A Critical Evaluation of PISA's Assessment of Science Literacy

Audrey B. Champagne
Professor Emerita, University at Albany, SUNY

June 2, 2009

INTRODUCTION

The charge of the conference organizers was to prepare a critical evaluation of the Programme for International Student Assessment's (PISA's) assessment of science literacy. I have addressed the charge using a general framework for science literacy to guide the analysis of the PISA perspective of science literacy as it is contained in *Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006* (Organisation for Economic Co-operation and Development [OECD] 2006) and then compared released items used to assess scientific literacy with its description. I argue that making policy decisions based on assessment results must be informed by an understanding of the construct measured, the components of science literacy, and the approach to measuring the constructs.

I have chosen rather to do a critical analysis of PISA's assessment of science literacy rather than an evaluation. Evaluation of an assessment has both conceptual and technical components. My analysis is on the conceptual components. I leave to the psychometricians and measurement experts all the technical issues. Furthermore, evaluation requires a standard for comparison. No such standard exists for science literacy. Science literacy is an ill-defined construct. Examples of the debate regarding its essential elements and importance can be found in the research literature of education, science, business, and philosophy and in the popular press (Burns, O'Connor, and Stocklmayer 2003; Collins and Pinch 1993; Gregory and Miller 1998; Hilgartner 1990; Irwin 2001; Miller 1998; Nisbet et al. 2002; Sturgis and Allum 2004; Wynne 1992). Among the most frequently cited definitions of science literacy are those contained in the *National Science Education Standards* (National Research Council 1966); the American Association for the Advancement of Science, *Benchmarks for Science Literacy* (AAAS 1993); and the frameworks for the Trends in International Mathematics and Science Study (TIMSS)

(Mullis et al. 2005) and the National Assessment of Educational Progress (NAEP) (National Assessment Governing Board 2007) assessments. While all of these have influenced how states have defined science literacy, not one of them is an acknowledged standard against which the PISA perspective on science literacy can be evaluated. One might arbitrarily choose one of these as a standard; however, a fairer approach is to analyze the match between PISA's definition in its documentation and its operational definition in the items it uses to measure science literacy.

CONTEXT FOR THE ANALYSIS

Student performance data from large-scale assessments are used for policy development and accountability. Typically, policy development is based on the performance of a country's, state's, or locality's students along a scale. Locations along the scales are used to delineate different levels of student performance and to order states' or countries' education systems by the performance of their students. Accountability judgments and policy development rely heavily on scale scores and less on the constructs that the assessment purports to measure or the statistical assumptions underlying the statistical translation of student performance on items to scale scores. The analytical approach taken in this paper illustrates how users of large-scale assessment data might take construct definition and its translation into items and testing time into consideration as parts of decision making. Ideally, such an analysis would include all items in the operational assessment; however, only released items are available for this analysis, and the paper is too short to do a complete analysis of all released items. This is a limitation of the paper.

DEFINING SCIENCE LITERACY

Definitions of science literacy abound, reflecting the values and educational philosophies of those who developed the definitions. The values and philosophies underlying the definitions are not always explicitly stated by the developers but include the value of being science literate to the individual and the value of a science literate citizenry to a nation or the world. The value of science literacy to individuals is based on the assumption that being science literate will enable "better" individual decisions regarding personal health, civic decisions, and the ability to perform well in the workplace or military. The value of science literacy to a nation is based on the assumption that a science literate citizenry contributes economic prosperity. The global perspective reflects the assumption that a science literate population will result in "better" decisions regarding issues relating to the Earth's environment, including moderating climate change or population growth. Perspectives on value influence assumptions about literacy levels required for optimal participation in personal, community, national, and global affairs.

LITERACY LEVELS

Venesky (1990) defines three levels of literacy: learned, competent, and capable of minimal function. These levels are not distinct but identify points along a continuum. From a societal perspective, literacy levels range from that necessary for functioning as a contributing member of society—earning a living, voting regularly and intelligently, attending to health matters—to that level necessary for functioning as a learned participant—a national political leader or a Nobel Prize winner. Ordinary literacy allows us to "get by" in the daily activities of life; other types of literacy are characteristic of the professions and the academic disciplines. Individuals functioning in academic disciplines and the professions are literate in the ordinary sense.[1] In addition, each discipline or profession has its own knowledge and practices that define the literacy of the profession.[2] A national goal for education is that individuals literate in the ordinary sense have some level of literacy in science and in other disciplines, including mathematics, engineering, and economics. The contemporary call for science literate citizens suggests that science literacy in the ordinary sense is necessary for active and intelligent participation in the workplace and in civic affairs. For science, the call goes beyond the basics to the ability to apply the knowledge of science and the reasoning characteristic of scientists to make reasoned decisions and develop logical and convincing arguments to defend those decisions.

An additional aspect of literacy is the cognitive interplay among different representations (Janvier 1987; Paivio 1986). Written science communication includes different forms of representation: verbal, diagrammatic (e.g., charts, drawings, tables, and graphs), and symbolic (e.g., formulas, calculations, and number sentences). The science literate person employs these representations to an advantage in arguments and explanations, often using one or more modes of representation and making connections among representations.

Definitions of science literacy are influenced by developers' assumptions about the value of science literacy to the individual and society, the ways in which science literary will influence an individual's decision making, and the level of literacy required for adequate decision making. These, in turn, influence the components of science literacy and the relative emphasis on each of the components in the definition.

COMPONENTS OF SCIENCE LITERACY

In simplest terms, science literacy has knowledge and abilities components. These components are often described separately, but as will be illustrated later, they are intertwined. Table 1 contains a list of knowledge components that are included in definitions of science literacy. Not all knowledge components are included in all definitions, and in many cases, knowledge is assumed that is not explicitly contained in the definition.[3] The kinds of knowledge are extensive, including not only knowledge related to the knowledge products of the science disciplines but also the practices that result in those knowledge products. Knowledge products of the science disciplines include theories, laws, and the facts, concepts, and principles of which the theories and laws are composed. The practices of science, including inquiry and experimentation, are complex, including the characteristics of how science inquiry is conducted and the attributes of scientific evidence, explanations, and investigations. Knowledge of these components of the processes and attributes of quality are components of the knowledge of science.

