Public Understand. Sci. **7** (1998) 203–223. Printed in the UK PII: S0963-6625(98)93921-2

The measurement of civic scientific literacy

Jon D. Miller

Building on two decades of national surveys in the United States and two Eurobarometer studies, the history, rationale, and structure of a measure of *civic scientific literacy* are described. Estimates of the proportion of adults who are very well informed or moderately well informed on the index of civic scientific literacy appear in the literature more frequently, and this paper provides the first comprehensive description and analysis of the civic scientific literacy measure. It is hoped that this analysis and discussion will encourage the inclusion and replication of the measure in a wider range of studies of the public understanding of and attitudes toward science and technology.

During the last two decades, there has been a growing recognition of the importance of increasing the proportion of citizens who are sufficiently scientifically literate to participate in the resolution of public policy disputes over issues involving science or technology. As the number of scientific and technological issues reaching public visibility and debate has increased in recent decades, the level of governmental and leadership awareness of public illiteracy about scientific and technological matters has increased,[1] and the number of new programs or initiatives has grown proportionally to concern.[2]

Despite this growing level of concern, there has been significantly less debate or agreement about the best methods to measure scientific literacy. And, the debate that has occurred has been primarily at the conceptual level with little or no empirical testing of these conceptualizations. This article will review the major conceptualizations of scientific literacy and explore alternative approaches to the measurement of the scientific literacy construct. In the development of any measure, there is a series of decisions that must be made, and this analysis will attempt to describe and discuss each step of the measurement process.

Data from both the United States and the European Union will be utilized to illustrate the impact of selected measurement approaches. The 1992 Eurobarometer survey of the then 12 member states of the European Union collected interview data from approximately 12,000 European adults, and provides the broadest set of knowledge and attitude measurements collected in Europe to date. The 1995 U.S. study is the most recent in a two-decade series of national surveys and includes the widest range of knowledge and attitude items ever collected about science and technology in the United States.

The concept of civic scientific literacy

The first and most basic conceptual issue concerns the scope of scientific literacy. Drawing from the basic concept of *literacy*, meaning the ability to read and write, scientific literacy

0963-6625/98/030203 + 21$19.50

might be defined as the ability to read and write about science and technology.[3] But, given the wide array of scientific and technical applications in everyday life, scientific literacy might include everything from reading the label on a package of food, to repairing an automobile, to reading about the newest images from the Hubble telescope. Approximately two decades ago, Shen suggested that the public understanding of science and technology might be usefully divided into practical scientific literacy, cultural scientific literacy, and *civic scientific literacy*.[4] In this context, civic scientific literacy refers to a level of understanding of scientific terms and constructs sufficient to read a daily newspaper or magazine and to understand the essence of competing arguments on a given dispute or controversy. Shen argued:

> Familiarity with science and awareness of its implications are not the same as the acquisition of scientific information for the solution of practical problems. In this respect civic science literacy differs fundamentally from practical science literacy, although there are areas where the two inevitably overlap. Compared with practical science literacy, the achievement of a functional level of civic science literacy is a more protracted endeavor. Yet, it is a job that sooner or later must be done, for as time goes on human events will become even more entwined in science, and science-related public issues in the future can only increase in number and in importance. Civic science literacy is a cornerstone of informed public policy.[5]

Through her studies of the nuclear power controversy in Sweden, Nelkin has provided a useful framework for thinking about the content of civic scientific literacy.[6] In the early 1970s, Sweden was seeking to develop a national policy on the use of nuclear power to generate electricity. To facilitate a broader public debate, the Swedish government provided small grants for “study circles” to discuss the nuclear power issue, usually in groups of 10 to 15 citizens with materials and a facilitator to provide a balanced presentation of the points of view. After months of discussions by approximately 80,000 Swedish citizens, the [Swedish] National Board of Civic Information conducted a study and found that the portion of Swedish adults who felt able to make a decision, having heard both arguments set forth, increased from 63 percent prior to the study circles to 73 percent after at least ten hours of study and discussion. Since it is primarily at the point of controversy that the public becomes involved in the resolution of scientific and technological disputes, it is clear that meaningful citizen participation requires a level of civic scientific literacy sufficient to understand the essential points of competing arguments and to evaluate or assess these arguments.[7]

Ziman and Wynne have attacked the basic idea of seeking to define and measure the understanding of scientific concepts, referring to this kind of analysis as based on a “deficit” model.[8] The general argument is that scientific meaning is socially negotiated and that it should not be presumed that the knowledge of scientists is better than the common sense or “local knowledge” of non-scientists. Durant, Evans, and Thomas have provided a thoughtful defense of the idea of defining and measuring public knowledge:

> [T]here remains the problem of stigmatization. Clearly to measure levels of scientific understanding within a population is inevitably to assign higher scores to some individuals than others. By analogy with the notoriously controversial issue of IQ testing, this may be seen as inherently normative. Surely, it may be said, by measuring scientific understanding we are automatically branding as inferior those who score badly? Not at all. It is worth remembering that the French psychologist Alfred Binet developed the IQ test in order to identify those pupils who were most in need of educational assistance. . . [demonstrating] that there is nothing necessarily

> prejudicial about the wish to find out how well individuals are doing in any particular area of educational and scientific attainment.
>
> We do not share Levy-Léblond's apparent willingness to divorce the ideals of democracy and literacy. On the contrary, we believe that the healthy functioning of democracy depends crucially upon the existence of a literate public; and in modern industrial societies, true democracy must embrace scientific literacy.[9]

In a recent discourse on the concept of scientific literacy, Shamos generally accepts the notion of a consumer scientific literacy and a civic scientific literacy, but, reflecting his own training in physics, insists on reserving the label of *true* scientific literacy for those individuals who understand the third law of thermodynamics in essentially the same terms as a physicist.[10] Although Shamos appears to accept the idea of civic scientific literacy at some points in his discourse, he ultimately concludes that citizens can never acquire sufficient understanding to participate in science and technology disputes, and embraces the long-discredited concept of a science court to remove science policy from the democratic process. Unable to step outside his own scientific training, Shamos fails to recognize that the general political institutions of society are extremely reluctant to exclude areas of decision-making from democratic influence, as shown in the uneasy experiment with independent regulatory commissions for securities, trade practices, and communications over the last four decades in the United States. Any effort to exclude science policy from the normal democratic processes would almost immediately foster similar demands for exclusive non-democratic arrangements from numerous other interest groups.

Given the strong likelihood that science and technology policy will remain within the normal democratic policy formulation process in most countries, it is important to develop usable measures of civic scientific literacy to better understand its origins and its function in modern democratic systems. Building on a series of national surveys initiated in 1979, Miller has attempted to build an empirical estimate of the proportion of American adults who qualify as being civic scientifically literate.[11] Since Miller's work reflects the only empirical effort to provide an estimate of the proportion of adults qualifying as civic scientifically literate, this analysis will utilize Miller's work as the basis of a review of the operational definition and measurement of civic scientific literacy.

The issue of dimensionality

One of the first issues in the definition of civic scientific literacy is whether it is a unidimensional or multidimensional construct. This is an issue that needs to be addressed first at the conceptual level, since it has important implications for measurement. In its simplest form, the issue is whether civic scientific literacy is a unidimensional construct, reflecting a set of positively correlated knowledge items, or whether there are distinct and substantively important clusters of knowledge or understanding that should be viewed as separate, but not necessarily independent, dimensions.

