculum designers, any articulation of the vague and general phrase "nature of science" is surely welcome. The notion of Whole Science echoes and extends efforts to characterize NOS inclusively. For example, science includes a variety of investigations, such as documenting, describing and organizing natural phenomena, mapping causes (not always explaining them), or producing certain "effects," as well as building theories and models. Scientists exchange material demonstrations and samples, as well as textual arguments. They assemble grant proposals and secure resources, as well as present claims and evidence. In addition, epistemic practices include not only cognitive and evidential methods but also social interactions (Latour, 1987; Longino, 1990; Rudwick, 1985; Shapin, 1996; Solomon, 2001). Nor is science just a conceptual exercise: It includes lab skills and quasi-autonomous work on experimental systems (Franklin, 1986; Hacking, 1983; Kohler, 1994; Pickering, 1995; Rheinberger, 1997). Most important, perhaps, a Whole Science approach underscores the role of rendering the integrity of scientific practice, or how all the various NOS strands interact toward epistemic ends.

Consensus on NOS tenets (like scientific claims) has varied culturally and changed historically. Discussion of NOS is hardly new. It was especially in vogue in the 1950s and 1960s. However, views then were shaped by post–World War II celebration of science and technology and by Cold War politics (Reisch, 2005; Rudolph, 2003). Characterizations of NOS (e.g., in the Nature of Science Scale of 1968 or the Nature of Scientific Knowledge Scale of 1976), despite reflecting a consensus of philosophers and educators and despite having motivated historical case studies as vehicles for NOS lessons (Conant, 1957), seem almost embarrassing by today's standards: woefully positivist and overly optimistic in idealizing science. Similarly, the VNOS list today (with its references to theory-ladenness and social construction) reads much like a post–Positivist manifesto: self-consciously declaring its intellectual distance from the "failures" of envisioning science through a formulaic hypothetico-deductive method. The confidence in that former portrayal of NOS was apparently misplaced. History may thus heighten awareness of *philosophical hubris*. Ideally, in a forward-looking approach, one can buffer against metascientific mood swings and philosophical fads, while also readily accommodating growth in understanding from science studies. Accordingly, characterizations of NOS may be transformed from a prepackaged list of declarative tenets to the more enduring dimensions of reliability that frame epistemic discourse (Clough, 2007).

Discussions of NOS also confront one further and potentially quite profound ambiguity. The very phrase "nature of science" (especially with its reference to "nature") tends to connote some inherent, universal essence. By contrast, science is widely recognized as a "constructed," contingent human endeavor. Does NOS describe science as it *ought* to be or science as it *is*? Does one interpret it normatively or descriptively? Does NOS refer to *idealized* science or *real* science as actually found?[6] The contrasting views resonate with philosophers, on the one hand, and historians and sociologists, on the other, accounting for much interdisciplinary wrangling over what science truly "is" (Allchin, 2004a). The same normative/descriptive ambiguity arises equally with other expressions, such as "scientific practices" or "science as a way of knowing." The theme of reliability in decision making helps to clarify that science ultimately includes *both*. Ideal methods are not always realized in practice. Even if the ideals remain epistemic guides. Normative and descriptive views cannot be easily reconciled or collapsed into single defining statements. One must thus abandon the notion that NOS can be expressed in unambiguous declarative statements of the form, "science *is* X." Properly viewed, the concept of Whole Science accommodates the complementary, sometimes contrasting perspectives.

The Whole Science framework fosters a responsible balance between the foundations for reliability and the limits of science. Blind skepticism is no better than blind faith, as Poincaré once reminded us. Here, neither incautious scientism nor antiscience cynicism gain traction. The understanding is *functional* in that students need to develop analytical tools to assess both the promoters and critics of science. One strives to interpret the reliability of knowledge to inform our decisions, both as individuals and as a society.

FRAMING AN ASSESSMENT INSTRUMENT: FROM DECLARATIONS TO WELL-INFORMED ANALYSIS

Clarifying NOS provides a basis for addressing corresponding assessment strategies. For orientation, consider the several dozen paper-and-pencil instruments for assessing NOS that have been developed over the past half-century (Halloun & Hestenes, 1998; Lederman, 2007; Lederman, Wade, & Bell, 1998; Liang et al., 2008; and below). Recently, the most widely used instrument has been VNOS (Lederman et al., 2002). However, VNOS was intended for educational research only, not for summative assessment in a classroom context (pp. 511, 517). Indeed, the use of both free response and interviews (to ensure validity) make it labor intensive and impractical, even as a model, for any large-scale application. In general, the instruments are designed to probe several explicit, declarative tenets about NOS. VNOS, for example, asks students the following questions:

- "What is an experiment?"
- "After scientists have developed a scientific theory, does the theory ever change?"
- "Is there a difference between a scientific theory and a scientific law?"
- "Do scientists use their creativity and imagination during their investigations?"

These largely reflect the kind of statements that appear in other instruments, although student response is often measured on a Likert scale.

First, most NOS instruments probe *beliefs*. This is implicit wherever students are asked to agree or disagree with certain NOS statements. VNOS, too, refers several times to what

[6] This tension seems to explain the residual disagreement in the otherwise strong consensus developed by Osborne et al. (2003, p. 713). For example, the authors acknowledged a trenchant debate about "the extent to which cultural and subjective factors impinge on the practice of science" (p. 714). Similarly, only marginal consensus was achieved on the "empirical basis of scientific knowledge" and the "cumulative and revisionary nature of scientific knowledge" (p. 713). These areas are where one finds the most striking divergence between norms and practice—and those that matter most to context-free claims about scientific authority.

a student believes. However, the educational aim is to foster understanding, not certain *views*. The goal is knowledge of how science achieves (or does not achieve) reliability, not indoctrination into a set of beliefs. Indeed, a consistent criticism of older instruments has been the use of particular and arbitrary standards, also likely behind recurring deficits in validity (Lederman et al., 1998). By focusing on functional understanding and dimensions of reliability, as profiled above, all such types of instruments become inappropriate. Rather, in assessing functional understanding, one seeks a *well-informed analysis*. That is, learning will be indicated, not by agreement with prescribed statements, but by *the degree, both in breadth and depth, to which a student is informed about the factors that shape the reliability of scientific claims*.

