Measurement and Operationalization of the "Science in the Service of Citizens and Consumers" Framework

REPORT OF THE 2010 NSF SCIENCE INDICATORS INSTRUMENTATION WORKSHOP

Prepared by:

Thomas M. Guterbock

with Nick Allum, John Besley, Frederick Conrad, Allyson Holbrook, Scott Keeter, Susan Losh, Jeff Mondak, Bryce Reeve, Deborah Rexrode, David Sikkink, Sally Stares, Roger Tourangeau, Chris Toumey

January 31, 2011

Table of Contents

List of Figures

Measurement and Operationalization of the "Science in the Service of Citizens and Consumers" Framework

REPORT OF THE 2010 NSF SCIENCE INDICATORS INSTRUMENTATION WORKSHOP

Introduction

NSF arranged with Chris Toumey of the NanoCenter at the University of South Carolina, and Thomas Guterbock, Director of the Center for Survey Research at the University of Virginia to conduct two scientific workshops during Fall 2010 to evaluate the questions that the National Science Board has used over the years as indicators of public knowledge of science.[1] The first workshop focused on the concept of adult science knowledge underlying the survey items, including the items that involve both scientific knowledge and religious conviction. The second workshop examined the 'instrumentation' of the survey, that is, the measurement adequacy of the items themselves. This is the report of the instrumentation workshop.

Background

Since 1979, NSF has sponsored recurrent surveys of the American public to gauge their understanding and awareness of science, sometimes referred to as scientific literacy. The results have appeared in *Science and Engineering Indicators*. However, the 2010 edition of *Science Indicators* omits the results from two true-false items because the National Science Board judged that they "don't properly reflect what Americans know about science and, thus, are misleading" (Bhattacharjee 2010).

At issue in particular were items in the survey that deal with evolution and the creation of the universe: (1) "Human beings as we know them today, developed from earlier species of animals," and (2) "The universe began with a huge explosion."[2] The chair of the NSB's SEI committee, Dr. Louis Lanzerotti, is quoted as saying that the questions are "flawed indicators of scientific knowledge because the responses conflated knowledge and beliefs" (Bhattacharjee 2010).

The action of the NSB in omitting these results from its annual report has drawn criticism from some science officials and social scientists who study scientific literacy; others find the

---

[1] We will refer to the changing sets of survey items that cover science knowledge and/or science literacy as "public science knowledge" items or questions.
[2] We will refer to these items as the "evolution" and "big-bang" questions, respectively.

action to be defensible. Echoing through the controversy is the long-standing issue in the United States of how conflicts between scientific discovery and religious belief should be resolved in education and in the public square, and, in particular, the continued and often acrimonious public debate about how evolution vs. 'creationism' and 'intelligent design' should be treated in public education.

While the content of these items may continue to fuel debate, another set of issues in the public science knowledge items warrants closer scientific attention. Is a 'true/false' test using dichotomous response choices the best available means of measuring scientific literacy? We do not know how a respondent processes these questions if some of the items do, in fact, raise conflicting potential responses based on the respondent's knowledge of science on the one hand, and his or her religious commitments of beliefs on the other.

Scholars in the field have given some attention to the measurement properties of the science literacy scales in use around the world, and some have employed item analysis and other psychometric methods to devise alternative scales (Pardo and Calvo 2004; Laugksch and Spargo 1996). Bauer, Allum and Miller (2007) suggest that measurement has shifted with changing 'paradigms' within the Public Understanding of Science survey research field. While issues of measurement and item validity have not gone unattended in this fertile field of social science inquiry, the renewed controversy over validity of the items used in the NSF surveys calls for a careful examination of the quality and utility of the full set of public science knowledge items, drawing on recent advances in survey methodology.