Miller has argued that civic scientific literacy is a multi-dimensional construct.[12] In Miller's original 1983 *Daedalus* article, he suggested that civic scientific literacy should be conceptualized as involving three related dimensions: (1) a vocabulary of basic scientific constructs sufficient to read competing views in a newspaper or magazine, (2) an understanding of the process or nature of scientific inquiry, and (3) some level of understanding of the impact of science and technology on individuals and on society. It was argued that the combination of a reasonable level of achievement on each of these three dimensions would reflect a level of understanding and competence to comprehend and follow arguments about science and technology policy matters in the media. In more recent cross-national studies of civic scientific literacy, Miller found the third dimension—the impact of science and technology on individuals and society—to vary substantially in content among different nations and adopted a two-dimensional construct for use in cross-national analyses.[13]

In his early work, Durant recognized a two-dimensional structure for scientific understanding, but opted to use a continuous index of 27 items to measure public understanding for analytic purposes, preferring to avoid use of the literacy concept and the establishment of a threshold that would classify individual respondents as literate and illiterate.[14] In more recent work, Durant and his colleagues have suggested a three-dimensional model, but have continued to utilize only the vocabulary or construct understanding dimension for analysis.[15]

Looking at the last 15 years of empirical work in this area, there appears to be agreement that civic scientific literacy can be usefully conceptualized as a two-dimensional measure, reflecting a vocabulary dimension of basic scientific constructs and a process or inquiry dimension. The desirability and feasibility of using a third dimension that reflects the social impact of science and technology in conceptualizing civic scientific literacy is still a point of some disagreement. There is general agreement among scholars engaged in national surveys, however, that a reliable two-dimensional measure of civic scientific literacy would be useful in a wide range of cross-national research.[16]

The construction of a durable measure

In conceptualizing and developing a measure of civic scientific literacy, it is important to construct a measure that will be useful over a period of years, providing a time-series indicator. If an indicator is revised periodically, it is often impossible to separate the variation attributable to measurement changes from real change over time. The current debate over the composition of consumer price indices in the United States and other major industrial nations is a relevant reminder of the importance of stable indicators over periods of time.[17]

The durability problem can be seen in the early efforts to develop measures of the public understanding of science in the United States. In 1957, the National Association of Science Writers (NASW) commissioned a national survey of public understanding of and attitudes toward science and technology.[18] Since the interviewing for this study was completed only a few months prior to the launch of Sputnik I, it is the only measure of public understanding and attitudes prior to the beginning of the space race. Unfortunately, the four major items of substantive knowledge were (1) radioactive fallout, (2) fluoridation in drinking water, (3) polio vaccine, and (4) space satellites. Twenty years later, at least three of these terms were no longer central to the measurement of public understanding.

Recognizing this problem, Miller attempted to identify a set of basic constructs, such as atomic structure or DNA, that are the intellectual foundation for reading and understanding contemporary issues, but which will have a longer durability than specific terms, such as the fallout of strontium 90 from atmospheric testing. In the late 1970s and the early 1980s, when the National Science Foundation began to support comprehensive national surveys of public understanding and attitudes in the United States, there was little experience beyond the 1957 NASW study in the measurement of adult understanding of scientific concepts. The first U.S. studies relied heavily on each respondent's self assessment of their level of understanding of various terms and concepts, building on a survey research literature that suggested that when respondents are offered a trichotomous set of choices—i.e., do you have a clear understanding of [construct A], a general sense of [construct A], or not much understanding of [construct A]—individuals selecting the clear understanding choice would be very likely to understand the concept, while individuals who were unsure about the concept or who did not understand it might select the middle or lower category.[19] The basic idea was that respondent inflation of their knowledge would occur primarily between the little understanding and general sense categories. This approach, which is still used in national studies in Japan and some other countries, provided useful estimates, but clearly lacks the greater precision provided by direct substantive inquiries.

In a 1988 collaboration between Miller in the United States and Thomas and Durant in the United Kingdom, an expanded set of knowledge items was developed that asked respondents direct questions about scientific concepts. In the 1988 studies, a combination of open-ended and closed-ended items were constructed that provided significantly better estimates of public understanding than had been collected in any prior national study. From this collaboration, a core set of knowledge items emerged that have been used in studies in Canada, China, the European Union, Japan, Korea, New Zealand, and Spain. To a large extent, these core items have provided a durable set of measures of a vocabulary of scientific constructs, with minor additions and deletions over the last decade.[20]

From the outset, it was recognized that the measurement of public understanding of the nature of scientific inquiry was more difficult. The 1988 U.K.–U.S. study utilized a single open-ended inquiry concerning the meaning of “scientific study,” and a joint coding exercise demonstrated the feasibility of double-blind open-ended coding, producing coefficients of reproducibility in the 0.9 range. Subsequently, Durant developed a set of closed-ended process items for use in the Eurobarometer, and Miller and Pifer developed an open-ended sequence for use in a 1993 [U.S.] Biomedical Literacy Study, which was subsequently incorporated into the NSF *Science and Engineering Indicators* series.[21]

The measurement of construct understanding

In the context of a search for durable measures, it is useful to begin with an examination of the development of measures of the public understanding of basic scientific constructs. Many of the measures used over the last decade to measure construct understanding emerged from the 1988 U.K.–U.S. collaboration, which produced a set of open-ended items, several multi-part questions, and a closed-ended true–false quiz. It may be useful to look briefly at some of these core items.

One of the core items that emerged from the 1988 U.K.–U.S. collaboration was an open-ended question concerning DNA. Typical of a series of open-ended questions used in later U.S. studies, the question began with a closed-ended inquiry:

> When you read the term DNA in a newspaper or magazine, do you have a clear understanding of what it means, a general sense of what it means, or little understanding of what it means?

Respondents who indicated that they had either a clear understanding or a general sense of the meaning of DNA were then asked: “Please tell me, in your own words, what is DNA?”

The interviewers—regardless of whether the interview was conducted in-person or over the telephone—were instructed to record the response verbatim, and these text responses were subsequently coded independently by teams of individuals knowledgeable about the definition and meaning of DNA. Standard double-blind coding procedures were employed, and in 1988, full sets of text responses were coded by both American and British coders to assure cross-national comparability. The results of this work demonstrated that double-blind coding practices could produce highly reliable data, and that there were few cross-national variations in coding judgments between the United Kingdom and the United States.

In subsequent U.S. studies, similar open-ended questions have been employed to measure the understanding of basic concepts such as molecule, radiation, acid rain, computer software, and the thinning of the ozone layer around the Earth. In general, open-ended questions provide a better measure of understanding than close-ended questions.

In addition to these open-ended items, a set of multi-part items was first developed in the 1988 U.K.–U.S. collaboration. A two-part question about the movement of the Earth and the Sun has been widely cited in the popular press. In this question, each respondent is asked whether "the Earth goes around the Sun, or the Sun goes around the Earth?" Those respondents who indicate that the Earth goes around the Sun are asked whether the Earth goes around the Sun "once a day, once a month, or once a year?" In 1988, approximately 47 percent of American respondents and 33 percent of British respondents were able to report that the Earth moves around the Sun once each year. The percentage of Americans able to answer this question correctly has remained stable since 1988.

Given the difficulty of asking too many open-ended and difficult questions to respondents, especially in a telephone setting where the respondent can terminate the interview by hanging up the telephone, it is important to use some less stressful forms of inquiry. In the 1988 U.K.–U.S. study, a series of items was constructed for use in a true–false format, with the invitation for a person who was unsure to indicate their uncertainty and continue to the next question. Examples of items in this true–false quiz include:

- Lasers work by focusing sound waves.
- All radioactivity is man-made.
- The earliest human beings lived at the same time as the dinosaurs.
- The center of the Earth is very hot.
- Antibiotics kill viruses as well as bacteria.
- The continents on which we live have been moving their location for millions of years and will continue to move in the future.
- Radioactive milk can be made safe by boiling it.

Finally, a few items were asked as direct inquiries. For example, respondents in the 1988 study and subsequent studies have been asked the question: "Which moves faster, light or sound?"

