Assessing inquiry skills as a component of scientific literacy

Carl J. Wenning, Physics Teacher Education Coordinator, Illinois State University, Normal, IL 61790-4560 wenning@phy.ilstu.edu

It is frequently said that achieving scientific literacy is the main goal of science education. It would seem reasonable, therefore, that an assessment instrument would exist for measuring progress toward that goal. Unfortunately, such an instrument does not appear to exist. Indeed, a single test of reasonable length would not likely be suitable for assessing scientific literacy in a comprehensive fashion. A battery of independent tests geared toward the task of assessing scientific literacy in its many dimensions is needed. Such battery tests can provide critical information to assess gaps in student knowledge and skills, guide instructional practice, hold schools accountable for achieving specific goals, and determine program and teacher effectiveness. After establishing a framework for assessing the skills of scientific inquiry, the author makes available the second in a series of tests being developed for comprehensively assessing a previously defined form of scientific literacy. The 35-item Scientific Inquiry Literacy Test (ScInqLiT) is introduced and explained.

Scientific literacy is multidimensional, and comes in a variety of types and degrees (Shen, 1975; Shamos, 1995; National Research Council [NRC], 1996). A relatively comprehensive form of scientific literacy that teachers might attempt to achieve among their students has been defined in the *National Science Education Standards* (NRC, 1996). The *National Science Education Standards* indicate that a scientifically literate individual will possess an understanding of six major elements of scientific literacy: (1) science as inquiry, (2) science content, (3) science and technology, (4) science in personal and social perspectives, (5) history and nature of science, and (6) unifying concepts and processes. Using this description as a guide, the author of this article has begun to develop a battery of standardized tests that can be used to measure progress toward attaining the goal of achieving scientific literacy so defined. In two previous articles, Wenning (2006a, 2006b) presented a framework for teaching and assessing Nature of Science (NOS) literacy. In the current article, the author proposes an operational definition of scientific inquiry suitable for guiding high school science teaching, reiterates a framework for teaching it, and describes a standardized test for assessing student knowledge and skills associated with scientific inquiry.

Operationally Defining Scientific Inquiry

Scientific inquiry – as a component of scientific literacy – has been variously defined. For instance, the *National Science Education Standards* (NRC, 1996) defines scientific inquiry as follows, “Scientific inquiry refers to the diverse ways in which scientists study the natural world and propose explanations based on the evidence derived from their work. Inquiry also refers to the activities of students in which they develop knowledge and understanding of scientific ideas, as well as an understanding of how scientists study the natural world” (p. 23).

Project 2061 gives a slightly different definition in *Benchmarks for Science Literacy* (American Association for the Advancement of Science [AAAS], 1993), “Scientific inquiry is more complex than popular conceptions would have it. It is, for instance, a more subtle and demanding process than the naive idea of ‘making a great many careful observations and then organizing them.’ It is far more flexible than the rigid sequence of steps commonly depicted in textbooks as ‘the scientific method.’ It is much more than just ‘doing experiments,’ and it is not confined to laboratories. More imagination and inventiveness are involved in scientific inquiry than many people realize, yet sooner or later strict logic and empirical evidence must have their day. Individual investigators working alone sometimes make great discoveries, but the steady advancement of science depends on the enterprise as a whole” (p. 9).

The National Science Teachers Association (NSTA, 2004) defines scientific inquiry somewhat differently still, “Scientific inquiry is a powerful way of understanding science content. Students learn how to ask questions and use evidence to answer them. In the process of learning the strategies of scientific inquiry, students learn to conduct an investigation and collect evidence from a variety of sources, develop an explanation from the data, and communicate and defend their conclusions” (p. 1).

While such statements are true – and several specific examples of scientific inquiry are given in the associated texts – these broad characterizations and specific examples are of little help to science teachers who are looking for a detailed operational definition that can guide science teaching. For the purpose of operationally defining scientific inquiry at a level appropriate for secondary schools, the author provides a listing of fundamental scientific inquiry skills in Table 1. These processes have been roughly organized into “stages” of scientific inquiry, and are patterned on the inquiry processes described in Wenning (2005a).

