Science Literacy, Critical Thinking, and Scientific Literature: Guidelines for Evaluating Scientific Literature in the Classroom

Karenann Jurecki[1] and Matthew C. F. Wander[2,a]

ABSTRACT

In this work, we present an approach for teaching students to evaluate scientific literature and other materials critically. We use four criteria divided into two tiers: original research, authority, objectivity, and validity. The first tier, originality and authority, assesses the quality of the source. The second tier, objectivity and validity, assesses the quality of the information presented in that source. The purpose of this work is to develop a concrete system of evaluation based on scientific terminology that can be taught to students over a wide variety of grade levels. Ultimately, the goal is to improve the overall quality of scientific literacy. © *2012 National Association of Geoscience Teachers*. [DOI: 10.5408/11-221.1]

Key words: first-year undergraduate/general, high school/introductory, geoscience education research, problem solving/decision making, nonmajor courses

INTRODUCTION

"Drowning in information" is a phrase that is often used to describe the proliferation of information in modern life, and not just in the sciences. It is in the sciences, however, where we fall alarmingly short. Jon D. Miller, when discussing his findings in a *New York Times* article, found that only 20% to 25% of Americans were scientifically aware while most of the rest "don't have a clue" (Dean, 2005, p. F3). In 1989, it was reported that less than half of Americans, and a third of Britons, knew that the Earth revolved around the sun (Culliton, 1989). The numbers vary, but it has been reported that between 40% and 50% of Americans reject evolution outright in spite of the preponderance of evidence supporting it (Alters and Nelson, 2002; Keeter et al., 2007). The emphasis on creating a scientifically literate population was supposed to help matters, but it has not. It is not a lack of knowledge or exposure to science that is the problem. The problem is the inability to evaluate information critically, to sort the good from the bad.

Scientific Literacy

Science literacy, as defined in *Science for All Americans* (SAA), means being familiar with the natural world, understanding key concepts and principles of science, having a capacity for scientific reasoning, and being able to use scientific knowledge for personal and social purposes (American Association for the Advancement of Science Project 2061 et al., 1994). Apart from being a very broad definition that few, if any, could live up to, nowhere does it say how we are to achieve science literacy at the practical level. *Science for All Americans* is a set of recommendations for ways of thinking that are essential for all citizens living in a world shaped by science and technology (American Association for the Advancement of Science Project 2061 et al., 1994). The report sets learning goals for material that students should learn, remember, and understand once they have left school. However, understanding is not just a linear progression of facts, it is also the connections among those facts (American Association for the Advancement of Science Project 2061 et al., 1994). For all the definitions and benchmarks that exist, there are few, if any, tools designed to help students evaluate scientific literature, even though there are guidelines that tell students how to read articles and Web sites. A model in which students are taught how to read a scientific article, what parts to read, and how carefully, does not go far enough to foster the higher-order critical-thinking skills that are such an integral part of science and would go a long way to cultivating science literacy.

For the past 40 y, the emphasis in science education can be best described as learning science by doing science. Students would learn the basic tenets of the scientific method by doing what scientists do, which is to observe, test, and record (O'Neill and Polman, 2004). The instructor would facilitate this hands-on approach to science, "depending on what is needed in order to keep students actively engaged in pursuit of a learning outcomes" (Center for Science Mathematics and Engineering Standards and Committee on Development of an Addendum to the National Science Education Standards on Scientific Inquiry, 2000, p. 202). Students, particularly young learners, would experience science in a way more in line with natural human curiosity, through experience and play (American Association for the Advancement of Science Commission of Science Education and Gagne, 1965). Discovery, process, and inquiry learning models eschew the traditional, teacher-driven lecture format in favor of a student-driven model where they learn that science is more than memorization (Gagne, 1963). Unfortunately, this emphasis has actually harmed conceptual understanding of science because of the students' poor knowledge of the basic facts upon which those concepts are based (Hodson, 1996).

The factor missing from that model of learning is that scientific knowledge is built on an understanding of prior scientific knowledge—that body of tested, agreed-upon, reliable knowledge (Bauer, 1992)—and so science becomes a chain of experiments with no history or connection to

Received 4 February 2011; revised 11 November 2011; accepted 20 December 2011; published online 13 June 2012.