By using various sets of these construct understanding questions, studies in numerous countries have been able to collect construct knowledge measures. Although the exact items have varied from study to study, the essential point is that each of these sets of items should be viewed as a sample of constructs from a universe of perhaps a hundred or more constructs that are important to civic scientific literacy. The range of constructs developed by Project 2061 provides a useful approximation of the range of substantive concepts that might constitute this universe of relevant constructs.[22]

The process of constructing reliable and comparable measures of construct vocabulary can be understood by examining the construction of these indices for the 1992 Eurobarometer study and the 1995 United States study. Looking first at the relevant items asked in the two studies, a common core of construct knowledge items was asked in both studies, but the U.S. study included more open-ended items and some items were not asked in both studies (see Table 1). An examination of the percent correct on the nine common items suggests that there is little difference between the European Union and the United States on this dimension. Looking at just the nine common items, Americans answered an average of 5.1 questions correctly, and citizens of the European Union answered 4.9 items correctly.

It is possible to limit cross-national comparisons to only those items asked in a comparable manner in each study, but this approach does not utilize the full array of information available from each study and may eliminate the possibility of cross-national comparison when there are minor variations in wording or translation.

Table 1. Vocabulary construct dimension of civic scientific literacy, United States, 1995.

	Percent correct	
	Europe 1992	United States 1995
Provide a correct open-ended definition of a molecule	†	9
Provide a correct open-ended definition of DNA	†	21
Disagree that "Antibiotics kills viruses as well as bacteria"	27	40
Disagree that "Lasers work by focusing sound waves"	36	40
Agree that "Electrons are smaller than atoms"	41	44
Indicate that the Earth goes around the Sun once each year through a pair of closed-ended questions	51	47
Disagree that "The earliest humans lived at the same time as the dinosaurs"	49	48
Disagree that "All radioactivity is man-made"	53	72
Indicate that light travels faster than sound	†	75
Disagree that "Radioactive milk can be made safe by boiling it"	66	61
Agree that "The continents on which we live have been moving their location for millions of years and will continue to move in the future"	82	78
Agree that "The center of the Earth is very hot"	86	78

† Not asked.

An alternative measurement approach utilizes a combination of factor analysis and item-response-theory (IRT) scores to construct a common metric suitable for comparing results across nations. This approach is built on the premise that any set of construct vocabulary items included in a national survey is only a sample of a larger range of items that would be considered important as a part of civic scientific literacy. To the extent that a given sample of items includes an appropriate scope of substantive concepts and reflects a unidimensional structure, a reasonable measure of construct vocabulary may be created. For this purpose, a confirmatory factor analysis represents the best means of assessing the scalar characteristics of a set of items. Applying the same confirmatory factor analysis procedures to the 1992 Eurobarometer and the 1995 U.S. items, a single construct vocabulary factor emerged, but the items that loaded on the factor and the strength of those loadings varied for the two studies (see Tables 2 and 3).[23] Setting aside for the moment the second dimension and focusing only on the construct vocabulary dimension, the results of both confirmatory factor analyses identified nine items in each study that constituted a unidimensional measure of construct vocabulary.

The construct vocabulary dimension for the 1995 U.S. study was anchored by the pair of open-ended questions requiring a definition of a molecule and of DNA, with factor loadings of 0.79 and 0.77, respectively. The two-part question about the movement of the Earth and the Sun had the third highest factor loading (0.69), followed by a pair of true–false questions on lasers and radioactivity that required a "false" response. The speed of light and sound question had a loading of 0.56, and a pair of true–false questions about humans and dinosaurs and about plate tectonics had loadings of 0.46. These nine items illustrate a range of knowledge from basic atomic structure (the definition of a molecule and the whole-part relationship of an atom and an electron) to basic biological constructs (definition of DNA)

Table 2. Dimensions of civic scientific literacy, United States, 1995.

	Construct knowledge dimension	Process knowledge dimension	Proportion of variance explained
Provide a correct open-ended definition of DNA	0.79	—	0.62
Provide a correct open-ended definition of a molecule	0.77	—	0.59
Indicate that the Earth goes around the Sun once each year through a pair of closed-ended questions	0.69	—	0.48
Disagree that "Lasers work by focusing sound waves"	0.65	—	0.42
Disagree that "All radioactivity is man-made"	0.59	—	0.35
Agree that "Electrons are smaller than atoms"	0.59	—	0.34
Indicate that light travels faster than sound.	0.56	—	0.31
Disagree that "The earliest humans lived at the same time as the dinosaurs"	0.46	—	0.22
Agree that "The continents on which we live have been moving their location for millions of years and will continue to move in the future"	0.46	—	0.21
Demonstrate an understanding of experimental logic by selecting a research design and explaining in an open-ended response the rationale for a control group	—	0.83	0.68
Provide an open-ended explanation of the meaning of studying something scientifically	—	0.68	0.46
Demonstrate an understanding of the meaning of the probability of one-in-four by applying this principle to an example of an inherited illness in four separate questions	—	0.63	0.38

$\chi^2 = 91.1/45$ degrees of freedom; Root mean square error of approximation (RMSEA) = 0.02;
Upper limit of the 90% confidence interval for RMSEA = 0.029;
Correlation between factors = 0.86; $N = 2{,}006$. See appendix for correlation matrix.

to earth sciences (plate tectonics). It is not an exhaustive set, but it taps basic scientific constructs from a broad spectrum of scientific disciplines.

The construct vocabulary dimension for the European Union tapped a similar array of basic scientific areas. The European factor included six items that were identical to items in the U.S. factor described above, and three items not included in the U.S. factor. All of the European items were closed-ended. The European construct vocabulary dimension was anchored by a true–false question on plate tectonics (0.70), followed by three true–false questions concerning radioactivity and lasers. The European dimension included three items—making radioactive milk safe by boiling it, antibiotics kill viruses, and the center of the Earth is very hot—that were asked in the 1995 U.S. study, but which did not load on the U.S. vocabulary dimension at the 0.40 level or higher. On balance, the two factors appear to reflect a similar underlying construct vocabulary dimension.

Since the two construct vocabulary factors include a slightly different mix of items, the next task is to create summary measures of these dimensions that are comparable across the two studies, and with other studies that include a unidimensional measure of an underlying construct vocabulary dimension. Multiple-group item-response-theory (IRT) methods, as implemented in the BILOG-MG program,[24] provide a means for computing item values and test scores that take into account the relative difficulty of the items and the different composition of each test and nation.[25] The program places the items from all tests on a common scale by jointly estimating the item parameters and the latent distribution of each group or nation, using a maximum marginal likelihood method. The method is capable of providing reliable results for tests or scales with fewer than ten items.[26]

The basic concept underlying the IRT approach is that the responses to any knowledge

Table 3. Dimensions of civic scientific literacy, European Union, 1992.

	Construct knowledge dimension	Process knowledge dimension	Proportion of variance explained
Agree that "The continents on which we live have been moving their location for millions of years and will continue to move in the future"	0.70	—	0.49
Disagree that "All radioactivity is man-made"	0.69	—	0.48
Disagree that "Lasers work by focusing sound waves"	0.69	—	0.48
Disagree that "Radioactive milk can be made safe by boiling it"	0.66	—	0.43
Disagree that "The earliest humans lived at the same time as the dinosaurs"	0.57	—	0.32
Agree that "The center of the Earth is very hot"	0.56	—	0.31
Disagree that "Antibiotics kill viruses as well as bacteria"	0.54	—	0.29
Indicate that the Earth goes around the Sun once each year through a pair of closed-ended questions	0.51	—	0.26
Agree that "Electrons are smaller than atoms"	0.47	—	0.22
Demonstrate an understanding of the meaning of the probability of one-in-four by applying this principle to an example of an inherited illness in four separate questions	—	0.57	0.33
Indicate that astrology is not at all scientific	—	0.52	0.27
Select a two-group experimental model in a closed-ended question	—	0.34	0.12

$\chi^2 = 32.3/23$ degrees of freedom; Root mean square error of approximation (RMSEA) $= 0.01$;
Upper limit of the 90% confidence interval for RMSEA $= 0.01$;
Correlation between factors $= 0.87$; $N = 12{,}127$. See appendix for correlation matrix.

item will form an item response curve (see Figure 1). Assuming that all respondents taking any given test could be arrayed in an order reflecting their knowledge of the domain being tested, the x-axis of the item response curve is an estimate of knowledge or ability. The y-axis is simply the probability that a respondent will answer the question correctly, given his or her level of knowledge or ability. The item-response curve indicates that few individuals with a low level of knowledge of the domain will be able to answer the hypothetical item in Figure 1, and that virtually all of the respondents with a high level of knowledge will be able to answer this hypothetical question.