Some definitions of science literacy include knowledge of the history and philosophy of science and concepts and processes that are characteristic of all the science disciplines. Because being science literate requires knowledge and skills associated with disciplines other than the natural sciences, definitions of science literacy often include knowledge from other disciplines. Knowledge of technology, construed as both engineering design and computer literacy, is often a component of science literacy.

One of the most important distinctions regarding knowledge of science is expressed as the difference between knowing and understanding. Especially in definitions of literacy originating in the education community is the idea that it is not sufficient to know something about science—the definition of a concept or a principle, for instance. The science literate person knows and understands. Evidence of understanding is the ability to apply the knowledge. For example, simply knowing that the boiling point of liquid is the temperature at which the vapor pressure of the liquid equals atmospheric pressure and exhibiting that knowledge by reciting it when asked are not indicative of understanding. Evidence of understanding is the ability to apply the principle to explain why it takes a potato longer to cook on Pikes Peak than to cook on Long Island.

The ability to apply science facts, concepts, principles, laws, and theories links the knowledge (definition of the boiling point) and ability (application of the principle) components of science literacy. Table 2 contains abilities that typically are included in definitions of science literacy. The application of science knowledge takes place in different situations, including learning in both formal and informal situations, making personal decisions, and participating in civic activities and the workplace. In all of these situations, it is assumed that science literate persons will apply their knowledge and think as scientists think when they engage in the practice of their discipline. The behaviors include: identifying scientific questions; the design, conduct, and critique of scientific studies (experiments, investigations); composing and critiquing explanations of natural phenomenon; making and critiquing predictions; evaluating the quality of evidence; and communicating results to others. Application requires literacy skills, reading and writing, and the ability to reason scientifically.

This science literacy framework provides a guide to help identify the knowledge and abilities included in the definition of literacy, the level of literacy that will be assessed, and assumptions about how individuals in the population being assessed will use the knowledge and skills.

SCIENTIFIC LITERACY IN PISA DOCUMENTATION

PISA's definition of scientific literacy and rationales for its components are contained in *Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006* (OECD 2006). The science portion of PISA 2006 was designed to answer the question: "What is it important for citizens to know, value, and be able to do in situations involving science and technology?" The answer to this question is contained in a definition of scientific literacy and its elaboration:

For the purposes of PISA 2006, *scientific literacy* refers to an individual's:

- Scientific knowledge and use of that knowledge to identify questions, acquire new knowledge, explain scientific phenomena and draw evidence-based conclusions about science-related issues
- Understanding of the characteristic features of science as a form of human knowledge and enquiry
- Awareness of how science and technology shape our material, intellectual, and cultural environments
- Willingness to engage in science-related issues and with the ideas of science, as a reflective citizen (OECD 2006, p. 23).

Three abilities, called competencies, are pivotal to the definition of science literacy. The competencies are the ability to identify scientific issues, explain phenomena scientifically, and use scientific evidence. Underlying these abilities are knowledge *of* science and knowledge *about* science.

The three competencies (Table 3) are central to the PISA conception of scientific literacy. Each of the competencies is built on a foundation of cognitive processes, abilities, and knowledge. Cognitive processes and abilities are listed at the top of Table 3. The three competencies are listed, along with their components, in Table 3. The components are stated as behaviors. The competency of using scientific evidence is elaborated further in *Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006* (OECD 2006, p. 30) than the other competencies. Scale scores are presented in the findings for each of the competencies.

Knowledge of science (Table 4) is organized by systems. Within each of these four systems, five to six topics are identified. The system names are associated with science disciplines, physical systems addressing topics in physics and chemistry, living systems addressing topics in biology and environmental science, earth and space systems addressing topics in physical and historical geology and astronomy, and technology systems addressing the relationships between science and technology and concepts and principles related to engineering design. Scale scores for physical systems, living systems, and earth and space systems are contained in the findings.

Knowledge about science (Table 5) addresses topics related to the practice of science. These practices are organized according to those related to scientific enquiry and scientific explanations. Attributes of scientifically valid enquiry, evidence, or explanation are not addressed explicitly as elements of knowledge about science.

The construct of the reflective citizen drives the PISA definition of science literacy. The importance of science and technology in daily life and the pressing issues that must be addressed locally, nationally, and globally underlie the choice of competencies that the authors assert will be required by the reflective citizen to make reasoned decisions regarding these matters. With regard to the knowledge component of scientific literacy, the competencies and the situations in which the reflective citizen will apply them influenced the representation on the PISA assessment of topics in each of the discipline-related systems and technology systems.

The knowledge of science and technology defining scientific literacy was not constrained by curricula in participating nations. Rather, content relevance to life situations, enduring utility, and developmental appropriateness to the target population were design principles for the assessment. Contexts and situations identified as design principles for the assessment are: health, natural resources, environment, hazard, and frontiers of science and technology in personal, social/community, and global contexts (OECD 2006, pp. 12, 17). These design principles have significant implications for the knowledge of science and competencies assessed.

Attitudes toward science are a component of PISA's definition of scientific literacy. The integration of attitudinal questions with questions assessing knowledge and abilities is a unique feature of PISA. The rationale for including attitudinal measures in science assessment is the subject of an in-depth analysis and critique by Loveless (2009). Loveless is critical of the justification in the PISA framework (OECD 2006) for including attitudes toward science on the assessment and of the conclusions drawn from the attitude measures in the analysis of PISA results (OECD 2007). Loveless recommends that "state policymakers should consider [several problems] before benchmarking their assessments to PISA" (2009, p. 17) Potential users of PISA results will be well advised to read Loveless's analysis.

SOME COMPARISONS, NOT JUDGMENTS

Comparisons of the science knowledge and abilities delineated in the 2007 TIMSS (Mullis et al. 2005) and 2009 NAEP (National Assessment Governing Board 2007) assessment frameworks with the knowledge and abilities that constitute scientific literacy as defined by PISA are invertible. In deciding which scores to use for benchmarking or policy, some general idea about the differences in the knowledge and abilities that define science literacy are essential to the decision-making process. Note that the descriptions for the 2007 TIMSS are for the eighth-grade assessment and the 2009 NAEP for the twelfth-grade assessment.[4] While the language used to describe the essential knowledge and abilities is different in the three frameworks, with few exceptions the components are similar. Essential differences are in the assessments, the characteristics of the items, and the testing time devoted to each component.