[1]MCA Solutions, Philadelphia, Pennsylvania 19102, USA
[2]Department of Chemistry, Drexel University, Philadelphia, Pennsylvania 19104, USA
[a]Author to whom correspondence should be addressed. Electronic mail: mcfwander@gmail.com.

existing scientific knowledge, instead of an understanding that, regardless of discipline, all scientists know the part of the puzzle to which the work belongs (Bauer, 1992). This is scientific consensus, and it plays a direct role in the evaluation of literature. In this context, consensus is expressed by means of theory and interpretation, specifically through the evaluation of the robustness of the information in general or the scientific idea in particular. Science and its evaluation are as much about the physical context, being able to draw connections between ideas, as they are about the experiment or measurement itself.

When students are taught that science is all about the experiment, they are led to believe that all experimental research becomes part of the scientific body of knowledge, which is not the case. It also leads students to believe that all science is experimental, or worse, that the only valid science is experimental science. Inquiry learning does not take into account the substantive evaluation process that scientific research goes through before the observations are made, the data are recorded, and the results are published. A balanced, representative, more accurate view of the way in which science is conducted must include a review of the literature as well as hands-on experimental, or theoretical, techniques.

The current science curriculum's de-emphasis of critical-thinking makes the goals of science literacy difficult to obtain. We intend to reintroduce critical thinking into learning science by providing guidelines on reading and evaluation of scientific literature: The ultimate goal is a more balanced, accurate depiction of the way in which science is actually undertaken and understood.

Social Science Standards and the Association of College and Research Libraries

In May 2007, the Association of College and Research Libraries (ACRL) published the latest revision of *Standards for Science and Engineering/Technology* (Association of College and Research Libraries, 2007). Like *Science for All Americans*, the ACRL standards are a set of learning goals, not curriculum guidelines. They were developed for libraries in order to help them evaluate their instruction programs. The literate student "selects information by articulating and applying criteria for evaluating both the information and its sources" (Stand. 3, Ind. 2, Association of College and Research Libraries, 2007). As mentioned earlier, the criteria available for evaluating scientific literature are limited, whereas criteria for social science literature and Web sites are more prevalent. It begs the question, what criteria are students to use?

The ACRL goes on to articulate some necessary criteria, but these are sometimes vague. The scientifically literate student is able to distinguish among "primary, secondary, and tertiary sources and recognizes how location of the information source in the cycle of scientific information relates to the credibility of the information" (Association of College and Research Libraries, 2007). On further examination, two aspects of this performance indicator are ill suited to the sciences. The first issue is the arbitrary division of material into primary, secondary, and tertiary categories. The second is the use of the word credibility. The assumption here is that students have learned the difference between primary, secondary, and tertiary sources, which is not always the case, particularly in a scientific context. Where a primary source in the social sciences revolves around an event, in the sciences it revolves around research. Evaluating a scientific primary source as if it revolves around an event, or creating an arbitrary distinction between secondary or tertiary does not make any sense.

With regard to the second problem, nowhere in the ACRL standards is there an explanation of credibility in a scientific context. There is an additional issue created by the juxtaposition of credibility and source in this performance indicator. It asks the reader to evaluate the credibility of a source based on whether it is a primary or secondary source, treating them differently. It does not seem to make any difference that, more often than not, secondary and tertiary sources make the inaccessible accessible and refer back to the original, primary source material. While students' needs may be better met by a secondary source, this is not a reflection of the quality of the information contained within that source. This ambiguity undermines the credibility of preliminary studies published in conference reports, or a U.S. government white paper, or a thesis.

Performance indicator 3.2 states that the scientifically literate student is supposed to be able to examine and compare information to evaluate: reliability, validity, accuracy, authority, timeliness, and bias (Association of College and Research Libraries, 2007). While these are useful and appropriate terms, once again the ACRL provides no definitions or guidelines on how to evaluate literature for any of these factors. While the ACRL differentiates among reliability, validity, and accuracy, it takes for granted that the average student knows the difference between these concepts and can make the connections among them.