To assist in resolution of these issues, NSF awarded special-purpose grants to the University of South Carolina (with Chris Toumey at the USC NanoCenter serving as PI) and to the University of Virginia (PI: Thomas Guterbock, Director of the Center for Survey Research) to conduct two scientific workshops during Fall 2010 to evaluate the public science knowledge questions in the Survey of Public Understanding of Science and Technology. The first workshop, convened by Toumey on October 22-23, focused on the conceptual content of the surveys and discussed the boundaries between scientific knowledge and religious conviction. The workshop suggested a new conceptual framework, dubbed SSCC, for "Science in the Service of Citizens and Consumers." The second workshop, convened by Guterbock on November 12, examined the 'instrumentation' of the survey, that is, the measurement adequacy of the items themselves, taking into account the newly defined SSCC conceptual framework. This is the report of the instrumentation workshop.

Meeting process summary

In consultation with NSF Program Officers and staff, Guterbock brought together a group of survey methodologists and substantive experts for the purpose of developing a set of specifications to identify the measurement qualities that would be desirable in the public science knowledge questions and outlining a protocol for creating additional questions and testing them. They met at NSF's Arlington headquarters on November 12, 2010. The workshop participants represented a wide range of expertise from the disciplines of sociology, communication, psychology, political science, and health policy and included survey researchers and methodologists. Some of the participants have been involved with the Science Engineering Indicators for many years. Others brought a fresh perspective stemming from their research on related topics. A list of participants appears at the end of this report.

The rich and wide-ranging discussions at the Instrumentation Workshop are briefly outlined here; a more detailed discussion of important issues and themes follows in the next sections of this report.

1) After welcoming remarks by Myron Gutmann (Assistant Director for Social and Behavioral Sciences at NSF), Robert Bell (Senior Science Resources Analyst in NSF's Division of Science Resources and Statistics [SRS]) provided background material regarding the Science and Engineering Indicators and discussed the recommendation from the National Science Board to conduct a thorough review of the public science knowledge indicators through these two workshops.

2) Chris Toumey gave an overview of the Conceptualization Workshop he convened in October and presented the proposed "Science in the Service of Citizens and Consumers" framework that was formulated by that group. He also reviewed that workshop's recommendations.

3) Tom Guterbock presented the following as guidelines for the task of the Instrumentation Workshop:

    a) Write specifications for Public Science Knowledge questions and define desirable measurement qualities.
    b) Write a protocol for finding additional questions and determining how to test them.
    c) Provide recommendations for NSF to proceed from here.

    In the sections below, this report summarizes the panel's work in considering specifications and developing ideas for how to develop new questions.

4) We reviewed in some detail the origin of the current items and the tests that have been conducted to date on their psychometric properties.

5) We attempted to match current public science knowledge items with the requirements of the framework developed by the conceptualization group. This process identified key gaps in the measurement of areas defined in the matrix, primarily in the areas of knowledge of science institutions and the purposes for which people use science knowledge.

6) The group discussed theoretical and practical issues involved in validating items of this kind.

7) Nick Allum gave a presentation about issues he has researched concerning the meaning of "don't know" responses in the public science knowledge items and whether "don't know" should be offered in the questions.

8) Sally Stares presented new results from her Item Response Theory analyses of public science knowledge items and contrasted the results for the UK with those of the US.

9) The group engaged in a wide-ranging discussion of the measurement of belief versus knowledge as it pertains to the existing questions on evolution and the "big bang" theory. We discussed the previous split-ballot experiments with these items, as well as other question formats that might be explored.

10) We learned about recent advances in telephone sampling including dual frames that include cell phones and discussed the potential value and difficulties of collecting public science knowledge items by different modes. It was noted that the new knowledge items introduced in the 2008 General Social Survey and reported in *Science and Engineering Indicators* 2010 depend on visual aids, making them impossible to administer by telephone.

The workshop participants closed the day by working up an initial draft of the group's recommendations, which are presented in finished form near the end of this report. Further discussion of the issues, and of an earlier draft of this report, was carried out via e-mail exchanges in December 2010 and January 2011, using an electronic collaboration tool provided to the group by the University of Virginia.