Table 6 summarizes the knowledge and abilities components of the three assessments. Physical, life, and earth sciences are science literacy components in all three assessments. PISA (OECD 2006), TIMSS (Mullis et al. 2005), and NAEP (National Assessment Governing Board 2007) report performance for each of the three.[5] PISA's knowledge component is different from those of NAEP and TIMSS in two respects. PISA includes knowledge *about* technology and knowledge *about* science as components of science literacy; neither TIMSS nor NAEP include these areas in their knowledge components.

Science-related abilities are called "science practices" in NAEP (National Assessment Governing Board 2007, p. 65). These practices, using science inquiry, describe abilities related to the practice of science, including the abilities to "design or critique aspects of scientific investigations and use empirical evidence to validate or criticize conclusions about explanations and predictions" (National Assessment Governing Board 2007, p. 80). The NAEP framework does not specify the knowledge underlying the practices related to science inquiry. In the TIMSS framework, three cognitive abilities are specified. These are knowing, applying, and reasoning. Each of the cognitive abilities has components associated with the practice of science. For instance, description is a component of knowing; using models is a component of applying; and hypothesizing and predicting are components of reasoning. Inquiry is mentioned in TIMSS at the very end of the science portion of the framework (Mullis et al. 2005, p. 75), acknowledging its importance in some of the countries participating in TIMSS. Each of the assessment frameworks uses different language to describe in the most general terms the knowledge and abilities that comprise science literacy. However, a careful search of each assessment's frameworks will reveal that, with a few exceptions, the same constructs appear in all three frameworks. Attitude is one significant exception.

An essential difference across the three assessments is the measurement of the knowledge and abilities—how much of the assessment time is devoted to each and the characteristics of the items used to evoke the test-takers' performances.

2006 PISA RELEASED ITEMS

When the New York State *Learning Standards for Mathematics, Science, and Technology* (Regents of the State University of New York 1996) were published in 1996, school districts in the state began a major effort to familiarize teachers with them. The teachers politely declined the opportunity, opting to wait until the tests were available. I must admit that I agree with the teachers' position. No matter how detailed standards or test frameworks may be, it is the items that are the reality.[6] Granted, looking at science items may be a daunting prospect for most nonscientists. However, the experience provides essential information about what those scale scores and proficiency-level descriptions really mean. The discussion that follows is based on a limited number of items and on the assumption that the items are representative of those on the operational assessment.

PISA items are contained in science units. Each unit is introduced with some text and, sometimes, diagrams that set the topic and context for the questions that follow. Excluding attitudinal questions, the number of questions in released units ranges from two to five. In addition to the text and diagrams in the unit introductions, the stems for questions within the units are more often than not introduced with additional text and, sometimes, diagrams, which are often graphs. Sometimes questions refer back to information contained in the introductory text and diagrams.

Placing items in context has both costs and benefits. Reading text and analyzing diagrams take testing time. Does that time contribute to better test results? Do test-takers try harder on tests with items that are in contexts that they find interesting? Is the test measuring more about individuals' understanding of context than their science knowledge and abilities? The PISA perspective on scientific literacy emphasizes the importance of science in life situations and applies that emphasis in the design of the assessment. The analysis of the released units provides some information about the implications of that design decision on the information the assessment provides about 15-year-olds' science literacy.

For individuals to use science principles and practices in life situations, they must first understand the principles and be able to apply the practices. While testing for understanding of principles requires placing items in context, the goal is to keep the testing time and ancillary knowledge and abilities (reading, for instance) required to understand the situation at a minimum so that the cognitive demand is on the science principles and abilities—and not on understanding the situation. Instances of situations in the PISA-released items on which an individual may spend testing time unnecessarily are in Starlight (Appendix A) and Ultrasound (Appendix A). In Starlight, the name "Toshio" appears. While it may seem trivial, for some 15-year-olds not familiar with foreign names, figuring out that Toshio is a name may take some testing time. A more substantive example is in Ultrasound. The science principle that Question 8.1 is testing is that distance equals rate times time ($D = r \times t$). The context in which the test-taker is asked to apply the principle is complex, involving an ultrasound wave traveling through, skin, muscle, and fluid to a fetus, where the wave is reflected to a "probe." Contrast this situation to one of a car traveling at a speed of 60 m/s through air. In each instance, application of the principle is placed in context. However, the processing time and knowledge required to decide if the principle is applicable in the case of the ultrasound wave is considerable. Is the purpose of the item to probe understanding of ultrasound waves? Do ultrasound waves behave in the same way as “normal” sound waves? How, if at all, is their speed affected by the medium through which they are traveling? Do they echo in the same way? Test-takers may know the principle and not associate it with ultrasound waves or may decide on the basis of their answers to the questions about ultrasound waves that the principle is not applicable to ultrasound waves. Granted, the situation is relevant and interesting to 15-year-olds. The test-taker may even learn something about ultrasound. (I did.) At issue are the testing time required in setting the context and what exactly the item is testing.

In some instances, the introductory text is not necessary to answer the first question in the Tobacco Smoking unit (Appendix A). The question can be answered easily without the text that precedes it:

Which one of the following is a function of the lungs?

A. To pump oxygenated blood to all parts of your body.

B. To transfer some of the oxygen that you breathe to your blood.

C. To purify your blood by reducing the carbon dioxide content to zero.

D. To convert carbon dioxide molecules into oxygen molecules.

A fair number of questions in the released units are simply measuring knowledge of science principles. These could easily be revised to be assessed as stand-alone items that focus on the principles and to reduce testing time devoted to reading.