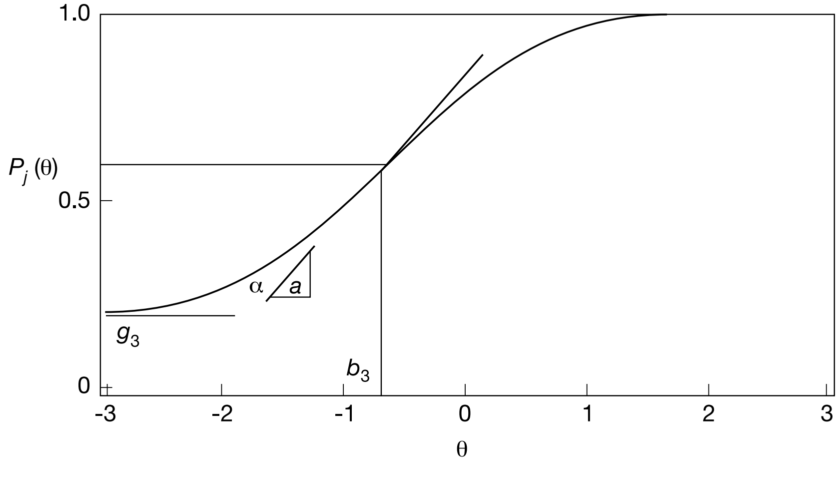

Figure 1. Item-response curve.

For each of the items used in either the 1992 Eurobarometer or the 1995 U.S. study, the BILOG-MG program calculates three separate IRT parameters: a threshold or location estimate, a slope estimate, and a guessing parameter (see Table 4). The threshold parameter is a measure of item difficulty, with higher values meaning that fewer respondents were able to answer it correctly. The slope parameter is an estimate of the measurement efficiency of the item, meaning that there is a positive regression coefficient between the probability of answering a specific question correctly and each individual's total estimated knowledge score. The guessing parameter provides a correction for guessing in closed-ended questions. In the ideal item-response curve shown in Figure 1, the guessing parameter raises the base from a zero correct level to a level that would be obtained by guessing alone with no substantive knowledge of the domain. The item parameters are estimated in a standardized form, assuming a mean of zero and a standard deviation of 1.0 in the combined latent distributions of the groups.

Table 4. IRT parameters for items measuring understanding of basic scientific constructs in the European Union and the United States.

	Slope parameter	Threshold parameter	Guessing parameter
Provide a correct open-ended definition of DNA	1.006	1.191	0.020
Provide a correct open-ended definition of a molecule	1.179	1.902	0.000
Indicate that light travels faster than sound	0.775	−0.872	0.234
Indicate that the Earth goes around the Sun once each year through a pair of closed-ended questions	0.600	0.066	0.077
Disagree that "Lasers work by focusing sound waves"	0.893	0.435	0.018
Disagree that "All radioactivity is man-made"	1.044	−0.185	0.117
Agree that "Electrons are smaller than atoms"	0.535	0.312	0.000
Disagree that "The earliest humans lived at the same time as the dinosaurs"	0.536	−0.074	0.000
Agree that "The continents on which we live have been moving their location for millions of years and will continue to move in the future"	0.750	−1.636	0.000
Disagree that "Radioactive milk can be made safe by boiling it"	0.853	−0.499	0.169
Agree that "The center of the Earth is very hot"	0.863	−1.887	0.000
Disagree that "Antibiotics kill viruses as well as bacteria"	0.534	1.158	0.000

For the computation of item parameters for this set of basic scientific constructs, the responses from all of the adults participating in both the 1992 Eurobarometer and the 1995 U.S. Science and Engineering Indicators Study were utilized, producing one set of item parameters that apply to all of the items asked in either study. The common, or linked, items provide a means for placing the parameters for items asked in only one of the two studies on the same scale as the other items. Since all of the item parameters are placed on a single metric, it is possible to compute comparable scores from each set of items, even though some of the items were asked in only one of the studies.

The computation of individual scores by BILOG-MG utilizes a standardized metric, with a mean of zero and a standard deviation of 1.0 in the combined pool of respondents. This standardized metric, however, is often confusing, since approximately half of the respondents would have a negative score. For expository purposes, Miller has set the mean for the combined pool of European and American respondents to a value of 50, with a standard deviation of 20. In practice, this means that for all respondents within 2.5 standard deviations of this collective mean, the score will vary between zero and 100. The approximately one percent of respondents who might fall outside this range are truncated into the zero to 100 scale. Using this metric, the mean construct vocabulary score for the

European Union is 49.3, and the mean score for the United States is 54.5. These results are similar to the mean percent correct on the nine common items, but the new metric provides more precise individual scoring and allows comparisons with other nations with overlapping, but not identical, sets of knowledge items.[27]

For the analytic identification of individuals capable of following and understanding a public policy dispute over a scientific or technological issue, it is necessary to determine the minimum score that would denote this ability. In previous estimates of civic scientific literacy, Miller has used a threshold level of 67 or more, reflecting the ability of a respondent to get two-thirds of the possible points on the construct vocabulary index.[28] When this standard is applied to the 1995 U.S. data, 27.2 percent of Americans score at or above the 67 point level, compared to 20.2 percent of Europeans.[29] This result suggests that approximately three of four adults in Europe and the United States would be unable to read and understand news or other information that utilized basic scientific constructs such as DNA, molecule, or radiation.

A typology of process understanding

Parallel to the measurement of construct understanding, efforts have been made to develop a measure of the public understanding of the nature of scientific inquiry. Recalling the conceptual discussion above, the basic idea of civic scientific literacy suggests that an individual should understand the empirical basis of scientific inquiry, ideally understanding science as theory building and testing, but minimally as the empirical testing of propositions.[30] The idea that scientific ideas are subjected to empirical scrutiny with the possibility of being falsified is an important component of understanding the nature of scientific inquiry.[31] Although some critics have challenged whether the potential for falsification uniquely defines scientific inquiry, the questioning and ultimate dismissal of, for example, cold fusion claims provides a public demonstration of the continuing empirical foundation of scientific claims.