There are other problems with using general declarative statements as benchmarks. Recently, many critics have noted that standard NOS items suffer from lack of context (Clough & Olson, 2008; Elby & Hammer, 2001; Ford, 2008; Osborne et al., 2003, pp. 712–713; Schwartz, Lederman & Crawford, 2004). That is, no qualifications are mentioned, or acknowledged as relevant in assessing student responses. In a Likert-scale analysis especially, the thoughtful student cannot respond, "well, it depends." As an example, consider the theme of tentativeness, central to NOS discourse for over four decades, in the context of global warming. The general claim that "science is tentative" can be (and has been) used unjustifiably in public discourse to dismiss the scientific consensus on the environmental dangers of anthropogenic climate change (Oreskes & Conway, 2010). Creationists have followed a similar tactic in alleging that Darwinians are "dogmatic," and hence fail to meet the scientific ideals of skepticism and open-mindedness (Allchin, 2001b). The declarative statement that "science is tentative," without context, can be grossly misleading. One needs skills to apply the knowledge properly—more than mere agreement or disagreement with certain tenets. Indeed, as these examples illustrate, declarative-type NOS may backfire in real-life contexts. Merely knowing that "science is tentative" does not solve the critical problem: how to interpret the degree of reliability of scientific claims relevant to key decisions. In the case of climate change, one might well argue, it is quite literally a matter of life or death.

The enduring theme of tentativeness is meant to capture the notion that while evidence and investigations can bolster our conclusions, they may also be limited in identifiable ways. One may measure this understanding concretely in context by how well a student assesses a case such as the revision in mammogram recommendations: is *this case* an example of tentativeness (unexpected but justified revision) or of political economics (limiting health care, with adverse consequences for women)? Ultimately, it is the fuller interpretation that reflects understanding of NOS, not any degree of agreement with a simple statement. *All* the standard "consensus" NOS tenets need such elaboration. The context is not merely background; it is the essential heart of the matter. That is, by articulating the context, one dissolves the role of any simple-minded statements as benchmarks or standards. In response, one cannot "just add context." The criticism about context exposes a fundamental weakness in the entire approach based on declarative tenets.

Nearly all the rhetoric about NOS (and history and philosophy of science) in science education seems to promise deeper analytical or critical thinking. But the targets of mainstream NOS assessment (and instruction) to date have largely been recall and comprehension of tenets. Skills in applying, analyzing, and/or evaluating NOS understanding are largely missing. In terms of Bloom's taxonomy, NOS is currently at the first two levels (remember and understand), while it purports (or aspires) to be at levels 3–5 (apply, analyze, and evaluate). Science educators need to live up to their rhetoric.—And in so doing, they need to set their sights well beyond a handful of simple NOS tenets. An effective alternative to the whole lineage of NOS instruments will address how to target directly the complexity that the simple statements cannot accommodate.

NOS assessment, therefore, should focus on *functional* understanding and analysis. Competence, not declarative knowledge, is the target (Pellegrino, Chudowsky, & Glaser, 2001, pp. 6–8). An instructor will want to know whether a student is prepared for the work at hand: assessing whether any *particular* claim—about the effectiveness of an antidepressant drug or the safety of a new food additive—is itself tentative or durable.

We may now address the core challenge. Based on the foregoing analysis of NOS (and the deficits in VNOS-type instruments), how might one frame an appropriate instrument for assessing functional NOS knowledge or NOS analytical skills? What features should it exhibit?

(1) *Authentic contexts*. In the spirit of authentic assessment (Nightingale et al., 1996), one ideally asks students to comment on cases similar to those they will encounter in life experiences. That is, students should straightforwardly demonstrate NOS understanding through a concrete example (Murcia & Schibeci, 1999; Nott & Wellington, 1998). Appropriate cases can easily be drawn from news reports in the local and national media—with the source of the report itself a possible object of comment—a strategy explored earlier by Norris and Phillips (1994), Glynn and Muth (1994), Korpan, Bisanz, and Bisanz (1997), Philips and Norris (1999), Norris, Phillips, and Korpan (2003), and Ford (2008). Cases might also be collected from science magazines and journals accessible to teachers or from history. Scenarios might be constructed for the student to adopt a specific perspective—possibly a researcher, policy-maker, consumer, or citizen, possibly reflecting a particular gender, ethnicity, or class.

An enduring tendency fostered by local institutional politics seems to be "teaching to the test." Often, this is viewed as a liability, taking valuable time away from "teaching." Authentic assessment, however, dissolves this problem. The "test" transparently embodies the goal. One should be clear and explicit about what is to be taught, as well as about how students may demonstrate what they have learned. Any assessment should be both meaningful and unsurprising.

(2) *Well-informed analysis*. Given the ultimate aim of functional scientific literacy, an ideal assessment will focus on analytical skills, not particular target conclusions. One aims to make the student's thinking visible (Pellegrino et al., 2001, p. 4). One is interested in depth and breadth of NOS knowledge. Reasoning and argumentation skills may be important at some level, but here the aim is to assess how well informed an analysis is. (1) Does the student recognize relevant factors shaping the reliability of claims? (2) Can they articulate their significance and how they shape specific interpretations of evidence? Perhaps also: (3) Can they identify information that would be important, but is missing—and where one would be likely to secure such information? Mere statements of general principles, without concrete examples or context (e.g., "all scientific knowledge is provisional"), are likely not sufficient to demonstrate fully developed analytical or critical skills.

Equally important, perhaps, an effective instrument will *not* ask students "What is your view?" or "What would you do?" Asking for a position statement, or worse, a justification, only tends to elicit post hoc rationalizations (Nickerson, 1998). An assessment must avoid values, ideology, and personal judgment—a recurring problem in earlier instruments. Rather, it should probe how well a student understands scientific practice in ways relevant to decision making (Kolstø, 2001, p. 307).

(3) *Adaptability to diagnostic, formative, or summative evaluation contexts*. The trend among professionals in assessment has shifted from summative to formative contexts. Still, politically, final evaluations seem to remain significant. An ideal assessment will thus fit

into a range of such contexts. VNOS and many others stipulate that they were designed for research only, not for practical purposes in classroom evaluation.

An appropriate format will also enable characterizing NOS knowledge on a fine scale along many dimensions (say, rating depth of understanding according to the dimensions noted in Table 2). That is, one should be able to develop an *NOS knowledge profile*, indicating level of competence in each category. Such detail can be valuable in formative assessment.

Of course, from the profile, one might *also* develop a numerical score for summative contexts. By weighting the various categories according to their relative importance or the number of concepts addressed, one can compile the individual's category scores into an *NOS knowledge index*. Such an index might give a crude, but still potentially useful indication of relative overall progress toward desired levels of performance.

One would certainly expect target achievement levels to be scaled to grade levels (or to institutionalized occasions for monitoring student progress). For example, AAAS partitions its benchmarks into four stages: K–2, 3–5, 6–8, and 9–12 (AAAS, 2009). In their framework, awareness that "reasoning can be distorted by strong feelings" is appropriately introduced in Grades 3–5 (p. 17). In Grades 9–12, students should know that "to avoid biased observations, scientific studies sometimes use observers who don't know what the results are 'supposed' to be" (p. 23). These are merely samples, to be validated by appropriate research, of how development in NOS understanding may be stratified, with educational goals associated with each level of education. (The proposal here does not try to resolve these details, but to provide a framework that can accommodate them.)