Another possible cost of putting units in context is illustrated in the Bread Dough example (Appendix A). As a baker, a chemist, and an individual familiar with yeast and its metabolism, I was puzzled by the text introducing the unit. When I bake bread, I do it under aerobic, not anaerobic, conditions. Does the author of the text make the yeast do its work without oxygen? Was fermentation introduced to get alcohol in the distracters? What are the chances that the 15-year-old taking the test will be knowledgeable about making bread? My guess is that the test-taker more likely will know about using yeast to make beer. Either way, the point is that the question's suggestion that bread is made under anaerobic conditions may confuse the test-taker.

Another consideration relates to the extent to which an assessment measures test-taking skills. Seasoned test-takers might choose not to read introductory text in units but go right to the questions. This is an example of strategic test-taking knowledge in action. These individuals have analyzed the situation and devised a more efficient plan to complete the task that does not involve following a set procedure of reading the item from top to bottom as it appears on the page of the test booklet.

A careful look at the released questions suggests that the language in the assessment framework describing the components of scientific literacy being assessed is more complex than the knowledge and abilities being measured. For instance, of the 51 questions in the released units, 30 are categorized as measuring the competency of explaining phenomena scientifically. Absent further definition of the competency, one might assume that questions measuring that competency would involve the test-taker writing explanations or evaluating the quality of explanations. However, the further description of the competency suggests abilities that are less challenging. These are the following: applying knowledge of science in a given situation; describing or interpreting phenomena scientifically and predicting change; and identifying appropriate descriptions, explanations, and predictions. Of the 30 questions categorized as explaining phenomena scientifically, only 1 required writing an explanation; the other 29 involved applying knowledge of science in a given situation.

The competency of using scientific evidence is applied to 13 of the released questions. This competency involves interpreting scientific evidence and making and communicating conclusions; identifying the assumptions, evidence, and reasoning behind conclusions; and reflecting on the societal implications of science and technological developments. An issue in some of these questions centers on what constitutes scientific evidence. Is the test-taker to assume that all the information in the introductory text, diagrams, and graphs meets the criteria of scientific evidence? Absent any information about sources, the science literate person might reasonably question the scientific validity of the evidence.

The three Lip Gloss questions (Appendix A) are all coded as using scientific evidence. The introduction to this unit contains recipes for lip gloss containing castor oil, beeswax, and palm wax. One question asks how to alter the recipes to make a softer product. Recipes are not scientific evidence. Is the item about analysis and reasoning without any science knowledge requirement? The lip gloss has a larger proportion of oil to wax than the lipstick. Lipstick is harder than lip gloss. Therefore, add more oil to the lipstick. Or is this task more about applying knowledge about the physical properties of castor oil, beeswax, and palm wax than using scientific evidence? You want softer lipstick, so you add more oil. Or is it simply measuring real-world knowledge of a difference between oils and waxes? What is the question designed to measure? What is it actually measuring?

Of the 51 released questions, 9 were coded as identifying science issues.[7] Components of the competency are: recognizing questions that it would be possible to investigate scientifically; identifying keywords to search for scientific information on a given topic; and recognizing key features of a scientific investigation. Released questions required the identification of questions in a list of questions that could be answered by experimentation; the identification of questions in a list of questions that should be answered before an organism is introduced into an environment; identifying the best-designed study from a list of possible studies; identifying reasonable questions for further research; and “identifying keywords to search for scientific information” (OECD 2006, p. 29). The complexity of these items was well matched to the complexity implied in the description of the competency.

One question coded as identifying science issues presented a detailed description of a procedure to study, Stickleback Behaviour (Appendix A), and asked: What is the question this experiment is attempting to answer? Questions of this form can be construed to mean that a description of a procedure without the question that the procedure is designed to answer is acceptable. Granted, the procedure describes what the student wants to investigate; that is, What will make the male stickleback show aggressive behavior? However, such a general statement is not adequate for a well-designed investigation. Further, there is an underlying assumption that only one question can be identified that is consistent with the procedure. These are serious concerns with this question type.[8]

A final observation regarding potentially misleading descriptors relates to questions whose application area was coded “Frontiers of science and technology.” Questions in Science Unit 3, Hot Work (about heat transfer), Lip Gloss, and Bread Dough units were coded “Frontiers of science and technology.” Specific heat, emulsifying agents, and yeast fermentation applied to such mundane situations as heating and cooling of objects, making cosmetics, and leavening bread are not applications at the frontiers of either science or technology.

As noted in the introduction, performance data from large-scale assessments are used to make accountability judgments and for policy development. These judgments rely heavily on scale scores and descriptions of student performance at levels defined by cut points along the scale. The numbers and description have little meaning absent understanding of the constructs measured and assumptions underlying the statistical translation of student performance on items to scale scores. This paper's goal was to give meaning to the scale scores and levels of student proficiency by an analysis of the construct, science literacy, and the relationship between the descriptions of the construct in *Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006* (OECD 2006) and items used to measure components of science literacy. Potential applications of PISA scores to policy or educational practice should give consideration to matters such as testing time devoted to the components of knowledge and abilities contained in the PISA definition of scientific literacy, especially the implications of including knowledge about science, technology, and attitudes as part of the assessment of science literacy. Also, attention needs to be given to how well items match descriptions of the skills and abilities the assessment claims to measure. Ultimately, while the value of science literacy to the individual and to society and the contribution of the various components of science literacy to the areas of application identified in the PISA framework will influence the choice, assessment is a resource-intensive process, and some choices regarding the allocation of resources are parts of the decision-making process.

NOTES

[1] Some would argue that not all professionals or academicians are necessarily literate in the ordinary sense, if being literate in the ordinary sense includes being science literate.

[2] Even within the professions, literacy is on a continuum. For instance, physicians' levels of medical literacy range from capable of sufficient function in the profession, including a history of practice with no malpractice suits, to learned, that is, holding a distinguished chair of medicine.

[3] Knowledge that is assumed is often associated with science abilities, such as the ability to evaluate evidence or the quality of evidence or the ability to construct an explanation. To evaluate the quality of evidence, one must know the criteria for the scientific quality of evidence. Similarly, to construct an explanation, one must know the attributes of a scientific explanation.

[4] In 2008, TIMSS conducted an Advanced Physics Assessment at the twelfth-grade level. Only physics was included in the content domain. The cognitive domain was the same as that for the eighth-grade assessment.