In his original U.S. studies, Miller utilized a combination of a single open-ended inquiry and a closed-ended question about astrology to identify respondents who held at least a minimal understanding of the process of scientific inquiry.[32] Following the two-part approach described above, respondents were asked whether they have a clear understanding, a general sense, or little understanding of what it means to study something scientifically. Those individuals who reported that they had a clear understanding or a general sense of it were then asked to describe, in their own words, what it means to study something scientifically. The responses were collected verbatim and were coded subsequently using teams of three or more independent coders. In the 1988 U.K.–U.S. study, teams of American and British coders coded all of the responses from both countries, and the final results had an inter-coder reliability coefficient above 0.90. Bauer and Schoon attempted to apply a multi-dimensional coding scheme to these data, but the limited number of probes in the original interviews and the large number of very short answers negates the feasibility of this approach for these data.[33]

In his U.S. studies, Miller classified those individuals with a correct response to this open-ended question and a closed-ended response indicating that astrology is not at all scientific as having a minimal understanding of the process of scientific inquiry. One of the common responses to the open-ended question about the meaning of scientific study was that a scientific study involved doing “an experiment.” Often, this was the only response provided, and it was coded as correct, but Miller and others wanted an expanded measure of the meaning of experimentation. In the 1993 Biomedical Literacy Study, Miller and Pifer were able to introduce a new question concerning experimentation:[34]

> Now, please think of this situation. Two scientists want to know if a certain drug is effective against high blood pressure. The first scientist wants to give the drug to 1000 people with high blood pressure and see how many experience lower blood pressure levels. The second scientist wants to give the drug to 500 people with high blood pressure, and not give the drug to another 500 people with high blood pressure, and see how many in both groups experience lower blood pressure levels. Which is the better way to test this drug?
>
> Why it is better to test the drug this way?

All respondents were asked the follow-up probe, regardless of which group they selected. This decision proved to be useful in assessing the level of understanding. While most of the 17 percent of American adults who selected the single group design in 1995 did not understand the rationale for a control group, a small number of respondents explained that they understood the logic of control groups and placebos, but that they could not ethically withhold medicine from a sick person. Setting aside the ethical argument, it is clear from this response that this small group of respondents—representing approximately one percent of the total sample—did have an adequate level of understanding of experimental logic and they were coded as understanding the nature of scientific inquiry.

Among the 69 percent of individuals who selected the two-group design in 1995, the open-ended probe found substantial misunderstanding of the rationale for experimental design. A majority of this group—representing approximately 40 percent of the total population—indicated that they selected the two-group design so that if the drug “killed a lot of people,” it would claim fewer victims since it would have been administered to fewer subjects. This is hardly the understanding of experimental logic that one would infer from the selection of the two-group design and illustrates one of the hazards of closed-ended questions. Approximately 12 percent of American adults selected the two-group design and were able to explain the logic of control groups. An additional 14 percent of Americans interviewed in the 1995 study selected the two-group design and provided a general rationale that included a “comparison” between the two groups, but lacked the language or logic of control groups.

In Miller’s 1995 analysis, all respondents who selected the two-group approach and who provided a research-oriented response were classified as providing a correct response. In addition, those respondents who selected the one-group approach for ethical reasons, but who demonstrated a minimal knowledge of the logic of experimentation, were also classified as providing a correct response. A total of 27 percent of Americans met this criterion in the 1995 study.

In the 1992 Eurobarometer, none of the open-ended items were included, but three sets of closed-ended items related to an understanding of the nature of scientific inquiry loaded on a second factor in a confirmatory factor analysis in a pattern similar to that described above for the United States (see Table 3). Given the importance of the two-dimension hypothesis to this analysis, it is important to look briefly at each of these three sets of questions.

First, the total 1992 Eurobarometer sample was randomly split into two groups, and a pair of closed-ended questions asked respondents to think about either a medical example or a machine tool example and determine how they would obtain information to assess the effectiveness of a drug or the likely durability of a metal. Each question offered the respondent three choices, which reflected asking the opinion of an expert in the field, using their own scientific knowledge, or doing an experiment. The experimental choice was coded as the correct choice. Approximately 38 percent of European adults provided a correct response.

An additional question was asked in the 1992 Eurobarometer to assess the understanding of experiments, utilizing the first part of the question written for the U.S. Biomedical Literacy Study. Each respondent in the 1992 Eurobarometer was asked:

> Let us imagine that two scientists want to know if a certain drug is effective against high blood pressure. The first scientist wants to give the drug to 1000 people with high blood pressure and see how many of them experience lower blood pressure levels. The second scientist wants to give this drug to 500 people with high blood pressure, and not give this drug to another 500 people with high blood pressure, and see how many in both groups experience lower blood pressure levels. In your opinion, which is the better way?

Approximately 65 percent of European respondents in 1992 selected the two-group model. The 1992 Eurobarometer did not use the follow-up probe employed in the United States in 1993 and 1995, thus it is likely that this response overstates the real level of public understanding of the logic of experimentation.

Although neither of these questions utilized an open-ended probe, the combination of the two items into a single indicator improves the quality of the measure. All respondents who selected the experimental option in the first closed-ended question and who selected the two-group model in the second question were classified as having at least a minimally acceptable level of understanding of experimentation. Approximately 28 percent of European adults qualified as knowledgeable about experimentation in the 1992 Eurobarometer.

Second, a split-ballot approach was employed with a question about the scientific or non-scientific basis of astrology. All respondents in the 1992 Eurobarometer were asked to rate “how scientific” a set of disciplines or activities were, using a scale that ranged from 1 for “not at all scientific” to 5 for “very scientific.” The list included biology, astronomy, history, physics, astrology, economics, medicine, and psychology. A random half of the respondents were given an additional sentence of explanation for each of the disciplines. For example, astronomy was defined as “the study of the heavenly bodies” and astrology was defined as “the study of occult influence of stars, planets, etc., on human affairs.” Nearly 40 percent of European adults interviewed in the 1992 Eurobarometer study indicated that astrology is not at all scientific, but a majority of European respondents thought that there was at least some scientific content in astrology.

Third, a closed-ended question assessed each respondent’s understanding of probability. The question posed a situation in which a doctor “tells a couple that their genetic makeup means that they’ve got a one-in-four chance of having a child with an inherited illness.” Each respondent was then shown a card with the following four choices and asked to select the correct response:

(a) If they have only three children, none will have the illness.
(b) If their first child has the illness, the next three will not.
(c) Each of the couple’s children has the same risk of suffering from the illness.
(d) If their first three children are healthy, the fourth will have the illness.

The (c) response was coded as correct, and 71 percent of European adults selected that choice in the 1992 Eurobarometer study.

To estimate the proportion of Europeans with an understanding of the nature of scientific inquiry, a simple typology was constructed, following the framework employed in the analysis of the U.S. data set. All respondents who demonstrated a minimally acceptable level of understanding of experimentation, who recognized that astrology is not at all scientific, and who provided a correct response to the probability question were coded as understanding the nature of scientific inquiry. Approximately 12 percent of European adults in the 1992 Eurobarometer study met this standard.

Confirmation of basic structure

Given the development of these indicators of construct vocabulary and process understanding, it is important to inquire into the relationship between these two dimensions. While it would be expected that these two measures would be positively correlated, the essential issue is whether there is sufficient differentiation between them to merit treatment as two dimensions. It is now appropriate to return to the two confirmatory factor analyses examined earlier in regard to the item composition and loadings on vocabulary dimension (see Tables 2 and 3).

Looking first at the confirmatory factor analysis results for the United States, the analysis confirms the existence of two factors in the data. The three items described above constitute a second factor, reflecting an understanding of the process of scientific inquiry. All three items are either open-ended or multi-part questions, and all three have loadings at the 0.6 level or higher (see Table 2). The two factors are correlated at 0.86, indicating that they are closely related, but sufficiently differentiated to reflect separable dimensions.

Although all of the three process items from the Eurobarometer loaded on a separate identifiable factor, they loaded at somewhat lower levels (see Table 3). These weaker loadings would appear to reflect the closed-ended character of all three of the questions, although the multi-part question on probability is similar to the item used in the U.S. study. The process factor is correlated with the vocabulary dimension at the 0.87 level in the Eurobarometer data—virtually identical to the pattern found in the United States.