The scientifically literate student is supposed to be able to recognize prejudice, deception, or manipulation in the use of information (Association of College and Research Libraries, 2007). This criterion is nearly impossible for the average student to fulfill. The vast majority of students are not going to recognize fraud or deception or manipulation of data in a scientific article on-sight. Furthermore, since students are not encouraged to read the methods, as they are too "difficult" to understand (Purdue University Libraries, 2008), there is no way to see intentional manipulation of the information. Most researchers do not recognize errors in the work of their peers simply by reading (Martinson et al., 2005). A possible error in a publication may arouse suspicions, and maybe a letter to the editor, but most of the time, errors in the literature go unrecognized until the methods are tested.

Finally, the scientifically literate student "recognizes the cultural, physical, or other context within which the information was created and understands the impact of context on interpreting of information" (Association of College and Research Libraries, 2007). While scientists agree that science does not take place in a vacuum, this standard does not take peer review or scientific consensus into account. By interpreting scientific information within a cultural context, rather than a physical one, one is actually introducing bias, not refuting it (Gould, 1996). This is not the same thing as recognizing the inherent bias of a source. For example, students can understand that studies of the effects of tobacco that are funded by tobacco companies should be treated with care. They can also understand that bias can lead scientists to read other scientists' work incorrectly. The process of removing the cultural context from scientific information allows for the objective evaluation of someone's error and/or bias. The collective nature of science determines what will be part of the scientific body of knowledge and what will not. This is why progress in science spans generations.

Peer Review as Literature Evaluation Criteria for the Classroom

The other evaluative criteria of note were given by James C. Lin, who put forth general principles for evaluating scientific literature that must be met before it can be accepted into the body of established scientific knowledge (Lin, 2002, 2004). The first three points revolve around objectivity in experimental and observational techniques, methods and the description of those methods, and data analysis. Unfortunately, he does not define objectivity. For most scientists, this is not a problem, but for an educator, such a definition is critical. The results should, "demonstrate the effect of the relevant variable at a proper level of statistical significance using the appropriate tests." They should also be "consistent, quantifiable in terms of relevant variables, and susceptible to independent confirmation by independent researchers" (Lin, 2004, p. 36). Lin's criteria stipulate the standards to be met for something to pass peer review. While there is a strong connection between the peer review and the postpublication evaluation process, the former must influence the latter.

We are not comparing apples and oranges, merely illustrating the disconnect between the way in which scientists actually evaluate literature and the way in which science, engineering, and technology students are taught how to evaluate literature. The ACRL encourages students to look for bias and the manipulation of data, while Lin argues that the results must be susceptible to independent confirmation by independent researchers. This is how science is done. The ACRL requires the scientifically literate student be able to recognize the context in which the information was created, while Lin, a scientist, has no such requirement.

In the end, neither system is suitable for a classroom setting. The ACRL results are derived from the social sciences. Lin's approach is more about peer-review guidelines for professional scientists. It is not at an appropriate level for most students. Since, neither are useful for students in the sciences, a middle ground needs to be found.

EDUCATIONAL EXPERIENCE AND MOTIVATION

This work is derived from experience teaching research methods in an alternative setting, the library. These were short courses, usually in concert with an established course or as a semester-long general education course. Students were frequently intimidated by the process of researching articles, and, in particular, what to do with the articles they found. Much of the problem revolved around a lack of understanding of the article, how the authors came to ask the research question in the first place, and how to read their findings. These courses were designed to improve the scientific literacy of students by explaining the scientific process and the way in which it leads to scientific research.

It has been our experience that students approach their assignments under the assumption that what they are researching has never been done before. As such, it was very important for students to learn that all scientific knowledge is part of a continuum, with current research being built upon prior research, which is where the body of literature comes in. In our short courses, we taught that there were certain things to look for when reading science and questions that needed to be asked. A primer on essential scientific language was provided, and a faculty member was invited to speak on his experience creating and publishing scientific literature. They learned that even negative results are just as valid as positive ones, and why. In short, a comprehensive, systematic overview of how science is done and disseminated successfully demystified science for our students. Demystifying science is critical if we are going to educate these students to the level necessary to succeed in today's world.