While the listing in Table 1 might at first appear to be based on a rather naïve understanding of the nature of scientific inquiry encountered in the secondary school classroom, it was developed in light of works by Kneller, Bauer, Wynn, Popper, Gould, Root-Berstein, Sayer and a number of others whose writings have been included in *Science and Its Ways of Knowing* edited by Hatton and Plouffe (1997). The author is fully cognizant of the fact that there is no “scientific method” per se, and that science more often than not develops along ways that are not consistent with the traditional Baconian approach. Further, this listing was developed in light of the fact that most scientific work at the secondary school level is not driven by hypothesis or model generation and

Stages of Scientific Inquiry

- Identify a problem to be investigated.
- Using induction, formulate a hypothesis or model incorporating logic and evidence.
- Using deduction, generate a prediction from the hypothesis or model.
- Design experimental procedures to test the prediction.
- Conduct a scientific experiment, observation or simulation to test the hypothesis or model:
  - Identify the experimental system
  - Identify and define variables operationally
  - Conduct a controlled experiment or observation
- Collect meaningful data, organize, and analyze data accurately and precisely:
  - Analyze data for trends and relationships
  - Construct and interpret a graph
  - Develop a law based on evidence using graphical methods or other mathematic model, or develop a principle using induction
- Apply numerical and statistical methods to numerical data to reach and support conclusions:
  - Use technology and math during investigations
  - Apply statistical methods to make predictions and to test the accuracy of results
  - Draw appropriate conclusions from evidence
- Explain any unexpected results:
  - Formulate an alternative hypothesis or model if necessary
  - Identify and communicate sources of unavoidable experimental error
  - Identify possible reasons for inconsistent results such as sources of error or uncontrolled conditions
- Using available technology, report, display, and defend the results of an investigation to audiences that might include professionals and technical experts.

Table 1. *A limited framework defining scientific inquiry skills as a part of scientific literacy. This framework is intended to be suggestive, not definitive.*

theory development, but that typically data are collected for the purpose of formulating principles or developing empirical laws. Finally, this listing was prepared with the understanding that not all inquiry processes will be experimental in nature. Sometimes evidence and logic alone will be used to draw scientific conclusions. Additionally, not all scientific inquiry skills will be used in any one investigation. Scientific inquiry based on observations will likely differ significantly form scientific inquiry based on experimentation. Geologist, biologists, chemists, and physicists, for example, all have different approaches to conducting scientific investigations and will use various elements of the listing to different degrees.

A Framework for Teaching Scientific Inquiry Skills

A framework must be provided if science teachers are to teach scientific inquiry skills systematically and a level appropriate to the intellectual maturity of their students. For instance, the approaches used for teaching early elementary children will differ remarkably from techniques used at the high school level. Teaching scientific inquiry skills effectively requires definitions of both the stages *and* levels of scientific inquiry suitable for students. Table 1 describes, roughly speaking, the levels of scientific inquiry (Wenning, 2005a). The most important features of the levels-of-inquiry spectrum are shown in Figure 1.

Using the stages and levels of inquiry sequences, teachers can implement inquiry practices in the science classroom. Teachers thereby help students learn inquiry skills by modeling successively more sophisticated forms of inquiry. Students develop increased understanding by moving through progressively more sophisticated levels of inquiry and carrying out various stages of inquiry repeatedly. As the level of intellectual sophistication required to conduct the various levels of inquiry grows, the locus of control shifts from teacher to student. For instance, during discovery learning the teacher directs students to make specific observations and guides them to draw specific conclusions using "funneling" questions (Wood, 1998). Inquiry lessons require the teacher to use a think aloud protocol to explain various scientific practices. While the teacher maintains control of equipment and the experiment, students are encouraged through "focusing" questions (Wood, 1998) that help them understand the nature of the scientific process. With inquiry labs, students take greater

Discovery Learning	Interactive Demonstrations	Inquiry Lessons	Guided Inquiry Labs	Bounded Inquiry Labs	Free Inquiry Labs	Pure Hypothetical Inquiry Applied Hypothetical Inquiry
Low		<= Intellectual Sophistication =>				High
Teacher		<= Locus of Control =>				Student

Figure 1. *The levels-of-inquiry spectrum. As students become more intellectually sophisticated, the level of inquiry utilized by teachers correspondingly can become more sophisticated. At the same time, the locus of control shifts gradually from the teacher to the student.*

control of the entire learning process, from answering a series of questions and developing problems, to designing experimental procedures and drawing conclusion on their own. Lastly, during the advanced levels of hypothetical inquiry, students identify their own problems, develop hypotheses or models, make predictions, conduct experiments or observations, and draw conclusions on the basis of logic using empirical evidence. Interested readers are referred to the article *Levels of inquiry: Hierarchies of pedagogical practices and inquiry processes* (Wenning, 2005a) for additional information and examples associated with each of the levels within the inquiry spectrum.