In addition to the separation of these dimensions in the two confirmatory factor analyses, in-depth analyses of the [U.S.] National Biomedical Literacy Study by Miller and Pifer have shown that the construct vocabulary dimension and the understanding of scientific inquiry dimension have different relationships with selected attitudes toward science and technology.[35] It is important to continue to examine the variables associated with the development of a threshold level of understanding on each dimension and the impact of each of these dimensions on subsequent attitudes and behaviors.

The computation of results

Working from the two-dimensional model described above, how might these dimensions be used to provide a single estimator of the level of civic scientific literacy? Conceptually, individuals who demonstrate a high level of understanding on both dimensions would be expected to be the most capable of acquiring and comprehending information about a science or technology policy controversy, and these individuals will be referred to as being “well informed,” or “civic scientifically literate.” At the same time, individuals who demonstrate either an adequate vocabulary of scientific constructs or who display an acceptable level of understanding of the nature of scientific inquiry would be expected to be relatively more capable of receiving and utilizing information about a science or technology policy dispute than citizens who understand neither dimension. This second group will be referred to as “moderately well informed” or “partially civic scientifically literate.” In the 1995 study, 12 percent of American adults qualified as well informed, or civic scientifically literate, and approximately 25 percent qualified as moderately well informed (see Table 5).

Table 5. Estimated percentage of adults qualifying as civic scientifically literate in the European Union and the United States, by component and total.

Country	Construct vocabulary score 67+	Understand scientific inquiry	Civic scientific literacy			*N*
			Literate	Partial	Not	
United States	27	21	12	25	63	2,006
European Union	20	12	5	22	73	12,147
Britain	27	20	10	26	64	1,000
Denmark	26	15	8	25	67	1,000
Netherlands	27	15	8	27	65	1,000
Italy	18	14	5	22	73	1,000
France	21	12	4	24	72	1,000
Germany	23	8	4	24	72	2,000
Spain	17	7	3	17	80	1,000
Belgium	16	10	3	20	77	1,000
Ireland	14	8	3	16	81	1,000
Greece	7	7	1	11	88	1,000
Portugal	8	2	1	8	91	1,000

Following the same procedure, all European respondents who earned a score of 67 or more on the Index of Scientific Construct Knowledge and who demonstrated at least a minimally acceptable level of understanding of the nature of scientific inquiry were classified as well informed, or civic scientifically literate. Individuals who qualified on one dimension, but not the other, were classified as moderately well informed. On the basis of the 1992 Eurobarometer, five percent of European adults were well informed, or civic scientifically literate, while an additional 22 percent were moderately well informed. Within Europe, the percentage of adults qualifying as civic scientifically literate ranged from 10 percent in Britain to one percent in Portugal.

The interpretation and use of estimates

Having reviewed the development of a set of estimates of the proportion of adults in the European Union and the United States who are civic scientifically literate, what conclusions should be drawn from these estimates and how might they be used to improve the formulation of public policy concerning issues with significant scientific or technological content? Given the technical nature of much of the preceding discussion of the construction of the estimates, it is appropriate to turn first to some methodological issues. This analysis will conclude with a discussion of the substantive and research implications of these estimates of civic scientific literacy.

Methodological issues

The last three decades have witnessed a substantial growth in the sophistication and accuracy of statistical tools to study the structure and content of measurements of knowledge. The great majority of this work has focused on the construction of high-stakes tests for populations of students at the pre-collegiate and collegiate levels, including the Scholastic Aptitude Test (SAT), the American College Test (ACT), and the Graduate Record Examination (GRE). Beginning with the work of Frederick Lord[36] of the Educational Testing Service in the 1960s, the technology for developing multiple versions of tests with a common metric has grown steadily, and the parallel growth of the accessibility and speed of microprocessors has made this technology widely available throughout the world.

Unfortunately, virtually no efforts have been made to apply this technology to the smaller knowledge item sets commonly collected in educational and social science surveys. The work reported above represents the first attempt to utilize the IRT technology for the construction of reliable cross-national estimates of the public understanding of science and technology. The relatively recent development of new software that allows the estimation of IRT parameters and scores for multiple groups has made the technology easily accessible for cross-national comparisons. Additional work is now underway to demonstrate the utility of the IRT technology in the construction of a metric for time series measures of civic scientific literacy in the United States.

A parallel issue concerns the use of open-ended and closed-ended items in the measurement of substantive knowledge, especially in the measurement of the public understanding of science. Although the availability of a guessing parameter in the IRT technology reduces the impact of respondent guessing on true–false or short multiple-choice questions, the preceding analysis demonstrated the stronger factor loadings and the higher IRT threshold values for open-ended questions. The collection and coding of open-ended responses requires a skillful field organization, and multiple languages create additional complications. Yet, despite the extra effort required in data collection and coding, the higher quality of the responses argues for the inclusion of some open-ended items in future cross-national studies of the public understanding of science. The protocols for multiple blind coding procedures are well-established, as are the procedures for translation and back-translation.

Finally, the future application of the IRT technology to cross-national studies of the public understanding of science requires the inclusion of some linkage items in each study, thus requiring continued coordination among the scholars and sponsoring organizations in the major industrial countries. To date, this coordination has been done primarily through the International Council for the Comparative Study of the Public Understanding of Science and Technology, which is a non-governmental group of scholars who have directed most of the studies in this area. The longer term stability of this work may require some sponsorship from established international organizations.

Implications for public policy

As the number and importance of public policy issues involving science or technology continues to increase, the attainment of an adequate level of civic scientific literacy becomes more critical to the long-term health of democracy. Looking ahead to the next 50 years, there can be little doubt that the number of public policy issues requiring some level of civic scientific literacy will increase, and increase markedly. The biotechnology revolution is at hand, and the number of public policy issues emanating from this technology alone will be larger than all of the science and technology-related public policy issues in the past. There is near unanimous agreement within the scientific community that modern societies will need to transition from fossil-based energy systems to new energy sources within the next century, and there will be numerous important public policy controversies associated with that transition. The continuing deterioration of the Earth’s environment will foster a broad array of public policy issues, almost certainly with more urgency than in previous decades. And, in the spirit of the present scientific revolution, there will almost surely be major public policy controversies over scientific and technical issues that cannot be imagined at this time.

While a full discussion of the impact of political specialization in modern political systems is beyond the scope of this article, it is clear that most public policy disputes involving science or technology are too specialized to become a major factor in electoral contests.[37] The exact processes through which the informed public will participate in the formulation of public policy will vary across political systems and new forms of participation may emerge from the current revolution in electronic communications. Regardless of the mode of participation, there can be little doubt that healthy democratic systems will require a significant number of citizens to be civic scientifically literate. Whether the optimal proportion of citizens who are civic scientifically literate is 20 percent, 30 percent, or more, there can be little doubt that the current levels of civic scientific literacy are too low in both the United States and the eleven European Union countries included in this analysis.[38]

The evidence suggests that the most effective path to a higher proportion of civic scientifically literate citizens is the improvement of pre-university and university education. The primary work in this area has been conducted in the United States, but the results of the recent Third International Mathematics and Science Study (TIMSS) point to the essential role of early learning in the development of competence in science and mathematics.[39] Again, a full discussion of the strategies for increasing the proportion of citizens who are civic scientifically literate is beyond the scope of this article, but it must be a part of any serious discussion of public policies to foster higher levels of public understanding of science and technology.

The need for continued research

The preceding review of the conceptualization and measurement of civic scientific literacy has demonstrated the feasibility of using current statistical technologies to build reliable cross-national measures of substantive knowledge. As suggested by the evolution of these measurement technologies over the last three decades, there are important methodological issues that merit continued investigation. And, the development of this measure of civic scientific literacy raises a wide range of substantive questions about its role and influence in the communications process and in the political process. It may be useful to discuss briefly some of the methodological and substantive issues that need continued research.