With concepts stratified in educational levels, assessments may well be reduced to simple scoring at each level: no visible progress; intermediate progress; acceptable level of performance. For example, a student may be able to articulate a critical lack of repeatability in a published experiment, but not notice that the study may be biased due to its commercial sponsorship and the researcher's institutional affiliation. Alternatively, one might institute, say, a simple five-point (or seven-point) scale. Such simple assessments may also help in selecting supplemental lessons and/or for evaluating the effectiveness of instructional programs.

(4) *Adaptability to single and mass and to local and large-scale comparative use*. Emerging expert perspectives indicate that ideally assessment will be implemented by teachers who have direct contact with the students, even while it generates information that can allow valid comparison across classrooms (Pellegrino et al., 2001, pp. 9, 13–14). Accordingly, one wants a simple scoring system (as noted above), which might additionally be accompanied by comments. Interviews—even if ideal from a research perspective—are not an option, given current teacher workloads and teacher–student ratios. At the same time, with an NOS knowledge index (above), administrators may assess groups of students at various levels (classrooms, teachers, schools, districts, states).

Of course, teachers who score the test will need to calibrate their own assessment judgments against a common standard. Review of sample scored responses is a possible approach, as developed by the Educational Testing Service in evaluating the free response section of AP exams and in the audit of AP syllabi.

(5) *Adaptability to performance-based assessment*. Another trend in educational assessment focuses on performance (as reflected in portfolios, say), with less reliance on single "high-stakes" tests. With a set of standardized practices, teachers may well be able to record levels of competence as they appear in classroom work. There need not be some universal "test day." With a well-articulated NOS knowledge profile, competency in individual dimensions of NOS can be achieved and recorded in piecemeal fashion. The prototype questions (appropriately adjusted) might equally well be regarded as projects to be completed

independently, not just as "tests" to be guarded from student awareness. Instruction and assessment merge.

(6) *Respect for relevant stakeholders*. An assessment instrument should be mindful of the perspectives, needs, and constraints of multiple stakeholders: (1) school teachers, (2) scientists, (3) scholars in science studies, (4) administrators and policy makers, and ultimately (5) students. Implementation and scoring should be practical, especially in terms of time, while still achieving the intended goals. The NOS content should be faithful to (not contradict) findings by historians, philosophers, and sociologists of science, as well as others who study science from a relatively remote vantage point. Researchers should also recognize expressions of their everyday practice and judgments. Finally, students should feel that the assessment allows them to convey what they have learned about NOS.

Prospective prototypes at the high school level meeting these six criteria (and again, modeled on the AP free response essay) are provided in Table 1 and the Appendix. This format or style of question I am calling KNOWS (Knowledge of the Nature Of Whole Science). Possible items for an itemized scoring rubric for each of the sample questions are indicated in Table 2. The prototypes seek a well-informed analysis, amenable to objective assessment. That is, they avoid soliciting personal views, positions, or judgments, and any consequent rationalizations. They seek transparency of thinking, allowing for evaluation of depth and breadth of knowledge. The style of question may thus be adapted to K–8 levels, echoing the pattern of analyzing concrete cases in a way relevant to a particular individual's perspective, while focusing on items from the NOS inventory appropriate to these grade levels.

Once again, these focus exclusively on NOS. One could well imagine a more expansive assessment where questions asked also for problem solving or for evaluation of a particular set of evidence. The samples here are designed specifically to focus on the challenge of assessing functional knowledge of NOS.

ALTERNATIVE AND INTERMEDIATE ASSESSMENTS

Alternatives to KNOWS are also possible. For example, the form of assessment may be less standardized, while still effective. Consider a teacher who uses a series of historical episodes, problem-solving cases, news stories, reflective lab activities, etc., to develop NOS knowledge, with lessons distributed throughout the school year. The ensemble of cases may be unique to the particular teacher or classroom. Here, students may synthesize the lessons from the individual cases in a capstone exercise or final assessment. A set of sample questions that I used for a high school biology course is presented in Table 3 (1981–1985, students age 15). These questions target relatively basic concepts: controlled experiment, measurement, making connections, factors in scientific belief, and the process of inquiry generally. Here, the relevant examples from the students' experience are conveniently recalled and listed. Students are also allowed some choice, so as to highlight their strongest knowledge. The responses are free response essays. Assessment of quality is based on three criteria, identified for the students:

1. *Organization*. Unity. See that everything supports one central idea.
2. *Breadth*. Completeness. Relate as many different ideas as possible together.
3. *Depth*. Proficiency. Convey your ideas effectively by using specifics.

Such criteria are widely applicable for assessing how well informed someone is in any particular domain, although scoring relies on the individual teacher's judgment. This approach also empowers individual teachers in a local setting. It accommodates them selecting case

TABLE 3 Sample NOS Questions

1. Discuss the role of **controlled experiments** using 4 examples:
   a. you and enzyme catalysis
   b. you and the effectiveness of exercise
   c. you and spontaneous generation
   d. you and nutrient indicators
   e. Mendel and inheritance
   f. Gause and the outcome of competition
   g. Moody and near-death experiences
2. Discuss how **making connections** is important in reaching conclusions using 4 examples:
   a. Darwin: domestic breeding and adaptation
   b. Darwin: Malthus's essay on population and nature
   c. Morgan: inheritance and chromosomes
   d. you: circulation in humans and squids
   e. Eijkman: beriberi and chicken diets
   f. Ingen-Housz: wind or sun and transport in plants
   g. Paul Broca: speech impediments
   h. enzyme deficiencies and inheritance
   i. associative learning (Pavlov, Skinner)
3. Discuss the significance of **measurement** in experimentation using 4 examples:
   a. Mendel: F_2 generation pea offspring **OR** you: fruit flies
   b. Pavlov: saliva secretion
   c. Harvey: blood flow through the heart
   d. Watson & Crick: X-ray photographs of DNA crystals
   e. you: map of a sugar maple forest
4. [Honors] Discuss the factors influencing **scientific belief** in three of the following controversies:
   a. genetic material: DNA or protein?
   b. human evolution: brain size or posture?
   c. human development: preformation or epigenesis?
   d. evolution: progressionism or natural selection?
   e. Cuvier's brain size: large or normal?
5. Discuss how three of the following are important to the **process of inquiry**:
   a. microscope/electron microscope
   b. models (any example from the video)
   c. pea plants (for Mendel) **OR** fruit flies (for Morgan)
   d. the social position of Darwin, Lamarck or Mendel **OR** the budget of the National Science Foundation

Taken from the author's final examinations for a high school biology course (students age 15) from 1981 to 1985.

studies appropriate to their students and that capitalize on the teacher's personal backgrounds and strengths. Second, it ensures that the teacher will be deeply familiar with the examples that a student may use in making comparisons or supporting general statements. While not designed for standardization across classrooms, this alternative certainly allows local assessment. It may also provide an intermediate assessment, synthesizing lessons before a KNOWS-style evaluation where a student analyzes unfamiliar new cases.