Scientific Inquiry Literacy Test (ScInqLiT)

Eight steps were followed in the development of the Scientific Inquiry Literacy Test *(ScInqLiT)* using general procedures outlined by DeVellis (1991). The first step was to develop a framework that clearly defines what is to be measured. The framework for *ScInqLiT* can be found in Table 1. This framework operationally defines what constitutes literacy in scientific inquiry at a level appropriate to the understanding of high school science students. This framework gave a clear statement about what needed to be included in the assessment that came to be based upon it. The framework was reviewed by several physics teaching majors, scientists, and educators to provide reasonable assurance of content validity.

A pool of 40 questions was then generated for possible inclusion in the final assessment instrument. Each item consisted of a multiple-choice question with four possible answers. One or more questions were generated for each of the specifications presented in the framework. A team of six reviewers consisting of senior level undergraduate physics teacher education majors then reviewed the items for accuracy and clarity. Each of these reviewers had a good understanding of scientific inquiry as demonstrated by multiple and varied assessments completed as part of their physics teacher education course work at Illinois State University.

An initial pilot test consisting of the 40 questions was administered to 425 high school science students enrolled in five different central Illinois high schools during early February 2007. The population generally consisted of freshmen enrolled in introductory lab science, biological science, or general science courses, sophomores and juniors enrolled in chemistry courses, and juniors and seniors enrolled in physics courses. The range of scores on the pilot test was 0 to 36 out of 40 possible. The test mean was 18.78 (46.95%) with a standard deviation of 7.90 and a standard error of measurement of 2.79. The KR20 reliability coefficient was an unexpectedly high 88%. An analysis was conducted of each test item examining difficulty, discrimination, and suitability of foils. The mean item difficulty for 4-response multiple-choice questions was 0.469, which is a bit low for multiple-choice questions with four responses each. To maximize item discrimination, desirable difficulty levels are slightly higher than the midpoint between random guessing (1.00 divided by the number of choices) and perfect scores (1.00) for the item. The ideal mean difficulty for the four response multiple-choice questions used in this test therefore should not deviate much from a value of 0.625.

Five poor performing (very high or very low difficulty and/or small to negative discrimination) and somewhat redundant test items were removed, and one non-redundant but poor-performing test item was revised. This question was rewritten for increased clarity, and better alternative answers were prepared. The revised pilot test was administered a second time during mid to late April 2007 to 61 entirely different high school students. It is believed that these students – four classes from among two teachers – were highly motivated, and the groups relatively homogeneous. The high/low scores were 12/31 out of 35 possible. The mean test score of these students was 23.7 (67.6%), which is slightly higher than expected for a test designed to produce the maximum possible spread among scores. The standard deviation of the sample was 4.62, and the standard error of the mean 2.49. The mean item difficulty was 0.68 meaning that, on average, 68% of students completing a question gave the correct response. This exceeds the ideal mean item difficulty for a test of this format and did so, ostensibly, due to the fact that this latter pilot group was both motivated and homogeneous. The mean item discriminability was 0.32. These facts, plus the fact that the frequency distribution of scores was positively skewed, support belief in the motivated/homogeneous assumption. This is an important factor in interpreting item analysis data. The KR20 reliability coefficient was 0.71, not unanticipated given the nature of the second pilot group.

Following the second pilot study, and as part of the final review process for publication, one question was replaced and several others had their wording revised for improved clarity. It is expected that the finalized version of *ScInqLiT* has increased validity and reliability as a result of these changes.

Administering ScInqLiT

ScInqLiT is an un-timed test requiring typically about 40 to 50 minutes for nearly all high school students to complete. *ScInqLiT* probably is best employed under pre-test, post-test conditions; it generally should not be used as an achievement test. Due to its nature as a diagnostic test, the results from any testing situation probably will be unacceptably low. Questions have been developed and selected to provide a maximum dispersion of scores. As can be seen from the pilot study samples, average scores on these tests hover in the vicinity of 47% to 68% for high school students. *ScInqLiT* is best used primarily for the purpose for which it was created – to serve as a research instrument for identifying weaknesses in student understanding, improving instructional practice, and determining program effectiveness in relation to teaching scientific inquiry skills. *ScInqLiT* can be used readily for educational research or during professional development workshops for both elementary- and secondary-level teachers to show learning gains among participants.

The author encourages widespread use of *ScInqLiT,* and urges that test results be forwarded to him along with participant demographics so that the test can be normed using a variety of study groups. Users are requested to keep the instrument secure as with other standardized tests, and collect copies from students following testing. Use of the names *Scientific Inquiry Literacy Test* and *ScInqLiT* should also be avoided with students to help prevent them searching the Web for background information.