Validity of the current public science knowledge items

A firm foundation

It must be said at the outset that the current questionnaire items that form the basis for the 2010 *Science Indicators* report rest on a solid social science foundation. These indicators of "scientific literacy" have proven durable and serviceable in practice, and have been widely applied both nationally and internationally. It is clear that they pick up important, theoretically predictable variations in science knowledge.

The science literacy scales incorporate items originally devised by Jon Miller, a giant in the field of public understanding of science, and a skilled and highly experienced survey researcher. Over the years, NSF's SRS Division has sponsored a series of studies and workshops to test, augment, and improve the public science knowledge item series; Robert Bell of SRS described these activities for the workshop. Several of these are of particular importance to the present inquiry.

When NSF collected data on public science knowledge by telephone in 2004, the survey included a split-ballot experiment devised by Roger Tourangeau in the wording of the two true/false knowledge questions that were omitted from SEI 2010. The experiment found greater percentages of correct answers to the items on evolution and the big-bang items where these were prefaced with the phrase, "According to the theory of evolution" or "according to astronomers." These results were reported in the 2006 and 2008 editions of *Science Indicators.*

In an effort to reduce survey time and cost, SRS commissioned a detailed study of the science knowledge items, carried out under contract by RTI International, a leading survey research firm (Bann and Schwerin, 2004). The RTI study showed that the knowledge items generally loaded well into a unified factor of knowledge, with several distinct subscales also discernible. Using statistical techniques drawn from both Classical Test theory and Item Response Theory, the study identified subsets of the items ("short forms") that could account for nearly all the variance in science knowledge, allowing the questionnaire to be shortened for the GSS without any significant loss of measurement accuracy.

For the 2006 survey, NSF moved the data collection to the General Social Survey [GSS], thus gaining access to an area probability sample of the U.S. population, interviewed in person. Prior to this change, and then again when new items derived from science tests like PISA were introduced for the 2008 GSS, items were carefully pre-tested before being fielded.

While the workshop participants agreed that further study of some of these items is warranted, as will be discussed below, we want to affirm that considerable scientific attention has already been paid to the assessment of the measurement properties of the existing science knowledge items used for *Science Indicators*, and that this existing instrumentation assessment work is of high quality and considerable value to our evaluation task.

Direct validation is possible

It is often the case that survey researchers must measure concepts for which no directly observable reference exists. For this reason, many of the most widely used tools for assessing measurement quality focus more on reliability (the degree to which a measure is internally consistent and replicable across instances of measurement) rather than directly on validity itself (the degree to which the measure corresponds to the underlying thing we want to measure). We know from the prior studies mentioned above that the science knowledge items have high inter-correlations and high scores on the Chronbach's alpha test of reliability. But a skeptic might still wonder: How do we know that these tightly correlated sets of items actually measure what they are intended to measure, i.e., a person's understanding of science?

A true test of a survey measure's validity requires the existence of a direct measure of the underlying concept, independent of the survey items that are under test. This is sometimes referred to as a 'gold standard' measure. Fortunately, the concept of "understanding of science" is, in principal, something that can be measured using accepted and well-developed tests. PISA, TIMSS and NAEP[3] are examples of widely accepted, broadly based tests of science knowledge, designed for use with broad populations of students. Granted, these tests were designed to test the knowledge of students who have recently studied science topics in school, but it is plausible that they might serve as one standard against which survey items of adult science knowledge could reasonably be tested

[3] Selected past questions from NAEP are accessible through the National Center for Education Statistics online. See: http://nces.ed.gov/nationsreportcard/itmrlsx/default.aspx.

These are lengthy tests that impose a considerable burden on the student who takes them, and they are clearly not suitable for use in a large-scale survey. However, they could be used in a specialized field study designed to validate directly the science knowledge items used by NSF. One would need only to recruit a reasonable number of adult respondents, randomly chosen from the general population. They would be compensated for their time and would, with informed consent secured, take one or more of the standard science tests, and at the same time respond to the survey items being tested. The analyst would then examine the degree to which scores on the survey items are correlated with scores on the standardized science test. We would expect such a test to show that the NSF public science knowledge items do indeed reflect a person's overall degree of scientific knowledge.