Assessments targeted to learning from individual case studies or other short-term activities may also be appropriate. In such cases, the assessment may also be an occasion to

TABLE 4
Sample Assessment for Reviewing NOS in a Case Study

What does the case of "Christian Eijkman & the Cause of Beriberi" show about the following aspects of doing biology?

- The role of chance or accident
- Theoretical perspectives in interpreting data
- The distinction between causation and correlation
- Growth of knowledge through small cumulative additions versus through major conceptual reinterpretations
- The role of individual versus groups in making a discovery
- Scientific communication and communities of researchers
- The cultural and economic contexts of science

help consolidate lessons by fostering further explicit reflection on the NOS features that student have encountered. Table 4 presents one sample for a historical case study (Hagen et al., 1996, p. 126). These questions resemble more closely the case-oriented design of KNOWS, but with the NOS elements already identified (or recalled) for students.

FROM ASSESSING TO TEACHING AGAIN

Assessment formats are ideally modeled in previous classroom activities and exercises, allowing teachers the opportunity to guide students toward appropriate target performance (Pellegrino et al., 2001, p.8). One may thus work "backward" from the form of the KNOWS assessment to the context of teaching, sketching horizons beyond the central focus of this paper. Here, the target implies introducing students to samples of Whole Science, or complex case studies, not merely vignettes or anecdotes. A focus on case studies reflects, of course, the earlier views of James Bryant Conant (1957), as well as recent consensus about teaching NOS. "Ideas about science," Osborne et al. (2003) concluded, are

> perhaps best. . . addressed through sets of well-chosen case studies of either a historical or contemporary nature and by more explicit reflection and discussion of science and its nature—an aspect that should emerge naturally from the process of scientific inquiry that is a normal feature of much classroom practice. (p. 716)

Accordingly, teaching needs to weave together experimental/material, conceptual, and social strands of doing science. Scientific practices, science as a way of knowing, and inquiry as curricular goals reflect inseparable parts of the scientific enterprise, even if they have independent contexts or meanings for educators (BOSE, 2010; Duschl et al., 2007). In addition, the NOS elements should ideally be coupled with, and rendered as arising from, content and inquiry. Whole Science may well designate a synthesis not only of relevant NOS elements, but also of scientific process and product in an educational context. Whole Science in a classroom embraces content, process of science skills, and broad-ranging NOS analysis.

The key element in all cases—whether contemporary, historical, or student-directed—is explicit reflection on how we know, how science works (and how it sometimes does not), and how methods and practice matter (Akerson, Abd-El-Khalick, & Lederman, 2000; Craven, 2002; Khishfe & Abd-El-Khalick, 2002; Scharmann, Smith, James & Jensen, 2005; Schwartz et al., 2004, Seker & Welsh, 2005). One cannot expect prepared lessons merely illustrating NOS tenets (McComas, 2008) to be effective. Reflection implies a constructivist teaching style—here, applying strategies common in teaching science to teaching NOS as well. Thus, science educators now need to profile more fully the anomalies or discrepant events (or other prompts of cognitive dissonance) that engage students in deeper NOS awareness and thinking. That is, educators need to focus more now on *problematizing* NOS in case examples (Allchin, 2010b). The trend toward problem-based and case-based learning in an open-ended format may well prove fruitful in informing NOS education (Cliff & Nesbitt, 2005; Lunberg, Levin, & Harrington, 1999).

One may be encouraged that children, even at the age they enter school, seem ready to engage in such explicit reflection (Duschl et al., 2007). While developmentally appropriate learning progressions have yet to be established, a Whole Science approach would seem fruitful throughout K-12 (with assessment scaled to each level). One challenge for teachers is to probe the capacities of students, working from a deep inventory of NOS dimensions (Table 2).

As noted, student inquiry—when pursued in an authentic style of uncertain outcomes and coupled with epistemic reflection—is one of three contexts for developing NOS understanding. Such an approach seems to fit comfortably within conventional teaching styles and resources. Such instruction can provide vivid experience for analytical skills in many NOS dimensions (in Table 2, dimensions 1–3, and parts of 6, 7, 9, and 10) (for a noteworthy example, see Lehrer, Schauble, & Lucas, 2008).

Contemporary case studies also provide NOS-teaching opportunities (Dimopoulos & Koulaidis, 2003; Elliot, 2006; Hodson, Kwan, & Yung, 2008; Jarman & McClune, 2007; McClune & Jarman, 2010; Wellington, 1991; Wong, Hodson, Kwan, & Yung, 2008). Indeed, a case-based approach has been adopted in a few special courses (Millar, 2000) and several textbooks (at the secondary level: Leonard, Penick, & Speziale, 1998/2008, and American Chemical Society, 2006; and at the college level: Postlethwait & Hopson, 2003, and Schwartz et al., 1997; although not always permeated with explicit NOS reflection). Given the emphasis in KNOWS on cases in the news, one might be tempted to see such cases as optimal and to use them almost exclusively. Yet modern cases may also be problematic in instruction. Emotion-laden content can distort perceptions and learning. The science is not yet resolved, fostering disillusionment and confusion (J. Thomas, 2000). Moreover, the key information on the internal mechanisms of science is largely unavailable through popular media (Dimopoulos & Koulaidis, 2003, p. 248). Most important, contemporary cases lack a clear solution by which to judge one's emerging problem-solving or interpretive efforts, critical in an instructional context. How can one learn what leads to reliable knowledge if one only addresses unresolved controversies (a problem exemplified in Collins & Pinch, 1993)? Understanding how an episode ultimately unfolded provides a benchmark for students to evaluate and adjust their own maturing skills. Reflection on historical outcomes contributes to self-regulation. Students also need the freedom to fail while they practice new analytical skills.

In addition, history seems essential for conveying lessons about NOS that rely on retrospect or larger perspectives. These include, most notably: the ways scientists can err, the nature of deep conceptual change and uncertainty, and the role of cultural context and potential bias in scientific ideas. In particular, the provisional nature of scientific knowledge, or "tentativeness," has been a prominent NOS learning goal for decades. To teach this, one needs examples of real, profound, and unanticipated conceptual change. To enable informative contrast of a reasonable "before" with the unexpected "after," a student problem-solving episode must be past and outcomes amenable to analysis. In a similar way, to appreciate gender or racial bias or other ways that cultural perspectives may sometimes become blindly naturalized in science, one must be at a relatively remote vantage point, to see the culture as culture. History and historical perspective are indispensable for complete NOS lessons. Thus, for example, a student's interpretation of the Climategate case (Prototype Question #2) would be greatly informed by earlier study of Millikan's oil-drop data (Franklin, 1986, chap. 5; Holton, 1978) or Mendel's data on inheritance in peas (Franklin, Edwards, Fairbanks, Hartl, & Seidenfeld, 2008); the "tricks" used to map chromosomes (Wimsatt, 2007, pp. 94–132) or a Mercator-style map (Turnbull, 1993); or the publication politics of Newton (White, 1997), Lavoisier, or geologist Roderick Murchison (Rudwick, 1985). The mammogram case (Prototype Question #1) would be informed by learning about how more data dramatically altered theories about the causes of pellagra or beriberi (Carpenter, 2000; Kraut, 2003) or the assessments of the risks of thalidomide or genetic engineering (Fredrickson, 2001; Hindmarsh & Gottweis, 2005; Stephens & Brynner, 2001). Facilitated communication (Prototype Question #3) brings to mind the earlier cases of Clever Hans and Uri Geller, or the recurring pitfalls of research on spiritualism and the "paranormal" (Gardner, 1990; Lyons, 2009, chap. 4–5). Such deep lessons cannot be learned through the students' own short-term inquiry experiences.