As noted above, this analysis is one of the first applications of the IRT technology to cross-national measures of substantive knowledge in adult populations utilizing survey interview data. In other analyses, Miller has extended this approach to 14 countries,[40] but there is a need to explore the application of the IRT technology to other adult survey data sets on other substantive areas. While the number of items available for this analysis was adequate to meet the minimal statistical requirements, it will be important to explore the optimal number of total items and the number of linkage items for IRT comparisons.

Similarly, while the preceding analysis demonstrated the feasibility of using a combination of open-ended and closed-ended items in the same analysis, future studies using this approach should seek to explore the relative impact of open-ended and closed-ended items on estimates and to explore various combinations of open-ended and closed-ended measures in obtaining cost-effective cross-national measures of adult understanding of science and technology. With the recent availability of digital technology that allows the capture of the verbatim response of respondents in telephone surveys, there are new opportunities to examine alternative approaches to the coding and classification of open-ended responses.

Substantively, it is essential to design studies to examine the sources and kinds of adult learning about science and technology throughout the life cycle. Current cross-sectional data suggest that individuals who have obtained a better understanding of science and technology through formal schooling tend to retain and enrich that understanding through the use of informal learning resources such as libraries, newspapers, magazines, television shows, and museums. It is important to learn more about the magnitude and dynamics of these adult learning processes, and about adults' selection of and trust in various kinds of communications.

In regard to participation in public policy disputes involving science or technology, studies need to focus on the impact of the political specialization process, the scope of citizen interest in selected scientific and technical controversies, and the role of civic scientific literacy in the decision about whether and how to participate in the particular dispute. While there have been some innovative studies of public participation in foreign policy disputes, there has been virtually no research focused on the role of civic scientific literacy, or other measures of knowledge, in this process.[41]

Finally, it is important to design panel studies to explore the stability of civic scientific literacy and how it might change during the course of a controversy. Miller's panel study of public attitudes toward space and the change of these attitudes during the Challenger disaster is the only study to follow a panel of individuals through a particular event or dispute.[41] Apart from disasters, it is essential to begin some panel studies that parallel the introduction of a new technology or issue and follow the behavior of large samples of adults as they become aware of the issue, acquire dispute-relevant information, assimilate new information into previous schemas, form substantive attitudes on the issue, and make decisions about the level of participation in the resolution of the issue or dispute that they wish to make. Most of the interesting questions about human behavior involve some understanding of the origins and sources of change, and the best measures of change will ultimately be obtained by measuring the same individuals periodically over some span of time.

Appendix

Table A1. Correlation matrix for the factor analysis displayed in Table 2 (United States).

	SCISTUDY	DNA	MOLECULE	KNOWEXP	RADIOACT	LASERS
SCISTUDY	1.00					
DNA	0.57	1.00				
MOLECULE	0.39	0.46	1.00			
KNOWEXP	0.59	0.55	0.53	1.00		
RADIOACT	0.30	0.46	0.51	0.37	1.00	
LASERS	0.34	0.52	0.44	0.38	0.50	1.00
ELECTRON	0.38	0.47	0.46	0.40	0.30	0.37
PLATETEC	0.24	0.41	0.30	0.43	0.23	0.40
DINOSAUR	0.30	0.38	0.35	0.40	0.23	0.30
LIGHT	0.33	0.38	0.49	0.18	0.36	0.44
EARTHSUN	0.39	0.49	0.52	0.48	0.31	0.43
UNDPROB	0.41	0.48	0.37	0.48	0.39	0.24
	ELECTRON	**PLATETEC**	**DINOSAUR**	**LIGHT**	**EARTHSUN**	**UNDPROB**
ELECTRON	1.00					
PLATETEC	0.26	1.00				
DINOSAUR	0.24	0.28	1.00			
LIGHT	0.28	0.20	0.19	1.00		
EARTHSUN	0.35	0.26	0.33	0.41	1.00	
UNDPROB	0.32	0.22	0.20	0.23	0.40	1.00

Table A2. Correlation matrix for the factor analysis displayed in Table 3 (European Union).

	HOTCORE	RADMILK	ELECTRON	PLATETEC	DINOSAUR	ANTIBIO
HOTCORE	1.00					
RADMILK	0.37	1.00				
ELECTRON	0.36	0.25	1.00			
PLATETEC	0.54	0.40	0.44	1.00		
DINOSAUR	0.32	0.37	0.26	0.39	1.00	
ANTIBIO	0.26	0.35	0.20	0.28	0.36	1.00
LASERS	0.35	0.46	0.32	0.36	0.41	0.37
RADIOACT	0.40	0.44	0.38	0.48	0.38	0.39
EARTHSUN	0.39	0.29	0.38	0.35	0.22	0.18
UNDPROB	0.34	0.34	0.24	0.43	0.28	0.27
ASTROL	0.25	0.21	0.21	0.27	0.21	0.24
KNOWEXP	0.19	0.19	0.16	0.30	0.16	0.18
	LASERS	RADIOACT	EARTHSUN	UNDPROB	ASTROL	KNOWEXP
LASERS	1.00					
RADIOACT	0.53	1.00				
EARTHSUN	0.29	0.35	1.00			
UNDPROB	0.32	0.30	0.31	1.00		
ASTROL	0.32	0.37	0.23	0.21	1.00	
KNOWEXP	0.19	0.27	0.15	0.31	0.18	1.00

Acknowledgments

The author wishes to acknowledge the support of the National Science Foundation in the conduct of the *Science and Engineering Indicators* surveys used in this analysis (NSF contract SRS-9217876). All views and conclusions expressed in this analysis are the responsibility of the author and do not necessarily reflect the views of the National Science Foundation.

References

1 See P. D. Hurd, “Science Literacy: Its Meaning for American Schools,” *Educational Leadership* **16** (1958): 13–16; B. V. Lewenstein, “The Meaning of ‘Public Understanding of Science’ in the United States after World War II,” *Public Understanding of Science* **1** (1992): 45–68; K. Prewitt, “Scientific Illiteracy and Democratic Theory,” *Kettering Review*, Summer (1985), 32–41; The Royal Society, *The Public Understanding of Science* (London: Royal Society, 1985); G. Sapp, “Science Literacy: A Discussion and Information-based Definition,” *College and Research Libraries*, January (1992), 21–29; M. Shortland, “Advocating Science: Literacy and Public Understanding,” *Impact of Science on Society* **152** (1988): 305–316.

2 W. Bodmer and J. Wilkins, “Research to Improve Public Understanding Programmes,” *Public Understanding of Science* **1**, (1992): 7–10; [U.S.] National Science Board, *Science and Engineering Indicators—1996* (Washington: Government Printing Office, 1996).

3 D. P. Resnick and L. B. Resnick, “The Nature of Literacy: An Historical Exploration,” *Harvard Educational Review* **47** (1977): 370–85; D. Harman, “Illiteracy: An Overview,” *Harvard Educational Review* **40** (1970): 226–30.

4 B. S. P. Shen, “Scientific Literacy and the Public Understanding of Science,” in *Communication of Scientific Information*, ed. S. Day (Basel: Karger, 1975), 44–52

5 Shen, “Scientific Literacy and the Public Understanding of Science,” p. 49.

6 D. Nelkin, *Technological Decisions and Democracy: European Experiments in Public Participation* (Beverly Hills, CA: Sage, 1977).

7 J. D. Miller, *The American People and Science Policy* (New York: Pergamon Press, 1983).

8 J. Ziman, “Public Understanding of Science,” *Science, Technology & Human Values* **16** (1991): 99–105; B. Wynne, “Knowledges in Context,” *Science, Technology & Human Values* **16** (1991): 111–121; A. Irwin and B. Wynne, eds., *Misunderstanding Science? The Public Reconstruction of Science and Technology* (Cambridge, U.K.: Cambridge University Press, 1996).