Age and generation would need to be taken account in such an assessment. Generation is at least as important as age, due to massive changes in science education over the past 30 years, reflected in emphases on science inquiry and knowledge of science processes. However, because cohort can easily be coded into the longitudinal NSF Surveys file (as Susan Losh [in press] has already done with the data through 2006 and demonstrated in several publications), it is now more feasible to compare student responses with those of adults. This work could serve as useful background to a validation study of the kind we are suggesting here. The extra value in such a test of the items would be its potential for identifying any specific items which do not correlate closely with a person's overall score on the more comprehensive knowledge test.

What else do we need to ask? Coverage of the current items

Is the full scope of the SSCC framework covered?

As has already been noted, the conceptualization workshop that convened under Chris Toumey's leadership on October 22-23 proposed a new conceptual framework for measuring science knowledge and attitudes, Science in the Service of Citizens and Consumers. The participants of the conceptualization workshop acknowledged the longitudinal and comparative value of the information that has been collected for over two decades but stressed the importance of adopting a modified framework that incorporates recent thought about relations between the science and the public. One insight that was especially salient to this group is that persons in the public have different reasons for acquiring scientific knowledge. Sometimes a person is in the role of acquiring information as a consumer. At other times a person is acting in a civic role and needs scientific knowledge for decision-making about public matters. Finally, acquiring scientific knowledge is often pursued for sheer interest, enjoyment, or personal fulfillment. All three purposes are encompassed within the SSCC framework.

A second consensus of the conceptualization workshop was that there are three principal categories of scientific knowledge that can serve persons pursuing these three identified purposes. First, factual scientific knowledge gives individuals a vocabulary of scientific information and scientific conclusions about the empirical world. Secondly, knowledge of scientific processes and standards enable us to comprehend certain intellectual practices and to recognize good science. Finally, institutional scientific knowledge enables us to know how scientific institutions operate; such knowledge is another key to evaluations of competing scientific (and non-scientific) claims by members of the lay public. (Toumey, 2010). This discussion led the participants of the conceptualization workshop to develop a three-by-three matrix to illustrate the intersection of these purposes and levels of knowledge (reproduced in Table 1.) This matrix served as a guiding tool for the instrumentation workshop in our evaluation of the current science knowledge questions.

Table 1. The SSCC Matrix

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="3">Purposes of public knowledge of science</td></tr>
<tr><td>Civic engagement with science</td><td>Practical/Individual Decision-Making</td><td>Cultural curiosity about the scientific worldview</td></tr>
<tr><td rowspan="3">Content</td><td>Factual Knowledge</td><td></td><td>How should antibiotic meds be used?</td><td>What is an electron?</td></tr>
<tr><td>Processes & Standards</td><td>How is probability relevant to a particular issue?</td><td></td><td>Principle of naturalistic explanation</td></tr>
<tr><td>Institutional Knowledge</td><td>Why does nanotechnology receive gov't funding?</td><td>Which experts and institutions can I trust?</td><td></td></tr>
</table>

Table 2 below shows how the workshop participants mapped the current public science knowledge items onto the three content categories that constitute the rows in the matrix above. As we reviewed the items, we found that many could serve more than one of the purposes described in the three columns of the SSCC matrix, so we did not attempt to classify the items as to purpose.

In 2006, NSF asked SRI to develop and pre-test a set of questions that would serve as indicators of knowledge of science and technology among American adults, for administration in the 2008 General Social Survey (GSS). These items reflected a mix of science field and topic (life science, physical science, earth and space science, science-related math, and science inquiry). Sixteen of the twenty-two items were selected for the GSS. Again, using the content side (rows) of the matrix, these items were classified by the instrumentation workshop participants as shown in Table 3.