While a role for history in NOS learning has been acknowledged for some time (Conant, 1957; Hodson, 2008, pp. 155–158), the KNOWS assessment format helps clarify the objectives of historically situated inquiry and thus how to assemble cases more effectively. Most notably, although the story is past, the history must be made "present" again. One must forsake ready-made science and restore science-in-the-making (Latour,1987). Students must experience a historically situated perspective, blind to the outcome, akin to the uncertainty in modern cases that one hopes to inform. For example, to learn about conceptual change, a student ideally engages in and experiences unforeseen change. Knowing the "right" answer in advance destroys the lesson, just as a spoiler ruins a mystery or suspense thriller (or any constructivist lesson). A case with open-ended problems, not rationally reconstructed, is essential (Allchin, 2002). Respect for historical context is not just the concern of some fussy historians; it is central to the NOS aims of using history at all (Allchin, 2004c, 2006). Many case studies designed to highlight Whole Science, situating students in historical scenarios and including NOS problems, have been developed and collected in recent years (Allchin and Minnesota Teachers, 1997; Allchin, 2010a, 2010b; Hagen, Allchin, & Singer, 1996; HIPST, 2010).

While complex case studies are desired endpoints, the notion of Whole Science supports development of NOS analytical skills through almost any fragment of authentic science. That is, even a teacher not yet deeply experienced in NOS can introduce students to a case, sharing with them a reflective posture of epistemic questioning, possibly learning together. (The bare prerequisite is awareness of the diversity of epistemic dimensions and openness to exploring whatever NOS problems a case may present.) A taxonomy of NOS dimensions (Table 2) can thus be a valuable reference. Case studies and other encounters with NOS will each highlight different constellations of NOS elements. Hence, an ensemble of cases should ultimately complement one another to convey a complete profile of NOS. Teachers may be guided by a full inventory of significant NOS features (Table 2; or Millar, 2000, Table 3). That is, the inventory may function as a curricular checklist for reviewing completeness of NOS across a set of various case studies (whether historical or contemporary) and other classroom investigations (e.g., Table 3).

Clough and Olson (2008) note that "while criticisms of common pencil-and-paper NOS assessments are well placed, attention is needed to creating viable, valid and reliable assessments that will encourage teachers to accurately and consistently implement developmentally appropriate NOS instruction" (p. 145). I hope that the prototypes sketched here help us move toward that goal. Many fear teachers teaching to the test. If the test is appropriately designed, however, it will transparently embody what needs to be taught. Should one not endorse (or even celebrate) teaching to such a "test"?

The author appreciates supportive discourse from Agustín Adúriz-Bravo, Rick Duschl, Fred Finley, Ami Friedman, Dietmar Höttecke, Cathrine Klassen, Andreas Henke, David Rudge, Cibelle Celestino Silva, Peter Taylor, participants of the 2009 SHiPS Teachers Workshop, the session on evaluating NOS at the 2009 SPSP meeting, the 2010 HIPST Conference, the 2010 New England Workshop for Science and Social Change, and the 8th International Conference on History of Science in Science Education, as well as two anonymous reviewers and the Editor.

APPENDIX: PROTOTYPE QUESTIONS FOR ASSESSING NOS, BASED ON CONTEMPORARY NEWS OR HISTORICAL CASES

"Climategate," Nov. 2009

In November 2009, the author of "The Air Vent," a blog critical of global warming claims, received an anonymous note:

> *We feel that climate science is, in the current situation, too important to be kept under wraps. We hereby release a random selection of correspondence, code and documents. Hopefully it will give some insight into the science and the people behind it.*

Included was a link to a file that contained over 1000 e-mails and other material apparently hacked from a server at the Climate Research Unit of the University of East Anglia in Britain. In the e-mails, climatologist Philip Jones, a leading member of the International Panel on Climate Change (IPCC), included comments about scuttling efforts to release data under a Freedom of Information Act request, a "trick" he used in graphing data and strategies to limit the publication of critics in peer-reviewed journals. James Delingpole, in a blog for England's *Telegraph*, promptly dubbed it "Climategate." The news sparked a flurry of comments by skeptics who presented this as proof of their repeated claims about fraud, collusion, and conspiracy in climate science. Within a week, the term "Climategate" could be found over 9 million times on the Internet.

While sitting at lunch with two coworkers, one mentions how the case just proves that global warming is a joke. The other, an avid environmentalist, contends that scientists do not do things like that, indicating that the posted documents themselves are probably fraudulent. Amid mutual accusations of being misinformed and biased, they ask you set the other straight. Comment on a what a well-informed interpretation of events in this case might indicate about the conduct of science and the evidence for climate change.

Resource documents

- *New York Times* (Nov. 21, 2009): Andrew Revkin, "Hacked E-mail is New Fodder for Climate Dispute"
  http://www.nytimes.com/2009/11/21/science/earth/21climate.html
- James Delingpole Blog, *Telegraph.co.uk* (Nov. 20, 2009) "Climategate: the final nail in the coffin of 'Anthropogenic Global Warming'?"
  http://blogs.telegraph.co.uk/news/jamesdelingpole/100017393/climategate-the-final-nail-in-the-coffin-of-anthropogenic-global-warming/
- *Nature* news brief (Nov, 26, 2009): "Storm clouds gather over leaked climate e-mails"

Facilitated Communication of Coma Patient, Nov. 2009

In November 2009, National Public Radio reported:

> Twenty-three years ago, a Belgian car-crash victim [Rom Houben] was diagnosed as being in a vegetative state. But doctors now say he appears to have been conscious the whole time. The man is now communicating using a special touchscreen. Neurologist Steven Laureys, who leads the Coma Science Group at the University of Liege in Belgium, says people in noncommunicative states are misdiagnosed up to 40 percent of the time.

Other major news media, including CNN, Fox News, and MSNBC carried the remarkable story.

Others were skeptical. In an Internet blog, Steven Novella, a neurologist at Yale University, acknowledged that Laureys had "impressive expertise in coma and disorders of consciousness," yet characterized the patient's observed responses as "bogus facilitated communication"—namely, that "the facilitator [who helped the patient to spell out words on the touch screen] is doing the communicating, not Houben."