[5] Note that while technology and science are more often than not mentioned in the same sentence in the PISA framework, and technology systems is included as a knowledge component, technology is not reported in the assessment results.

[6] This idea is consistent with that expressed by the philosopher of science, Alfred North Whitehead: that science concepts are defined by the way in which they are measured (Whitehead 1925).

[7] One question was coded both “Explaining phenomena scientifically” and “Identifying scientific issues.”

[8] The example unit in *Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006* (OECD 2006) titled "The School Milk Study" requires identification of questions that are likely research questions for the study.

REFERENCES

American Association for the Advancement of Science. (1993). *Benchmarks for Science Literacy*. Washington, DC: Author.

Burns, T.W., O'Connor, D.J., and Stocklmayer, S.M. (2003). Science communication: A contemporary definition. Public Understanding of Science, *12*, 183-202.

Collins, H., and Pinch, T. (1993). The Golem: What Everyone Should Know About Science. Cambridge: Cambridge University Press.

Gregory, J., and Miller, S. (1998). Science in Public: Communication, Culture, and Credibility. New York: Plenum.

Hilgartner, S. (1990). The Dominant view of popularization: Conceptual problems, political uses. Social Studies of Science, *20*(3), 519-539.

Irwin, A. (2001). Constructing the scientific citizen: Science and democracy in the biosciences. Public Understanding of Science, *10*, 1-18.

Janvier, C. (1987). Problems of Representation in the Teaching and Learning of Mathematics. Hillsdale, NJ: Erlbaum.

Loveless, T. (2009). The 2008 Brown Center Report on Education: How Well Are American Students Learning? Vol. II, No. 3. Washington, DC: Brookings Institution. Retrieved from http://www.brookings.edu/reports/2009/0225_education_loveless.aspx.

Miller, J.D. (1998). The measurement of civic scientific literacy. Public Understanding of Science, *7*, 203-223.

Mullis, I.V.S., Martin, M.O., Ruddock, G.J., O'Sullivan, C.Y., Arora, A., Erberber, E.B. (2005). TIMSS Assessment Frameworks. Boston: International Association for the Evaluation of Educational Achievement. TIMSS and PIRLS International Study Center, Lynch School of Education, Boston College.

National Assessment Governing Board. (2007). *Science Framework for the 2009 National Assessment of Educational Progress* (Prepublication ed.). Washington, DC: National Assessment Governing Board.

National Research Council. (1966). *National Science Education Standards*. Washington, DC: National Academy Press.

Nisbet, M.C., Scheufele, D.A., Shanahan, J.E., Moy, P., Brossard, D., and Lewenstein, B.V. (2002). Knowledge, reservations, or promise? A media effects model for public perceptions of science and technology. *Communication Research*, *29*(5), 584-608.

Organisation for Economic Co-operation and Development. (2006). *PISA 2006 Assessment Framework.* Paris, France: Author. Retrieved from http://www20.gencat.cat/portal/site/Educacio/menuitem.0abe0881c305d9a1c65d3082b0c0e1a0/?vgnextoid=229aa47a169a6110VgnVCM1000008d0c1e0aRCRDandvgnextchannel=229aa47a169a6110VgnVCM1000008d0c1e0aRCRDandvgnextfmt=defaultandnewLang=en_GB.

Organisation for Economic Co-operation and Development. (2006). *PISA 2006 Science Competencies for Tomorrow's World: Vol. 1. Analysis.* Paris, France: Author. Retrieved from http://www.oecd.org/document/2/0,3343,en_32252351_32236191_39718850_1_1_1_1,00.html.

Paivio, A. (1986). *Mental Representations: A Dual Coding Approach*. New York: Oxford University Press.

Regents of the State University of New York. (1996). *Learning Standards for Mathematics, Science, and Technology.* Albany, NY: State Education Department.

Sturgis, P., and Allum, N. (2004). Science in society: Re-evaluating the deficit model of public attitudes. *Public Understanding of Science, 13*(1), 55-74.

Venezky, R.L. (1990). Definition of literacy. In R.L. Venezky, D.A. Wagner, and B.S. Ciliberti (Eds.), *Toward Defining Literacy* (pp. 2-16). Newark, DE: International Reading Association.

Whitehead, A.N. (1925). *Science and the Modern World*. New York: MacMillan.

Wynne, B. (1992). Misunderstood misunderstanding: Social identities and public uptake of science. *Public Understanding of Science, 1*, 281-304.

APPENDIX A

Science Unit 7	Starlight
Science Unit 8	Ultrasound
Science Unit 6	Tobacco Smoking
Science Unit 11	Bread Dough
Science Unit 9	Lip Gloss
Science Unit 5	Stickleback Behaviour

Science Unit 7

Starlight

Toshio likes to look at stars. However, he cannot observe stars very well at night because he lives in a large city. Last year Toshio visited the countryside, where he observed a large number of stars that he cannot see when he is in the city.

Question 7.1

Why can many more stars be observed in the countryside than in large cities?

A. The moon is brighter in cities and blocks out the light from many stars.

B. There is more dust to reflect light in country air than in city air.

C. The brightness of city lights makes many stars hard to see.

D. The air is warmer in cities due to heat emitted by cars, machinery, and houses.

Scoring and Comments on Question 7.1

Full Credit

Code 1: C. The brightness of city lights makes many stars hard to see.

No Credit

Code 0: Other responses

Code 9: Missing

Item type: Multiple choice

Competency: Explaining phenomena scientifically

Knowledge category: Earth and space systems (Knowledge of science)

Application area: Environment

Setting: Social

In this item, students' knowledge of the effect of extraneous light on their ability to resolve light from stars is needed to select the correct response. Consequently, the classification is "Explaining phenomena scientifically—Earth and space systems." The item performed quite well in the field trial, showing adequate discrimination and with minimal evidence of gender or cultural bias. It was answered correctly by about 65% of students.

Science Unit 8

Ultrasound

In many countries, images can be taken of a foetus (developing baby) by ultrasound imaging (echography). Ultrasounds are considered safe for both the mother and the foetus. The doctor holds a probe and moves it across the mother's abdomen. Ultrasound waves are transmitted into the abdomen. Inside the abdomen, they are reflected from the surface of the foetus. These reflected waves are picked up again by the probe and relayed to a machine that can produce an image.