Table 2. Review of Basic Science Knowledge Questions

Item No.	Question	Identified Matrix Category
E1.	When you read or hear the term scientific study, do you know what it means?	Process
E2.	What does it mean to study something scientifically?	Process
E3.	Scenario on the best way to test a drug	Process
E4.	Why is it better to test the drug this way?	Process
E5.	Scenario on genetic makeup and inherited illness	Process
E6.	Will each of a couple's children have the same risk of illness?	Process
E7.	The center of the Earth is very hot.	Facts
E8.	All radioactivity is man-made.	Facts
E9.	Whose gene decides whether the baby is a boy or girl.	Facts
E10.	Lasers work by focusing sound waves.	Facts
E11.	Electrons are smaller than atoms.	Facts
E12.	Antibiotics kill viruses as well as bacteria.	Facts
E13.	The universe began with a huge explosion.	Facts
E14.	The continents have been moving for millions of years.	Facts
E15.	Human beings developed from earlier species of animals.	Facts
E16.	Does the Earth go around the Sun, or the reverse?	Facts
E17.	How long does it take for the Earth to go around the Sun?	Facts

Table 3. Review of Items from Revised Science and Technology Module

Item No.	Question	Identified Matrix Category
Q1.	What property of water is most important for living organisms?	Facts
Q2.	Which combination of bodily features is BEST suited to a small animal that needs to minimize heat loss?	Process (graphical literacy)
Q3.	Which of the following is a key factor that enables an airplane to lift?	Facts
Q4.	Lightning and thunder happen at the same time, but why do you see the lightning first?	Facts
Q5.	Results of combining acid and base solutions.	Facts
Q6.	Does temperature affects the behavior of goldfish?	Process (graphical literacy)
Q7.	Results of using ocean water to water vegetables.	Facts
Q8.	Which of the following is NOT an example of erosion?	Facts
Q9.	Traits are transferred from one generation to another.	Facts
Q10.	How do most fish get the oxygen they need to survive?	Facts
Q11.	Why do people experience shortness of breath at the top of a mountain?	Facts
Q12.	Looking at a graph, in which time periods did the most errors occur?	Process (graphical literacy)
Q13.	Which is the BEST method to report the weight of the leaf?	Process
Q14.	Which one of the pictures show what should be used for a pot of plants?	Process
Q15.	Looking at a picture of an experiment, what is the scientist trying to find out?	Process
Q16.	Why did you choose the answer in Q15?	Facts

This exercise of evaluating the current knowledge questions according to the newly proposed conceptual framework suggests that both content (facts) and process items are being covered well by the items in current use by NSF. However, it was noted that no relevant questions exist in the institutional knowledge category. Examples of this type of knowledge might be items asking about the Federal government's role in funding basic research, the role of universities, differences in credibility of independently funded research versus that funded by for-profits,[4] and so on.

The items currently being used to assess the characteristics Americans associate with scientific practices may have some potential as a source for developing items reflecting "institutional knowledge." These items ask about a series of features of a study, and ask the respondent to rate each on how important it would be in making the study scientific. (For

[4] SRS actually included a question on this topic in the 2006 GSS, but results were not reported in *Science Indicators*.

example: "It is done by scientists employed in a university setting." "The people who do it have advanced degrees in their field.")[5] While these items do not measure institutional knowledge of science directly in their current form, the list of features used in this set of items is suggestive of the type of understanding of scientific institutions that a citizen would need in order to make informed assessments of competing claims around science issues. One excellent suggestion to come out of the workshop is that the list of items characterizing science ought to include one or more items that refer to the protection of human subjects in research or to other salient ethical obligations of scientists.

Measuring the purposes for which people use scientific knowledge

The SSCC framework is built on the idea of science knowledge as something that potentially empowers a citizen. Three different purposes of science knowledge are identified, each of which can help to better people's lives and the well-being of society as a whole. People's lives can be improved if they learn to make their consumer choices with scientific findings in mind. With adequate science knowledge, the citizen can take a role as an active participant in deliberation on policy issues where science is involved. And a lay person's pursuit of science knowledge may aid him or her in understanding the world and can even serve as a means of self-expression.