The head of the Belgian hospital enlists your help as an external reviewer and provides a modest budget for additional research on this important question. What is your initial perspective on this case and how would you propose resolving the issue? If you propose further investigation, describe your experimental design and how it will help determine the legitimacy of the claims more decisively. Alternatively, explain how the available information is sufficient to provide a conclusive answer.

Resource documents

- Nov. 24, 2009 NPR story
  http://www.npr.org/templates/story/story.php?storyId=120784397
- CNN story
  http://www.cnn.com/2009/HEALTH/11/24/coma.man.belgium/index.html
- Steve Novella blog, Nov. 2009
  http://www.theness.com/neurologicablog/?p=1286#more-1286

Autism and the Measles Vaccine, Feb. 2010

In January 2010, the prestigious British medical journal *Lancet* formally retracted a 1998 article that linked a widely used measles–mumps–rubella vaccine to autism, a serious disorder of the nervous system. Studies criticizing the original report were published almost every year until 2004, when 10 of the 13 original coauthors withdrew their support for it. Meanwhile, amid concerns about the widely reported risks, many parents decided not to have their children vaccinated. In 2006, a 13-year-old boy became the first person in 14 years to die from measles in Britain. In 2008, the British government estimated that less than half the children in London and 3 million children nationally had not had the recommended two doses of the vaccine. Charges of misconduct against the lead researcher were filed in 2006, and it now appears that he received substantial funds to sponsor research that would support several patients' legal actions, although he did not report these. The researcher continues to claim no wrong-doing.

You have been appointed as a citizen-member of a panel to review this case and make recommendations: both about specific actions regarding the lead researcher and about general guidelines for reviewing and publishing scientific research and press coverage of health issues. If certain actions seem to have been warranted at certain dates (and were not done), provide information that justifies your view, and indicate how the system of review in science and/or science journalism could be changed accordingly to initiate action at an appropriate time.

Resource documents

- Overview: "Retracting a Medical Journal's Autism Study" (*New York Times*, Feb. 8, 2010)
  well.blogs.nytimes.com/2010/02/08/did-the-media-inflame-the-vaccine-autism-link/
- Time line: "The 12-year controversy over a vaccine" (*Telegraph*, Jan. 29, 2010)
  www.telegraph.co.uk/health/healthnews/7091683/The-12-year-controversy-over-a-vaccine.html
- "MMR-autism link doctor Andrew Wakefield defends conduct at GMC hearing" (*Telegraph*, Mar. 27, 2008)
  www.telegraph.co.uk/news/uknews/1582980/MMR-autism-link-doctor-Andrew-Wakefield-defends-conduct-at-GMC-hearing.html
- "A Short Form FAQ about the Wakefield GMC Case" (Age of Autism Web site, Jan. 28, 2010)
  www.ageofautism.com/2010/01/a-short-form-faq-about-the-wakefield-gmc-case.html

Beriberi in Java, 1896

It is 1895. You are the governor in Java, a large island of the Dutch territory in Southeast Asia, important in trading spices, coffee, sugar, and other products. Recently, there has been a marked rise in the incidence of beriberi, a nervous degenerative disease, sometimes causing death. It is prominent in prisons and insane asylums and among the fleet crews that load the Dutch ships. (In Japan, it seems prevalent in the Army and Navy.) Dutch physician Christian Eijkman believes that the disease is caused by a bacterium in rice that, once in the stomach, transforms starch into a neurotoxin. An antitoxin, he also claims, is present in the coating to the rice (which is "polished" off to yield white rice). Eijkman has just completed a joint study of the incidence of beriberi in Java prisons. The results are presented in the accompanying table.

1. Identify at least one other possible explanation for beriberi, consistent with these results. Describe the design of an investigation to help determine which explanation is more justified.
2. Given the possibility of alternative explanations, what action, if any, is warranted by the results in the interest of public health?

Resource documents

- Table of Eijkman's results
- Brief discussion of other theories about beriberi in 1895

REFERENCES

Akerson, V. L., Abd-EI-Khalick, F. S., & Lederman, N. G. (2000). The influence of a reflective activity-based approach on elementary teachers' conceptions of the nature of science. Journal of Research in Science Teaching, 37, 295–317.

Allchin, D. (1999). Do we see through a social microscope? Credibility as a vicarious selector. Philosophy of Science, 60(Proceedings), S287-S298.

Allchin, D. (2001a). Error types. Perspectives on Science, 9, 38 – 59.

Allchin, D. (2001b). The emperor's old clothes. American Biology Teacher, 63, 625 – 636.

Allchin, D. (2002). How *not* to teach history in science. The Pantaneto Forum, 7 (July). Retrieved December 17, 2010, from www.pantaneto.co.uk/issue7/allchin.htm.

Allchin, D. (2004a). Should the sociology of science be rated X? Science Education, 88, 934–946.

Allchin, D. (2004b). Error and the nature of science. ActionBioscience. Retrieved September 20, 2010, from www.actionbioscience.org/education/allchin2.html.

Allchin, D. (2004c). Pseudohistory and pseudoscience. Science & Education, 13, 179–195.

Allchin, D. (2006). Why respect for history—and historical error—matters. Science & Education, 15, 91–111.

Allchin, D. (Ed.). (2010a). Teaching science through history (2nd. ed.; CD-rom). Also retrieved September 20, 2010, from ships.umn.edu.

Allchin, D. (2010b). From rhetoric to resources: New historical problem-based case studies for nature of science education. Paper presented at History and Philosophy in Science Teaching Conference, Kaiserslautern, Germany, March, 2010.

Allchin, D. (Ed.), and Minnesota Teachers. (1997). Teaching history and nature of science: Curriculum modules. Minneapolis, MN: The Bakken Museum and Library.

American Association for the Advancement of Science. (2009). Benchmarks for scientific literacy (revised). Retrieved September 20, 2010, from www.project2061.org/publications/bsl/online/index.php.

American Chemical Society. (2006). Chemistry in the community (5th ed.). New York: W. H. Freeman.

Bill & Melinda Gates Foundation. (2010). Primary sources: America's teachers on America's schools. New York: Scholastic.

Board on Science Education, U.S. National Academies of Sciences. (2010). Draft of the conceptual framework for the new science education standards. Washington, DC: National Academies. Retrieved July 30, 2010, from www7.nationalacademies.org/bose/Standards_Framework_Preliminary_Public_Draft.pdf.

Carpenter, K. (2000). Beriberi, white rice and vitamin B. Berkeley: University of California Press.

Cliff, W. H., & Nesbitt, L. M. (2005). An open and shut case? Contrasting approaches to case study design. Journal of College Science Teaching, 34(4), 14–17.

Clough, M. P. (2007). Teaching the nature of science to secondary and post-secondary students: questions rather than tenets. The Pantaneto Forum, 25(January). Retrieved September 20, 2010, from www.pantaneto.co.uk/issue25/clough.htm.