Question 8.1

To form an image, the ultrasound machine needs to calculate the *distance* between the foetus and the probe. The ultrasound waves move through the abdomen at a speed of 140 m/s. What measurement must the machine make so that it can calculate the distance?

Scoring and Comments on Question 8.1

Full Credit

Code 1: It must measure the time the ultrasound wave takes to travel from the probe to the surface of the foetus and reflect back.

- The time of travel of the wave.
- The time.
- Time. Distance = speed / time. (Although the formula is incorrect, the student has correctly identified “time” as the missing variable.)
- It must find when the ultrasound finds the baby.

No Credit

Code 0: Other responses

- The distance.

Code 9: Missing

Item type: Open-constructed response

Competency: Explaining phenomena scientifically

Knowledge category: Physical systems (Knowledge of science)

Application area: Frontiers of science and technology

Setting: Personal

Science Unit 6

Tobacco Smoking

Tobacco is smoked in cigarettes, cigars and pipes. Research shows that tobacco-related diseases kill nearly 13,500 people worldwide every day. It is predicted that, by 2020, tobacco-related diseases will cause 12% of all deaths globally. Tobacco smoke contains many harmful substances. The most damaging substances are tar, nicotine, and carbon monoxide.

Question 6.1

Tobacco smoke is inhaled into the lungs. Tar from the smoke is deposited in the lungs, and this prevents the lungs from working properly.

Which one of the following is a function of the lungs?

A. To pump oxygenated blood to all parts of your body.

B. To transfer some of the oxygen that you breathe to your blood.

C. To purify your blood by reducing the carbon dioxide content to zero.

D. To convert carbon dioxide molecules into oxygen molecules.

Scoring and Comments on Question 6.1

Full Credit

Code 1: B. To transfer oxygen from the air that you breathe to your blood.

No Credit

Code 0: Other responses

Code 9: Missing

Item type: Multiple choice

Competency: Explaining phenomena scientifically

Knowledge category: Living systems (Knowledge of science)

Application area: Health

Setting: Personal

Science Unit 11

Bread Dough

To make bread dough, a cook mixes flour, water, salt, and yeast. After mixing, the dough is placed in a container for several hours to allow the process of fermentation to take place. During fermentation, a chemical change occurs in the dough: the yeast (a single-celled fungus) helps to transform the starch and sugars in the flour into carbon dioxide and alcohol.

Question 11.1

Fermentation causes the dough to rise.

Why does the dough rise?

A. The dough rises because alcohol is produced and turns into a gas.

B. The dough rises because of single-celled fungi reproducing in it.

C. The dough rises because a gas, carbon dioxide, is produced.

D. The dough rises because fermentation turns water into a vapour.

Scoring and Comments on Question 11.1

Full Credit

Code 1: C. The dough rises because a gas, carbon dioxide, is produced.

No Credit

Code 0: Other responses

Code 9: Missing

Item type: Multiple choice

Competency: Explaining phenomena scientifically

Knowledge category: Physical systems (Knowledge of science)

Application area: Frontiers of science and technology

Setting: Personal

Science Unit 9

Lip Gloss

The table below contains two different recipes for cosmetics you can make yourself. The lipstick is harder than the lip gloss, which is soft and creamy.

Lip gloss ingredients: 5 g castor oil 0.2 g beeswax 0.2 g palm wax 1 tsp of colouring substance 1 drop of food flavouring **Instructions:** Heat the oil and the waxes in a container placed in hot water until you have an even mixture. Then add the colouring substance and the flavouring, and mix them in.	**Lipstick ingredients:** 5 g castor oil 1 g beeswax 1 g palm wax 1 tsp of colouring substance 1 drop of food flavouring **Instructions:** Heat the oil and the waxes in a container placed in hot water until you have an even mixture. Then add the colouring substance and the flavouring, and mix them in.

Question 9.1

In making the lip gloss and lipstick, oil and waxes are mixed together. The colouring substance and flavouring are then added. The lipstick made from this recipe is hard and not easy to use.

How would you change the proportion of ingredients to make a softer lipstick?

Scoring and Comments on Question 9.1

Full Credit

Code 1: Responses indicating that you would add less wax and/or add more oil.

- You could use a bit less beeswax and palm wax.
- Add more castor oil.
- Put in 7 g of oil.

No Credit

Code 0: Other responses

- Heat the mixture for longer which will soften it.
- By not heating the waxes as much. (The question asks how you would change the proportion of ingredients.)

Code 9: Missing

Item type: Open-constructed response

Competency: Using scientific evidence

Knowledge category: Scientific explanations (Knowledge about science)

Application area: Frontiers of science and technology

Setting: Personal

The context of cosmetics has everyday relevance for students of this age group, although it could be expected that this unit would generate more interest among females than males. This item can be answered by comparing the quantities of ingredients used in the two recipes to conclude why one recipe produces a softer substance than the other one. The item is therefore classified as "Knowledge about science," category: "Scientific explanations." However, it helps to have knowledge of the properties of the main ingredients (oil and wax) and a case can be made for classifying the item as "Knowledge of science," category: "Physical systems"; competency: "Explaining phenomena scientifically." In the field trial, about 65% of students answered the item correctly, and it displayed good discrimination. Females were much more likely to answer it correctly than were males.

Question 9.2

Oils and waxes are substances that will mix well together. Oils cannot be mixed with water, and waxes are not soluble in water.

Which one of the following is most likely to happen if a lot of water is splashed into the lipstick mixture while it is being heated?

A. A creamier and softer mixture is produced.

B. The mixture becomes firmer.

C. The mixture is hardly changed at all.

D. Fatty lumps of the mixture float on the water.

Scoring and Comments on Question 9.2

Full Credit

Code 1: D. Fatty lumps of the mixture float on the water.