APPROACH: THE TWO TIERS

Overview

We developed a two-tier set of criteria for evaluating scientific literature that can be used in traditional and nontraditional learning environments, and can be presented to science majors and nonscience majors alike. The first tier is the more concrete of the two; either the article is original scientific research or it is not, and either it is published in an authoritative source or it is not. Authority is a measure of the reputation of the publication and the authors it publishes. We have found this to be too vague and have settled on peer review as the indication of authority. These are not statements of value, but they are designed to get students thinking about the nature and types of literature in the cycle of scientific knowledge.

The purpose of this first tier is to establish a hierarchy of evidence. This tier breaks sources down into four levels of scientific materials. The first level would be an article of original scientific research published in a peer-reviewed journal. This is an article that can truly be said to be part of the accepted body of scientific discussion. The next level is an article that is not original scientific research, such as a review article, meta-analysis, or a literature review, that is published in a peer-reviewed journal. The review article is still subject to the rigors of the peer-review process, and it will direct readers to the original research. The third level or evidence would be an original scientific article that is not published in a peer-reviewed journal, such as a thesis, U.S. government white paper, or a book chapter. The fourth level would be an article that is not original research and is not published in a peer-reviewed publication, such as a newspaper or magazine article or a popular science book. These publications are reviewed by editors for content, but they are not reviewed by peers. They put scientific concepts into layman's terms and direct readers to the related source material should they choose to read it.

The purpose of the second tier is to encourage students to read scientific materials for the quality of the information, not just the quality of the source. This tier is more subjective than the first tier and gets into issues of credibility and trust within the context of objectivity and validity. These two criteria, objectivity and validity, strike at the heart of the scientific method. Objectivity refers to the nature of the test that is performed in the primary study, or the fairness with which a review of material is presented; in layman's terms, it is the credibility of the study. Validity refers to the strength of the information presented based on sound scientific principles of reliability, rigor, and robustness, which we will define subsequently. Critical thinking is a key requirement of this second tier, in addition to being a goal of this work. Students cannot leave their critical-thinking skills at the door.

Application

Tier I: Originality (Original Research) and Authority (Peer Review)

To establish whether or not an article is original scientific research, we propose the following criteria:

1. The article focuses on a single, well-defined topic (the hypothesis), which is the starting point of the research.
2. The article describes the experimental or computational design.
3. The article outlines the methods.
4. The article contains statistical/quantifiable data that either support or refute the hypothesis.
5. The article discusses the results.
6. The article suggests a course for future research.

The next step is to determine whether or not the article was published in an authoritative source. The standard of authority among scientists and the academic community alike is publication in a peer-reviewed, refereed, or juried journal. Throughout academia, the fact that a paper has been accepted for publication in a well-known refereed journal is probably the best indication that it reports quality research (Braun, 2004). Peer-reviewed journals are the primary means of communicating ideas within the scientific community, which, when accepted, are disseminated in books, textbooks, primers, and other forms of scholarly communication. An article in a journal is evaluated on specific criteria established by the journal, including validity, rigor, and scope. Reviewers and editors ensure that the authors used established methods and protocols, and that the finished product contributes to the body of scientific knowledge. A textbook or primer is only as good as the material it is citing, but it can be a good indicator of the current scientific consensus.

Tier II: Objectivity and Validity

The second tier requires more careful reading of the language, data, methods, and results. The first step is for students to think of scientific objectivity as a set of open-minded, self-correcting procedures designed to ensure reasonable results that will be able to withstand scrutiny (Lett, 1997). To help students judge the objectivity, we propose the following criteria:

1. The author(s) poses a hypothesis that can be tested.
2. The objectives and methods are written clearly and explained adequately.
3. The results are written in language that is unambiguous and free of bias.
4. It is easy to ascertain who sponsored the research.
5. It is easy to ascertain the author's credentials.

Assessing validity requires careful examination of the whole study and related research for experimental and methodological design, presentation and interpretation of results, and the conclusions. Given that students have different levels of capability, we recommend that when implementing this method, it should be done with the instructor guiding the process. The exercise itself is of value, even if students are not able to fully implement all of the criteria by themselves. We define validity as having evidence that is reliable and relevant to the hypothesis. Specifically, we propose the following criteria for assessing validity:

1. Responsiveness: The evidence answers the question put forth in the hypothesis.
2. Robustness: The researcher(s) used established methods and techniques.
3. Reliability: Other researchers performed the same tests on the same instruments and got the same results.
4. Rigor: The interpretation of the results makes sense and is consistent with the results of similar work. If the results are not consistent, the authors explain those inconsistencies adequately. Ultimately, scientists will evaluate this criterion by surveying the literature, but students can use the introduction of the work itself and its treatment of its references as a valuable tool to answering this question.