Clough, M. P., & Olson, J. K. (2008). Teaching and assessing the nature of science: An introduction. Science & Education, 17, 143–145.

Collins, H., & Pinch, T. (1993). The golem. Cambridge, England: Cambridge University Press.

Conant, J. B. (Ed.). (1957). Harvard case histories in experimental science. Cambridge, MA: Harvard University Press.

Craven, J. A. (2002). Assessing explicit and tacit conceptions of the nature of science among preservice elementary teachers. International Journal of Science Education, 24, 785–802.

Dimopoulos, K., & Koulaidis, V. (2003). Science and technology education for citizenship: The potential role of the press. Science Education, 87, 241–256.

Duschl, R. A., Schweingruber, H. A., & Shouse, A. W. (Eds.). (2007). Taking science to school. Washington, DC: National Academies Press.

Elby, A., & Hammer, D. (2001). On the substance of a sophisticated epistemology. Science Education, 85, 554–567.

Elliot, P. (2006). Reviewing newspaper articles as a technique for enhancing the scientific literacy of student–teachers. International Journal of Science Education, 28, 1245–1265.

Ford, M. (2008). 'Grasp of practice'as a reasoning resource for inquiry and nature of science understanding. Science & Education, 17, 147–177.

Franklin, A. (1986), The neglect of experiment. Cambridge, England: Cambridge University Press.

Franklin, A., Edwards, A. W. F., Fairbanks, D., Hartl, D., & Seidenfeld, T. (2008). Ending the Mendel–Fisher controversy. Pittsburgh, PA: Pittsburgh University Press.

Fredrickson, D. S. (2001). The recombinant DNA controversy, a memoir: Science, politics, and the public interest, 1974–1981. Washington, DC: ASM Press.

Friedman, S. M., Dunwoody, S., & Rogers, C.L. (1999). Communicating uncertainty. Mahwah, NJ: Erlbaum.

Gaon, S., & Norris, S. P. (2001). The undecidable grounds of scientific expertise: science education and the limits of intellectual independence. Journal of Philosophy of Education, 35, 187–201.

Gardner, M. (1990). Science: Good, bad and bogus. Amherst, NY: Prometheus Books.

Glynn, S. M., & Muth, K. D. (1994). Reading and writing to learn science: Achieving scientific literacy. Journal of Research in Science Teaching, 9, 1057–1069.

Goldman, A. I. (1999). Knowledge in a social world. Oxford, England: Oxford University Press.

Goldman, A. I. (2002). Pathways to knowledge: Public and private. Oxford, England: Oxford University Press.

Guinta, C. J. (2001). Using history to teach scientific method: the role of errors. Journal of Chemical Education, 78, 623–627.
Hacking, I. (1983). Representing and intervening. Cambridge, England: Cambridge University Press.
Hagen, J. B., Allchin, D., & Singer, F. (1996). Doing biology. Glenview, IL: Harper Collins. Also available at doingbiology.net.
Halloun, I., & Hestenes, D. (1998). Interpreting VASS dimensions and profiles. Science & Education, 7, 553–577.
Hardwig, J. (1991). The role of trust in knowledge. Journal of Philosophy, 88, 693–708.
Hindmarsh, R. & Gottweis, H. (Eds.). (2005). Recombinant regulation: The Asilomar legacy 30 years on. Science as Culture, 14(4).
HIPST. (2010). History and philosophy in science teaching: HIPST developed cases. Retrieved December 17, 2010, from hipstwiki.wetpaint.com/page/hipst+developed+cases.
Hodson, D. (2008). Towards scientific literacy. Rotterdam, The Netherlands: SensePublishers.
Holton, G. (1978). Subelectrons, presuppositions and the Millikan-Ehrenhaft dispute. Historical Studies in the Physical Sciences, 9, 166–224. Reprinted in The scientific imagination (pp. 25–83). Cambridge, England: Cambridge University Press (1978).
Jarman, R., & McClune, B. (2007). Developing scientific literacy: Using news media in the classroom. Maidenhead, England: Open University Press.
Khishfe, R., & Abd-El-Khalick, F. (2002). Influence of explicit and reflective versus implicit inquiry-oriented instruction on sixth graders' views of nature of science. Journal of Research in Science Teaching, 39, 551–578.
King, P. M., & Kitchener, K. S. (1994). Developing reflective judgment. San Francisco: Jossey-Bass.
Kohler, R. E. (1994). Lords of the fly. Chicago: University of Chicago Press.
Kolstø, S. D. (2001). Scientific literacy for citizenship: Tools for dealing with the science dimension of controversial socioscientific issues. Science Education, 85, 291–300.
Korpan, C. A., Bisanz, G. L., & Bisanz, J. (1997). Assessing scientific literacy in science: Evaluation of scientific news briefs. Science Education, 81, 515–532.
Krajik, J. S., & Sutherland, L. M. (2010). Supporting students in developing literacy in science. Science, 328, 456–459.
Kraut, A. M. (2003). Goldberger's war: The life and work of a public health crusader. New York: Hill & Wang.
Kuhn, D., Cheney, R., & Weinstock, M. (2001). The development of epistemological understanding. Cognitive Development, 15, 309–328.
Latour, B. (1987). Science in action. Cambridge, MA: Harvard University Press.
Lederman, N. G. (2007). Nature of science: Past, present, and future. In S. K. Abell & N. G. Lederman (Eds.), Handbook of research on science education (pp. 831–880). Mahwah, NJ: Erlbaum.
Lederman, N. G., Abd-El-Khalick, F., Bell, R. L., & Schwartz, R. (2002). Views of nature of science questionnaire: Toward valid and meaningful assessment of learner's conceptions of nature of science. Journal of Research in Science Teaching, 39, 497–521.
Lederman, N. G., Wade, P., & Bell, R. L. (1998). Assessing understanding of the nature of science. Science & Education, 7, 595–615.
Lehrer, R., Schauble, L., & Lucas, D. (2008). Supporting development of the epistemology of inquiry. Cognitive Development, 23, 512–529.
Leonard, W. H., Penick, J. E., & Speziale, B. (1998/2008). BioComm: Biology in a community context. Cincinnati, OH: Thomson; Armonk, NY: It's About Time, Herff Jones Education Division.
Liang, L. L, Chen, S., Chen, X., Kaya, O. N., Adams, A. D., Macklin, M., & Ebenezer, J. (2008). Assessing preservice elementary teachers' views on the nature of scientific knowledge: A dual-response instrument. Asia-Pacific Forum on Science, Learning and Teaching, 9(1), 1. Reprinted as "Preservice teachers' views about nature of scientific knowledge development: An international collaborative study." International Journal of Science and Mathematics Education, 7, 987–1012.
Longino, H. (1990). Science as social knowledge. Princeton, NJ: Princeton University Press.
Lunberg, M. A., Levin, B. B., & Harrington, H. L. (1999). Who learns what from cases and how? Mahwah, NJ: LEA.
Lyons, S. L. (2009). Species, serpents, spirits and skulls. Albany: State University of New York Press.
McClune, B,, & Jarman, R. (2010). Critical reading of science-based news reports: Establishing a knowledge, skills and attitudes framework. International Journal of Science Education, 32, 727–752.
McComas, W. C. (2008). Seeking historical examples to illustrate key aspects of the nature of science. Science & Education, 17, 1249–1263.
McComas, W. F., & Olson, J. K. (1998). The nature of science in international science education standards documents. In W. F. McComas (Ed.), The nature of science in science education: Rationales and strategies (pp. 41–52). Dordrecht, The Netherlands: Kluwer.