9 J. Durant, G. Evans, and G. Thomas, “Public Understanding of Science in Britain,” *Public Understanding of Science* **1** (1992): 161–182.

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

11 See J. D. Miller, “Scientific Literacy: A Conceptual and Empirical Review,” *Daedalus* **112** (1983): 29–48; J. D. Miller, “Scientific Literacy in the United States,” in *Communicating Science to the Public*, ed. D. Evered and M. O’Connor (London: Wiley, 1987); J. D. Miller, “Scientific Literacy for Effective Citizenship,” in *Science/Technology/Society as Reform in Science Education*, ed. R. E. Yager (New York: State University Press of New York, 1995).

12 *Ibid.*

13 J. D. Miller, R. Pardo, and F. Niwa, *Public Perceptions of Science and Technology: A Comparative Study of the European Union, the United States, Japan, and Canada* (Madrid: BBV Foundation, 1997).

14 See J. R. Durant, G. A. Evans, and G. P. Thomas, “The Public Understanding of Science,” *Nature*, **340** (1989): 11–14; J. R. Durant, G. A. Evans, and G. P. Thomas, “Public Understanding of Science in Britain: The Role of Medicine in the Popular Presentation of Science,” *Public Understanding of Science* **1** (1992): 161–182; G. A. Evans and J. R. Durant, “The Relationship between Knowledge and Attitudes in the Public Understanding of Science in Britain,” *Public Understanding of Science* **4** (1995): 57–74.

15 M. Bauer, J. R. Durant, and G. A. Evans, “European Public Perceptions of Science,” *International Journal of Public Opinion Research* **6** (1994): 163–186.

16 Miller, Pardo, and Niwa, *Public Perceptions of Science and Technology*.

17 D. Fixler, “The Consumer Price Index: underlying concepts and caveats,” *Monthly Labor Review* **116** (1993): 3–12; B. R. Moulton, “Bias in the Consumer Price Index: What is the Evidence?” *Journal of Economic Perspectives* **10** (1996): 159–77.

18 R. C. Davis, *The Public Impact of Science in the Mass Media* (Ann Arbor: University of Michigan Institute for Social Research, 1958).

19 See A. N. Oppenheim, *Questionnaire Design and Attitude Measurement* (New York: Basic Books, 1966); P. J. Labaw, *Advanced Questionnaire Design* (Cambridge, MA: Abt Books, 1980); J. M. Tanur, ed., *Questions about Questions: Inquiries into the Cognitive Bases of Surveys* (New York: Russell Sage Foundation, 1992).

20 J. D. Miller, “Public Understanding of Science and Technology in OECD Countries: A Comparative Analysis” (paper presented to the OECD Symposium on the Public Understanding of Science and Technology, Tokyo, Japan, November 1996); Miller, Pardo, and Niwa, *Public Perceptions of Science and Technology*.

21 National Science Board, *Science and Engineering Indicators—1996*.

22 Project 2061, American Association for the Advancement of Science, *Benchmarks for Science Literacy* (New York: Oxford Univ. Press, 1993).

23 A confirmatory factor analysis is a technique for testing a hypothesis about the relationships among a set of items. In this case, it was hypothesized that civic scientific literacy was organized into two dimensions, reflecting construct understanding and process understanding. This technique examined the correlations among all of the items and identified two sets of items—factors—with high intercorrelations among the items on each dimension, but not necessarily with the items on the other dimension. In this case, the confirmatory factor analysis found that the two factors were positively correlated. This note will not be repeated on subsequent tables describing the results of other confirmatory factor analyses utilized in this analysis.

24 M. F. Zimowski, E. Muraki, R. J. Mislevy, and R. D. Bock, *BILOG-MG: Multiple-group IRT Analysis and Test Maintenance for Binary Items* (Chicago: Scientific Software International, 1996).

25 R. D. Bock and M. F. Zimowski, “Multiple-group IRT,” in *Handbook of Modern Item Response Theory*, ed. W. J. van der Linden and R. K. Hambleton (New York: Springer-Verlag, 1997), 433–448.

26 R. D. Bock and M. Aitkin, “Marginal Maximum Likelihood Estimation of Item Parameters: Application of an EM-algorithm,” *Psychometrika* **46** (1981): 443–459.

27 Miller, Pardo, and Niwa, *Public Perceptions of Science and Technology*.

28 Miller, “Scientific Literacy in the United States”; Miller, “Scientific Literacy for Effective Citizenship”; Miller, Pardo, and Niwa, *Public Perceptions of Science and Technology*.

29 In a separate analysis of 14 industrial nations, Miller found the proportion of adults meeting this standard within the European Union ranged from 27.2 percent in the Netherlands to 6.6 percent in Greece. See Miller, “Public Understanding of Science and Technology in OECD Countries.”

30 T. S. Kuhn, *The Structure of Scientific Revolutions* (Chicago: University of Chicago Press, 1962).

31 K. R. Popper, *The Logic of Scientific Discovery* (New York: Basic Books, 1969).

32 Miller, "Scientific Literacy: A Conceptual and Empirical Review"; Miller, "Scientific Literacy in the United States"; Miller, "Scientific Literacy for Effective Citizenship"; Miller, Pardo, and Niwa, *Public Perceptions of Science and Technology*.

33 See M. Bauer and I. Schoon, "Mapping Variety in Public Understanding of Science," *Public Understanding of Science* **2** (1993): 141–155, and J. D. Miller, "Theory and Measurement in the Public Understanding of Science: A Rejoinder to Bauer and Schoon," *Public Understanding of Science* **2** (1993): 235–243.

34 J. D. Miller and L. K. Pifer, *The Public Understanding of Biomedical Science in the United States, 1993* (Chicago: Chicago Academy of Sciences, 1995).

35 *Ibid.*

36 See F. M. Lord, "A Theory of Test Scores," *Psychometric Monographs*, no. 7 (1952); F. M. Lord, *Applications of Item Response Theory to Practical Testing Problems* (Hillsdale, NJ: Erlbaum, 1980).

37 For a discussion of political specialization, see G. A. Almond, *The American People and Foreign Policy* (New York: Harcourt, Brace and Company, 1950); J. D. Miller, *The American People and Science Policy* (New York: Pergamon Press, 1983); J. A. Rosenau, *Citizenship Between Elections* (New York: Free Press, 1974).

38 The Eurobarometer survey interviews at least 1000 respondents in each member country, except that only 500 respondents are interviewed in Luxembourg and Northern Ireland. Since some items are asked in a split-ballot format, the effective number of responses for some questions drops to 250 in these two areas. For reasons of statistical reliability, these two countries (areas) were omitted from the analysis.

39 U.S. Department of Education, National Center for Education Statistics. *Pursuing Excellence* (Washington, D.C.: U.S. Government Printing Office, 1996); A. E. Beaton *et al.*, *Mathematics Achievement in the Middle School Years* (Boston: Boston College, 1996); A. E. Beaton *et al.*, *Science Achievement in the Middle School Years* (Boston: Boston College, 1996).

40 Miller, "Public Understanding of Science and Technology in OECD Countries."

41 Rosenau, *Citizenship Between Elections.*

42 J. D. Miller, "The Challenger Accident and Public Opinion," *Space Policy* **3** (1987): 122–140.

Author

Jon D. Miller is Professor of Political Science at Northern Illinois University and Vice President of the Chicago Academy of Sciences. He serves as Director of the International Center for the Advancement of Scientific Literacy and has studied the public understanding of and attitudes toward science and technology for 25 years. He also directs the Longitudinal Study of American Youth, which examines the development of scientific and quantitative literacy in middle school, high school, and college-aged young adults in the United States. He can be reached at 2060 North Clark Street, Chicago, IL 60614, USA, e-mail: jdmiller@icasl.org.