No Credit

Code 0: Other responses

Code 9: Missing

Item type: Multiple choice

Competency: Using scientific evidence

Knowledge category: Scientific explanations (Knowledge about science)

Application area: Frontiers of science and technology

Setting: Personal

This item has less everyday relevance than other items in this unit. Students must reason from the information provided in the stimulus in selecting an appropriate prediction from those on offer. Thus the item is classified as “Knowledge about science,” competency: “Scientific explanations.” About 70% of students answered the item correctly. As with Question 9.1, females were much more likely to answer it correctly than were males.

Question 9.3

When substances called emulsifiers are added, they allow oils and waxes to mix well with water.

Why do soap and water remove lipstick?

A. Water contains an emulsifier that allows the soap and lipstick to mix.

B. The soap acts as an emulsifier and allows the water and lipstick to mix.

C. Emulsifiers in the lipstick allow the soap and water to mix.

D. The soap and lipstick combine to form an emulsifier that mixes with the water.

Scoring and Comments on Question 9.3

Full Credit

Code 1: B. The soap acts as an emulsifier and allows the water and lipstick to mix.

No Credit

Code 0: Other responses

Code 9: Missing

Item type: Multiple choice

Competency: Using scientific evidence

Knowledge category: Scientific explanations (Knowledge about science)

Application area: Frontiers of science and technology

Setting: Personal

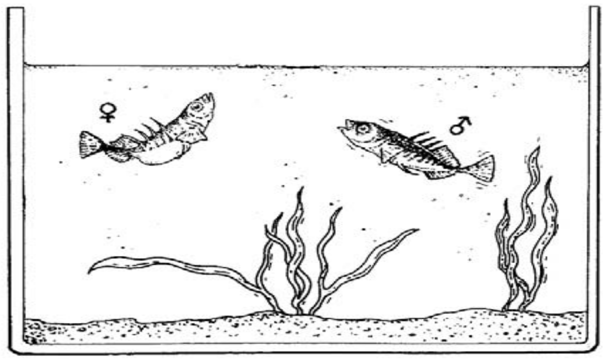
Unlike other items in the unit, there was no discernible difference in performance between males and females on this item in the field trial. Like the previous item, an explanation that accords with the information supplied has to be selected from the four options.

Consequently, this item has the same knowledge and competency classifications. The item performed well in the field trial with good discrimination and had medium difficulty.

Science Unit 5

Stickleback Behaviour

The stickleback is a fish that is easy to keep in an aquarium.

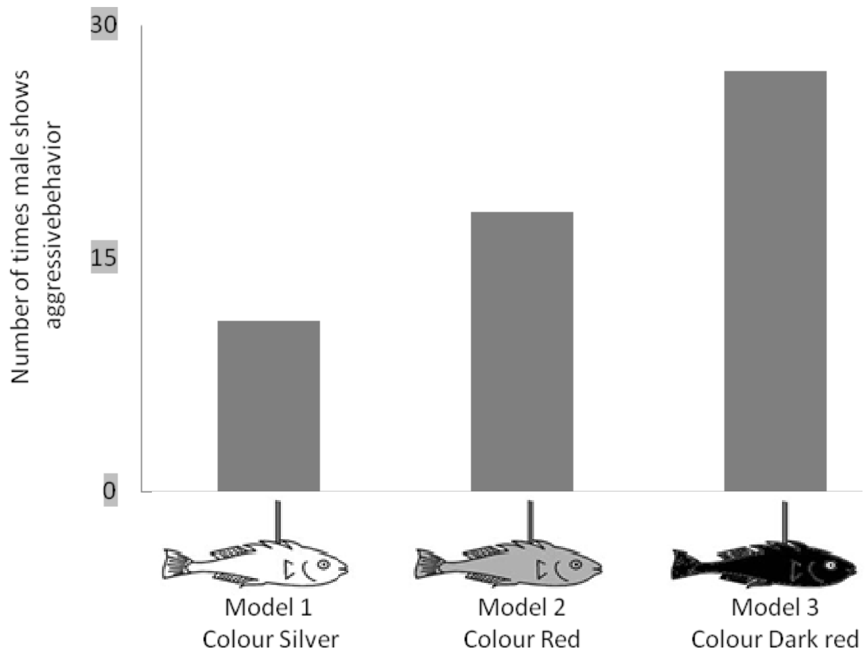


- During the breeding season, the male stickleback's belly turns from silver coloured to red.
- The male stickleback will attack any competing male that comes into his territory and try to chase it away.
- If a silver-coloured female approaches, the male will try to guide her to his nest so she will lay her eggs there.

In an experiment, a student wants to investigate what will make the male stickleback show aggressive behaviour.

A male stickleback is alone in the student's aquarium. The student has made three wax models attached to pieces of wire. He hangs them separately in the aquarium for the same amount of time. Then the student counts the number of times the male stickleback reacts aggressively by pushing against the wax figure.

The results of this experiment are shown below.



Question 5.1

What is the question that this experiment is attempting to answer?

Scoring and Comments on Question 5.1

Full Credit

Code 1: What colour elicits the strongest aggressive behaviour by the male stickleback?

- Does the male stickleback react more aggressively to a red-coloured model than to a silver-coloured one?
- Is there a relationship between colour and aggressive behaviour?
- Does the colour of the fish cause the male to be aggressive?
- What fish colour does the stickleback find most threatening?

No Credit

Code 0: Other responses (including all responses that do not refer to the colour of the stimulus/model/fish).

- What colour will elicit aggressive behaviour in the male stickleback. (No comparative aspect.)
- Does the colour of the female stickleback determine the aggressiveness of the male? (The first experiment is not concerned with the gender of the fish.)
- Which model does the male stickleback react to most aggressively? (Specific reference must be made to the colour of the fish/model.)

Code 9: Missing

Item type: Open-constructed response

Competency: Identifying scientific issues

Knowledge category: Scientific enquiry (Knowledge about science)

Application area: Frontiers of science and technology

Setting: Personal

All relevant information about the experiment is supplied, and hence the "Knowledge about science" classification. The context classification ("Personal"; "Frontiers of science and technology") is in accord with the framework descriptor "expanding one's understanding of the natural world."

In the field trial, the item demonstrated an adequate discrimination, but it was generally difficult, with about 25% of students gaining credit. This unit was not included in the main study because it was considered less relevant to 15-year-olds' daily lives than other units vying for inclusion and because of its high overall reading load.