Students should read the methods section of the paper to learn how the experiment was done. Although it is not necessary for the student to understand every aspect of the methods, the attempt itself is educational. If they want to conduct the experiment themselves, then a more careful reading of the methods is required. To determine if the researchers used established techniques, students should compare and contrast with other articles in their literature search, or with the articles cited by the authors and articles that cite the article they are reading.

APPLICATION AND DISCUSSION

The indicators of authority in scientific research are peer review and impact factor. In order for an article to be published in a peer-reviewed journal, it must be approved by a majority of the review panel and the editor. Students can find out whether or not a source journal is peer reviewed either directly from the journal or from *Ulrich's Periodicals Directory*. Impact factor is not included in our criteria, but it should be discussed in conjunction with peer review. It would be beneficial for students to know that it is another tool used by academics and administrators to rank, evaluate, and compare journals as a gauge of their authority in a given field; the higher the impact factor, the more authoritative the journal is considered to be. Impact factors are not without controversy, so they should be used carefully. Students do not need to know the formula used to calculate the impact factor. The reason it is not part of our criteria is simple: Not every journal has an impact factor.

We contemplated clarity as a possible criterion but decided that it is part of objectivity. If the authors cannot, or will not, describe their methods or discuss their results clearly and without unnecessary jargon, the question must be raised as to whether or not the authors are trying to obfuscate deliberately. Terminology in itself is not jargon, but the use of flowery language could be. As a result, we incorporated clarity into our criteria for objectivity.

Scientific validity has been described as "the use of accepted principles and methods, including statistical techniques to produce reliable and [accurate] data'' (Emanuel et al., 2000, p. 2703). Validity determines if the evidence presented answers the question satisfactorily (Gott and Duggan, 2003). We considered many ways for students to examine validity and decided that they should evaluate validity on four points: responsiveness, rigor, robustness, and reliability. These points are easy to determine with the average literature review. While not required, a literature review is a very valuable exercise for a student to undertake. As mentioned before, responsiveness is a measure of how well the results answer the question put forth in the hypothesis. One of the reasons scientists consult the literature beforehand is to see what methods and techniques their peers are using, and then they apply the appropriate technique(s) to their research. This is a measure of the robustness of the methods. Robustness is also a measure of the flexibility of the methods. If the methods do not stand up to repeated, independent testing, then they are invalid, and the results are invalid. This is an indicator of rigor. Scientists also consult the literature to review the results. Consistent results, over a period of time, are the indicator of reliability.

Timeliness is a common criterion for evaluating literature in the humanities and social sciences, as well as in medicine. It acts as a measure of currency, asking if a particular article is the most current research, as well as providing an historical time line. We considered timeliness as a criterion but decided against it, as it would be too difficult to apply. Determining timeliness would require near-expert knowledge of methods, techniques, and even instrumentation on the part of the students. It is more difficult to determine timeliness because it would require the same level of knowledge of the extant literature on their subject, including the history and development, which is a tall order. We decided that it would be easier, and more beneficial, for students to be able to make the connections among responsiveness, reliability, rigor, robustness, and validity by reading the literature and comparing the methods, techniques, and results.

CONCLUSION

The purpose of this project is to foster critical-thinking skills in science students and nonscience students alike by providing them with a means by which to undertake critical reading of scientific literature. Whether a student is attempting to read a peer-reviewed publication or a scientific newspaper article, an understanding of the physical context of the material is essential to its evaluation. Our goal is to teach students at the college level how to read and evaluate science literature of all kinds. No science experiment or theoretical calculation is performed without a thorough review of the literature first. Students must learn about the whole scientific process. They need better, more concise definitions for important concepts like objectivity, validity, reliability, hypothesis, and theory. They also need to learn how these concepts work together.