While these purposes are important conceptually, most of the existing public science knowledge items do not map clearly onto any one of these purposes. However, it would be relatively easy to develop new items that would directly measure whether or not a person deploys scientific knowledge for these purposes in everyday life. For example, items could probe whether the respondent regularly consults scientific or technical sources or published data in making important consumer purchases. Other items could measure whether the respondent holds opinions on policy issues where science is relevant, and whether she or he relies on science knowledge in forming such opinions. A third set of items could focus simply on the extent to which a person enjoys hearing or reading about scientific studies, or learning about science. Three distinct "science-purpose" scales could be constructed from these items.

Once these scales are developed, one could analyze the degree to which specific science-knowledge items are correlated with each of the purpose scales. Such an analysis would put to a direct test the plausible but untested assumptions that underlie the SSSC framework: i.e., that science knowledge actually *does* empower citizens or improve their lives by making them more able consumers, more effective citizens, or better able to comprehend the world. This analysis would also make it possible to map some of the current knowledge items onto one or more of the purposes (columns in the SSCC matrix). This mapping exercise would then allow identification of gaps in the public science knowledge item set. For example, this analysis might show that more items are needed that measure science knowledge relevant to consumer purchase decisions.

[5] The respondent is asked: "Now I'm going to read you some statements about science and scientists. How important are each of the following in making something scientific? *[Item is read]*. Would you say this is very important, pretty important, or not important at all in making something scientific?"

How to ask and how to improve the questions

Dealing with non-opinions, and "don't knows"

A significant amount of educational testing literature has explored the use of "don't know" as a response alternative in knowledge items, while survey methodologists have explored the use of "don't know" in opinion items (see Krosnick, 2002) and in political knowledge items. Mondak and others argue that encouraging "don't know" invites a guessing response set. Because some individuals are more likely than others to attempt an answer on the basis of limited or partial information[6], apparent differences in knowledge levels between groups may, in fact, reflect nothing more than a differential "propensity to guess."

Allum and his colleagues have conducted experimental studies to explore this further. They question if respondents are taking blind guesses on these items, does forcing all respondents to guess really improve on the conventional item format? (Allum et al., 2007). They conclude that the answer to this question depends on whether the "don't know" conceals any underlying knowledge. They further find little reason to believe that don't know responses conceal anything more than a trivial amount of underlying information. As with other research (Delli Carpini and Keeter, 1996), they further encourage opinion researchers and political scientists to continue using "don't know."

This workshop did not achieve full consensus on the use of "don't know" responses; the relevant literature is not in consensus either on whether "don't know" should be explicitly offered to respondents, or not. Since Allum's work was concerned with political knowledge items, it would seem worthwhile to undertake similar experiments on the science knowledge items. However, the issue must be considered not only in terms of pure measurement accuracy, but also in terms of usability. Since a set of knowledge items is experienced by respondents as a kind of test, it can be awkward in an interview NOT to allow a "don't know" response when a difficult item is asked. In the absence of conclusive research on the particular items being used to measure science knowledge, it might be best to continue the current treatment of "don't know" in NSF's public science knowledge items, for the sake of continued comparability across years.

In general, guessing becomes more difficult (that is, less likely to hit upon the correct answer by chance) when a larger number of alternatives is offered to the respondent. This suggests that True/False questions are particularly vulnerable to the effects of guessing. While this format might ideally be avoided, the need for comparability with prior years of the NSF surveys suggests that it would be unwise to abandon the True/False format altogether in future years.