Table 1. Knowledge Components of Science Literacy

Science Literacy: Knowledge Components	
Disciplinary (including physics, chemistry, biology, earth science)	• Knowledge about the facts, concepts, principles, theories, and laws that scientists apply to the natural world • Knowledge about how science is practiced by scientists • Knowledge about the characteristics of science investigations inquiries, and experiments • Knowledge about the characteristics of scientific evidence • Knowledge about the characteristics of scientific explanations • Knowledge about the history of science • Knowledge about the philosophy of science
Cross-disciplinary (natural sciences)	Knowledge about concepts and practices common across the natural science disciplines (e.g., systems, energy, controlled experiment)
Supporting disciplines	Knowledge related to the application of the language arts, mathematics, and technology (engineering, computer science) to the practice of science

Table 2. Abilities Components of Science Literacy

Science Literacy: Abilities Components		
Abilities	**Activity**	**Scientific Practices**
Apply knowledge Read science text and communicate science ideas (basic literacy applied to science) Reason scientifically	Learning Making personal decisions Making social and civic decisions Making workplace decisions	Design, conduct, critique scientific studies (experiments, investigations, etc.) Compose and critique explanations of natural phenomenon Make and critique predictions Evaluate the quality of evidence (does it meet the standards for scientific evidence?)

Table 3. PISA 2006 Competencies and Their Components (pp. 29–30)

Competencies	
Underlying Cognitive Processes and Abilities	
Identify scientifically oriented issues; describe, explain, or predict phenomena; interpret evidence and conclusions; use scientific evidence to make and communicate decisions; inductive/deductive reasoning; critical and integrated thinking; transforming representations; constructing and communicating arguments and explanations; thinking in terms of models; using mathematics, logic, reasoning, critical analysis	
Competencies	**Components**
Identifying science issues	**Recognizing questions that it would be possible to investigate scientifically**
	Identifying keywords to search for scientific information on a given topic
	Recognizing key features of a scientific investigation
Explaining phenomena scientifically	**Applying knowledge of science in a given situation**
	Describing or interpreting phenomena scientifically and predicting change
	Identifying appropriate descriptions, explanations, and predictions
Using scientific evidence	**Interpreting scientific evidence and making and communicating conclusions**
	Identifying the assumptions, evidence, and reasoning behind conclusions

Competencies	
	Reflecting on the societal implications of science and technological developments
	Accessing scientific information and producing arguments and conclusions Selecting from alternative conclusions in relation to evidence Giving reasons for or against a given conclusion in terms of the process by which the conclusion was derived from the data provided Identifying the assumptions made in reaching a conclusion Reflecting on the societal implications of scientific or technological developments Presenting clear and logical connections between evidence and conclusions or decisions

Table 4. PISA 2006 Categories of Knowledge of Science (pp. 31−33)

Categories of Knowledge of Science
Physical systems
• Structure of matter (e.g., particle model, bonds) • Properties of matter (e.g., changes of state, thermal and electrical conductivity) • Chemical changes of matter (e.g., reactions, energy transfer, acids/bases) • Motions and forces (e.g., velocity, friction) • Energy and its transformation (e.g., conservation, dissipation, chemical reactions) • Interactions of energy and matter (e.g., light and radio waves, sound and seismic waves)
Living systems
• Cells (e.g., structures and function, DNA, plant and animal) • Humans (e.g., health, nutrition, subsystems [i.e., digestion, respiration, circulation, excretion, and their relationship], disease, reproduction) • Populations (e.g., species, evolution, biodiversity, genetic variation) • Ecosystems (e.g., food chains, matter, and energy flow) • Biosphere (e.g., ecosystem services, sustainability)
Earth and space systems
• Structures of the earth systems (e.g., lithosphere, atmosphere, hydrosphere) • Energy in the earth systems (e.g., sources, global climate) • Change in earth systems (e.g., plate tectonics, geochemical cycles, constructive and destructive forces) • Earth's history (e.g., fossils, origin and evolution) • earth in space (*e.g.*, gravity, solar systems)

Categories of Knowledge of Science
Technology systems • Role of science-based technology (e.g., solve problems, help humans meet needs and wants, design and conduct investigations) • Relationships between science and technology (e.g., technologies contribute to scientific advancement) • Concepts (e.g., optimization, trade-offs, cost, risk, benefit) • Important principles (e.g., criteria, constraints, innovation, invention, problem solving)

Table 5. PISA 2006 Categories of Knowledge About Science (p. 33)

Categories of Knowledge About Science
Scientific enquiry
• Origin (e.g., curiosity, scientific questions) • Purpose (e.g., to produce evidence that helps answer scientific questions; current ideas/models/theories guide enquiries) • Experiments (e.g., different questions suggest different scientific investigations, design) • Data type (e.g., quantitative [measurements], qualitative [observations]) • Measurement (e.g., inherent uncertainty, replicability, variation, accuracy/precision in equipment and procedures) • Characteristics of results (e.g., empirical, tentative, testable, falsifiable, self-correcting)
Scientific explanations
• Types (e.g., hypothesis, theory, model, law) • Formation (e.g., data representation, role of extant knowledge and new evidence, creativity and imagination, logic) • Rules (e.g., must be logically consistent; based on evidence, historical and current knowledge) • Outcomes (e.g., produce new knowledge, new methods, new technologies; lead to new questions and investigations)

Table 6. Components of Science Literacy: A Comparison

Components of Science Literacy: A Comparison			
	PISA	**TIMSS**	**NAEP**
Knowledge	**Knowledge of science** • Physical systems • Living systems • Earth and space systems • Technology systems **Knowledge about science** • Scientific enquiry • Scientific explanations	**Content domains** • Biology • Chemistry • Physics • Earth science	**Content statements (Principles)** • Physical science • Life science • Earth and space • Science
Abilities	**Competencies** • Identifying science issues • Explaining phenomena scientifically • Using scientific evidence	**Cognitive domains** • Knowing • Applying • Reasoning	**Practices** • Identifying science • Principles – Using science principles – Using science inquiry – Using technological design