It is unfortunate that in a time when science is such a part of daily life, students are not encouraged to read it critically. Those that advocate for inquiry learning put forth that students ''re-experience scientific inquiry and discovery as the original 'first-thinkers' experienced it'' (Elrod and Somerville, 2007, p. 689). If we really want students to get the same sense of discovery that the scientists do, we should have them do science the way the experts do: formulate the hypothesis, review the literature, and then observe, record, analyze, and evaluate. In the well-intentioned effort to keep students engaged by making the sciences more accessible, we have put the sciences further out of reach.

SUPPLEMENTAL MATERIALS

Additional materials for use in the classroom are available at: http://nagt.org/nagt/teaching_resources/materials.html or by emailing the authors.

Acknowledgments

The authors would like to thank Barry Bickmore for the opportunities to share his class to try some of these ideas, and Steve Borrelli, Elaine Brady, Aurora Clarke, Betty Galbraith, Corey Johnson, and Stephanie Westcott for their helpful comments.

REFERENCES

Alters, B.J., and Nelson, C.E. 2002. Perspective: Teaching evolution in higher education. *Evolution*, 56:1891–1901.

American Association for the Advancement of Science Commission of Science Education, and Gagne, R.M. 1965. The psychological basis of science: A project approach. Washington, DC: AAAS. p. 35.

American Association for the Advancement of Science Project 2061, Rutherford, F.J., and Ahlgren, A. 1994. Science for all Americans. New York: Oxford University Press. p. 272.

Association of College and Research Libraries. 2007. Information literacy standards for science and engineering/technology, volume 2008. Chicago IL: Association of College and Research Libraries.

Bauer, H.H. 1992. Scientific literacy and the myth of the scientific method. Urbana, IL: University of Illinois Press. p. 180.

Braun, T. 2004. Keeping the gates of science journals: Gatekeeping indicators of national performance in the sciences. *In* Moed, H.F., Glanzel, W., and Schmoch, U., eds., Handbook of quantitative science and technology research: The use of publication and patent statistics in studies of S&T systems. Boston, MA: Kluwer Academic. p. 95–114.

Center for Science Mathematics and Engineering Standards, and Committee on Development of an Addendum to the National Science Education Standards on Scientific Inquiry. 2000. Inquiry and the National Science Education Standards: A guide for teaching and learning. Washington, DC: National Academy Press. p. 202.

Culliton, B. 1989. The dismal state of scientific literacy. *Science*, 243:600.

Dean, C. Scientific savvy? In the US not much. *New York Times*, 23 August 2005, F3.

Elrod, S.L., and Somerville, M.M. 2007. Literature-based scientific learning: A collaboration model. *The Journal of Academic Librarianship*, 33:684–691.

Emanuel, E.J., Wendler, D., and Grady, C. 2000. What makes clinical research ethical. *JAMA*, 283:2701–2711.

Gagne, R.M. 1963. The learning requirements for enquiry. *Journal of Research in Science Teaching*, 1:144–153.

Gott, R.A., and Duggan, S. 2003. Understanding and using scientific evidence: How to critically evaluate data. Thousand Oaks, CA: Sage. p. 195.

Gould, S.J. 1996. The mismeasure of man. New York: Norton. p. 432.

Hodson, D. 1996. Laboratory work as scientific method: Three decades of confusion and distortion. *Journal of Curriculum Studies*, 28:115–135.

Keeter, S., Smith, G., and Masci, D. 2007. Religious belief and public attitudes about science in the US; International indicators of science and the public. London, UK: The Pew Research Center.

Lett, W.J. 1997. Science, reason, and anthropology: The principles of rational inquiry. Lanham, MD: Rowan & Littlefield.

Lin, J.C. 2002. Scientific literature on biological effects of radio frequency radiation: Criteria for evaluation. *IEEE Antennas and Propagation Magazine*, 44:140–142.

———. 2004. Evaluating scientific literature on biological effects of microwave radiation. *IEEE Microwave Magazine*, 5:34–37.

Martinson, B.C., Anderson, M.S., and deVries, R. 2005. Scientists behaving badly. *Science*, 435:737–738.

O'Neill, D.K., and Polman, J.L. 2004. Why educate ''little scientists?'' Examining the potential of practice-based scientific literacy. *Journal of Research in Science Teaching*, 41:234–266.

Purdue University Libraries. 2008. How to read a scientific article, volume 2008. West Lafayette, IN: Purdue University Libraries.