The consensus of the workshop participants is that the use of "don't know" as a response option in the public science knowledge questions represents a ripe area for experimentation in the near future. This could take the form of split-ballot experiments on the next round of the GSS, or—a less costly and risky alternative—in separate surveys by various investigators working in the PUS research area. Another promising experimental approach is to first ask

[6] On an educational test, nearly everyone wants to achieve the highest score possible and it is probable that everyone guesses as much as they can. On an opinion survey, however, items testing knowledge may be approached differently by different groups, resulting in different propensities to guess.

the question with an offered DK response, and then invite each respondent who chooses DK to guess at their answer. Yet another possibility is to experiment (with appropriate human subjects protections in place) using questions in which none of the offered alternatives is correct, to more clearly identify the propensity to guess (i.e., an insistence on guessing vs. a willingness to say DK). To be clear, there is substantial research in the existing literature on non-opinions, the varying meaning of "don't know" responses, and the effect on survey results of explicitly offering the DK response or not; what is lacking is research that test these effects on public science knowledge questions in general, or on the NSF items in particular.

It was noted in the workshop that there is evidence of acquiescence bias in the response patterns on true-false items. That is, some respondents tend to agree with statements made by the interviewer, especially when unsure of their own opinion on the matter. As a result, there is a measurable tendency to say that the true-false items are 'true,' resulting in higher percents correct for those items where 'true' is the correct response and lower percents correct for the 'false' items. However, there was no consensus in the workshop that the true-false format should therefore be abandoned.

IRT is a useful approach to appraising survey items

Bryce Reeve raised the additional option of using Item Response Theory (IRT -- also known as latent trait) models to adjust for guessing. IRT models the relationship between a person's response to each item in a scale and the underlying or "latent" variable being measured by the questionnaire. People's responses are taken to be explained by their positions on the underlying knowledge continuum. IRT models responses to each item in terms of a set of properties that describe their performance in the scale (Presser, 2004). There are at least two different IRT models that can incorporate the different response sets depending on whether "don't know" or other options are used. The 3-parameter logistic IRT model adjusts for guessing when a respondent of low knowledge level gets a difficult question correct. The Nominal IRT model can actually model the "don't know" response and align it along with the other response options on the same knowledge continuum.

Nick Allum showed with an IRT model that the evolution item is more "difficult" in the US than in the UK (that is, one needs to have a higher level on the knowledge scale in order to answer it correctly). It is also a relatively weak discriminator for overall science literacy in the UK, but a much stronger one in the US. However, some further IRT analyses by Sally Stares suggest that this strong discrimination may be part and parcel of its generally different measurement behavior when compared with the other items in the set. A two-trait model of US items from 2005/06 showed notable differences in the measurement behavior of the evolution and "big bang" items compared with other knowledge items. In a similar model of Eurobarometer data from 2005, the evolution item did not seem to behave very differently compared with other items for UK respondents. Although the evolution measure is interesting in its own right, this analysis suggests that it is unsuitable for direct cross-national comparison as an indicator of overall scientific literacy.[7]

[7] IRT models of Eurobarometer data from 2002 and 2005 at the European level suggest that the items posed on general science have better scaling properties than those focused on knowledge about biotechnology in particular. Amongst the Eurobarometer biotechnology items, some were designed specifically to capture elements of *affective* responses to or "images" of biotechnology as well as levels of knowledge. The

Some items might be improved through cognitive testing

Most of the item-testing work on the knowledge items that SRS has sponsored has been quantitative in nature. However, some useful insights might be gained if some of the items were tested through cognitive interviews. In cognitive interviews, respondents are typically asked to "think aloud" as they answer a question, or (alternatively) to recount their internal thought processes after they finish a section of the interview. Results are analyzed qualitatively to achieve insight into problems in question wording and to suggest potential fixes. More quantitative approaches to testing question usability could also be tried, such as "behavioral coding" in which observers record any unusual significant problems for interviewers or respondents during live interviews.