Limitations of ScInqLiT

Ideally, assessing procedural knowledge will be done use ing performance tests. *ScInqLiT* is a paper-and-pencil test. As such, it is limited in its ability to authentically assess student abilities to conduct scientific inquiry. Ideally, a test of scientific inquiry abilities would include materials with which a student would create and conduct a scientific experiment and draw legitimate conclusions. Alternatively, observational data could be provided to students who would then interpret that data to draw scientific conclusions. As a paper-and-pencil test, *ScInqLiT* should be thought of as only an indicator of student ability to conduct scientific inquiry. Researchers would do well to develop authentic tests including manipulatives that might be used to more fully assess student ability to conduct scientific inquiry in each of the various science disciplines.

The Importance of ScInqLiT

The No Child Left Behind Act of 2002 requires that all 50 states develop challenging goals in science and assess student progress toward the goals outlined in the *National Science Education Standards.* The required assessment in science must be in place for the 2007-2008 school year. States are now working toward developing their responses to the Federal mandate. The problem before them is to prepare and implement quality science assessments by the deadline. In an effort to provide assistance with this effort, the National Science Foundation asked the National Research Council to formulate guidelines for this work. The NRC responded by producing *Systems for State Science Assessment* (NRC, 2005). This work was predicated on the fundamental position of the *National Science Education Standards:* scientific literacy should be the goal for all K-12 science education. Any science assessment should therefore include not only content knowledge assessment, but also the critically important idea that it is important "...for students to understand science as a specific way of knowing..." (p. 1). As can be seen from a review of the historical development of a definition for scientific literacy (Wenning, 2006), one of the central themes has almost always been an understanding of how one conducts scientific inquiry.

If the main goal of science education is indeed the attainment of scientific literacy, then understanding the processes of scientific inquiry is critically important to achieving the stated goal. A scientific inquiry literacy assessment instrument – an instrument for measuring a fundamental dimension of scientific literacy – could have a significant impact on both curriculum design and instructional practice. For instance, assessments and their frameworks provide important data required for informed decision making, for holding schools accountable for meeting achievement goals, and for determining program effectiveness. Additionally, such assessments and their associated frameworks can help classroom teachers, school administrators, and educational agencies to exemplify their goals for student learning. *ScInqLiT* is currently being used as part of a *Student Teacher Effectiveness Reporting System* at Illinois State University that will be the subject of a future article.

Teachers, teacher educators, and science education researchers wishing to obtain a copy of the *Scientific Inquiry Literacy Test* (*ScInqLiT*) may download it as a password-protected portable document file (PDF) from the *Journal of Physics Teacher Education Online* Web site at the following URL: http://www.phy.ilstu.edu/jpteo/ScInqLiT.pdf. The associated *Nature of Science Literacy Test* (*NOSLiT*) is similarly available at http://www.phy.ilstu.edu/jpteo/NOSLiT.pdf. The passwords for both tests may be obtained directly from the author of this article by e-mailing him at wenning@phy.ilstu.edu.

References

American Association for the Advancement of Science (1993). *Benchmarks for Science Literacy.* New York: Oxford University Press.

DeVellis, R. F. (1991). *Scale Development: Theory and Applications.* London: Sage Publications.

Hatton, J., & Plouffe, P. B. (1997), *Science and Its Ways of Knowing.* Upper Saddle River, NJ: Prentice Hall.

National Research Council (1996). *National Science Education Standards.* Washington, DC: National Academies Press.

National Research Council (2005). *Systems for State Science Assessment.* Mark R. Wilson and Meryl W. Bertenthal (Eds.), Committee on Test Design for K-12 Science Achievement. Washington, DC: National Academies Press.

Shamos, M. (1995). *The Myth of Scientific Literacy.* New Brunswick, NJ: Rutgers University Press.

Shen, B. S. P. (1975). Scientific literacy: The public need. *The Sciences,* Jan.-Feb., 27-29.

Wenning, C. J. (2005b). Implementing inquiry-based instruction in the science classroom: A new model for solving the improvement-of-practice problem. *Journal of Physics Teacher Education Online, 2*(4), 9-15.

Wenning, C. J. (2005a). Levels of inquiry: Hierarchies of pedagogical practices and inquiry processes. *Journal of Physics Teacher Education Online, 2*(3), 3-11.

Wenning, C. J. (2006a). A framework for teaching the nature of science. *Journal of Physics Teacher Education Online, 3*(3), 3-10.

Wenning, C. J. (2006b). Assessing nature-of-science literacy as one component of scientific literacy. *Journal of Physics Teacher Education Online, 3*(4), 3-14.

Wood, T. (1998). Alternative patterns of communication in mathematics classes: Funneling or focusing? In *Language and Communication in the Mathematics Classroom,* eds. Heinz Steinbring, Maria G. Bartolini Bussi, and Anna Sierpinska. Reston, VA: NCTM, 167-78.