Millar, R. (2000). Science for public understanding: Developing a new course for 16–18 year old students. Critical Studies in Education, 41, 201–214.
Millar, R., & Osborne, J. (1998). Beyond 2000: Science education for the future. London: King's College.
Murcia, K., & Schibeci, R. (1999). Primary student teachers' conceptions of the nature of science. International Journal of Science Education, 21, 1123–1140.
National Research Council. (1996). National Science Education Standards. Washington, DC: National Academy Press.
Nickerson, R. S. 1998. Confirmation bias: A ubiquitous phenomenon in many guises. Review of General Psychology, 2, 175–220.
Nightingale, P., Te Wiata, I., Toohey, S., Ryan, G., Hughes, C., & Magin, D. (Eds). (1996). Assessing learning in universities. Sydney, NSW, Australia: UNSW Press.
Norris, S. P. (1995). Learning to live with scientific expertise: Toward a theory of intellectual communalism for guiding science teaching. Science Education, 79, 201–217.
Norris, S. P. (1997). Intellectual independence for nonscientists and other content-transcendent goals of science education. Science Education, 81, 239–258.
Norris, S. P., & Phillips, L. M. (1994). Interpreting pragmatic meaning when reading popular reports of science. Journal of Research in Science Teaching, 31, 947–967.
Norris, S. L., Phillips, L. M., & Korpan, C. A. (2003). University students' interpretation of media reports of science and its relationship to background knowledge, interest, and reading difficulty. Public Understanding of Science, 12, 123–145.
Nott, M., & Wellington, J. (1998). Eliciting, interpreting and developing teachers' understandings of the nature of science. Science & Education, 7, 579–594.
OECD. (2009). PISA 2009 assessment framework. Paris: Author. Retrieved October 6, 2010, from http://www.oecd.org/document/44/0,3343,en_2649_35845621_44455276_1_1_1_1,00.html.
Oreskes, N., & Conway, E. M. (2010). The merchants of doubt: How a handful of scientists obscured the truth on issues from tobacco smoke to global warming. New York: Bloomsbury Press.
Osborne, J. (2007). Science education for the twenty-first century. Eurasian Journal of Mathematics, Science and Technology Education, 3, 173–184.
Osborne, J., Collins, S., Ratcliffe, M., Millar, R., & Duschl, R. (2003). What "ideas-about-science" should be taught in school science? A Delphi study of the expert community. Journal of Research in Science Teaching, 40, 692–720.
Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). Knowing what students know: The science and design of educational assessment. Washington, DC: National Academies Press.
Phillips, L. M., & Norris, S. P. (1999). Interpreting popular reports of science: What happens when the reader's world meets the world on paper? International Journal of Science Education, 21, 317–327.
Pickering, A. (1995). The mangle of practice. Chicago: University of Chicago Press.
Postlethwait, J. H., & Hopson, J. L. (2003). Explore life. Pacific Grove, CA: Brooks/Cole.
Reisch, G. (2005). How the cold war transformed philosophy of science to the icy slopes of logic. Cambridge, England: Cambridge University Press.
Rheinberger, H. J. (1997). Toward a history of epistemic things. Stanford, CA: Stanford University Press.
Rudolph, J. L. (2000). Reconsidering the 'nature of science' as a curriculum component. Journal of Curriculum Studies, 32, 403–419.
Rudolph, J. L. (2003). Portraying epistemology: School science in historical context. Science Education, 87, 64–79.
Rudwick, M. J. (1985). The great Devonian controversy. Chicago: University of Chicago Press.
Rutherford, J. F., & Ahlgren, A. (1990). Science for all Americans. New York: Oxford University Press.
Scharmann, L. C., Smith, M. U., James, M. C., & Jensen, M. (2005). Explicit reflective nature of science instruction: evolution, intelligent design, and umbrellaology. Journal of Science Teacher Education, 16, 27–41.
Schwartz, R. S., Lederman, N. G., & Crawford, B. A. (2004). Views of nature of science in an authentic context: An explicit approach to bridging the gap between nature of science and scientific inquiry. Science Education, 88, 610–645.
Schwartz, T., Bunce, D. M., Silberman, R. G., Stanitski, C. L., Stratton, W. J., & Zipp, A. P. (1997). Chemistry in context (2d ed.). Dubuque, IA: William C. Brown/American Chemical Society.
Seker, H., & Welsh, L. C. (2005). The comparison of explicit and implicit ways of using history of science for students understanding of the nature of science. Paper presented at Eighth International History, Philosophy, Sociology & Science Teaching Conference, Leeds, UK.
Selinger, E., & Crease, R. P. (2006). The philosophy of expertise. New York: Columbia University Press.
Shapin, S. (1996). A social history of truth. Chicago: University of Chicago Press.
Solomon, M. (2001). Social empiricism. Cambridge, MA: MIT Press.

Stephens, T., & Brynner, R. (2001). Dark remedy: The impact of thalidomide and its revival as a vital medicine. New York: Basic Books.
Thomas, J. (2000). Using current controversies in the classroom: Opportunities and concerns. Critical Studies in Education, 41, 2, 133–144.
Thomas, L. (1981). Introduction. In H. F. Judson (Ed.), The search for solutions (pp. ix–x). New York: Holt, Rinehart, Winston.
Turnbull, D. (1993). Maps are territories: Science is an atlas. Chicago: University of Chicago Press.
Wellington, J. (1991). Newspaper science, school science: friends or enemies? International Journal of Science Education, 13, 363–372.
White, M. (1997). Isaac Newton: The last sorcerer. Reading, MA: Perseus Books.
Wimsatt, W. C. (2007). Reengineering philosophy of science for limited beings. Cambridge, MA: Harvard University Press.
Wong, S. L., & Hodson, D. (2009). From the horse's mouth: What scientists say about scientific investigation and scientific knowledge. Science Education, 93, 109–130.
Wong, S. L., Hodson, D., Kwan, J., & Yung, B. H. Y. (2008). Turning crisis into opportunity: Enhancing student–teachers' understanding of nature of science and scientific inquiry through a case study of the scientific research in severe acute respiratory syndrome. International Journal of Science Education, 30, 1417–1439.
Ziman, J. (1978). Reliable knowledge. Cambridge, England: Cambridge University Press.