An example of an item that might be improved through such testing is the item that asks: "Does the Earth go around the Sun, or does the Sun go around the Earth?" In 2008, only 58 percent of men and 42 percent of women are able to correctly answer both this question and the follow-up for those who give the correct response: "How long does it take the Earth go around the Sun?" It might be worthwhile to see—by means of a split-ballot experiment—if the answers are improved if the order of the two alternatives in the first question are reversed; something about the order of the suggested answers may be leading some respondents into error.[8] If many respondents ask for the question to be repeated, that might signify issues with mental processing of the question elements, simple though they may appear when viewed in written form. It might also be worthwhile to ask the test respondents some general follow-up questions about the solar system (e.g., "Which of these three planets is closest to the Sun?") and then to come back to this question to see if they can give a correct answer after their thoughts about the solar system as a whole have been activated.

Knowledge vs. belief

Although the instrumentation workshop had the task of reviewing the public science knowledge questions as a whole, the workshop participants did devote some attention to the two items that had drawn criticism from the NSB and were eventually deleted from the list of knowledge indicators in the 2010 *Science Indicators* report: the true/false questions regarding evolution of humans from lower forms of life and the origins of the universe in the "big bang." While these questions are stated as simple factual propositions without any direct religious content, it is clear that some respondents respond to the items based on doctrines of the religious belief systems to which they are committed. In particular, conservative Christians who hold the Bible to be inerrant would be reluctant to endorse these items as being "true." The continued strength of conservative Christianity in the United States would then explain the lower scores on these items in the United States compared with other developed countries.

---

measurement behavior of these items was found to be notably different compared to the other items in the set—rather like the contrasts between evolution and "big bang" items and the other general science items, in the US context.

[8] On the other hand, if random error were intruding greatly in the processing of this item, it would not correlate as well as it does with the other public science knowledge items. We thank Bob Bell for making this point.

It is fortunate that the NSF science knowledge questions have been incorporated into the General Social Survey, because that survey includes a very rich battery of questions about religion and religiosity.[9] To our knowledge, the relationship of various aspects of religiosity with answers to the evolution and big-bang items has been explored informally by several researchers, but has not yet been fully treated in published research. These informal explorations find, as expected, that fundamentalist respondents are less likely to endorse these items. In earlier surveys that did not incorporate direct questions about religion, lower scores on these items are found in southern states, where fundamentalist and conservative Christian religious traditions are strongest. As Tourangeau's split ballot experiment showed, the items achieve far higher levels of endorsement if they are preceded with the phrase "According to… ." And Sally Stares' item analysis comparing UK and US response patterns (presented at the workshop) shows that, among US respondents, these items are located in a different area of a two-dimensional factor space than are the other general knowledge items.

An interesting point raised in the workshop by David Sikkink, a specialist in the sociology of religion, is that respondents are not necessarily passive subjects in an interview situation. A highly committed fundamentalist Christian may devote considerable energy in daily life separating self, family, and community from certain elements of the larger society that conflict with his or her world view. This orientation of separation or social distancing may carry over into the interview situation. Hearing a series of questions asking for direct endorsement of the truth of what scientists say, such respondents may be eager to assert their religious difference from the secularized mainstream. For these respondents, an answer of 'false' to the evolution and big-bang items is an assertion to the interviewer of their choice to live out a 'devoted' life.

On the other hand, the issue of knowledge vs. belief need not be specific to any one belief system or domain of science. One can easily imagine a variety of beliefs that people might hold that would conflict with specific elements of accepted scientific teaching in some domains.

With all this said, it is true that the evolution and big-bang items do still correlate fairly well with the other general-knowledge items. Thus, they do function in part as measures of science knowledge, but they are also clearly picking up another dimension (i.e., personal commitment to Christian—or perhaps other—religious orthodoxy). In a word, these are not really bad indicators of science knowledge, but they are not the best.

Since the other general science knowledge items that have been in use for years do correlate highly with one another, one might argue that is not necessary for good measurement of science knowledge to ask respondents specifically about either human evolution or the big bang. However, it is the strong consensus of the workshop participants that the notion of evolution is too fundamental and broad-reaching a concept in science to be left out of the

---

[9] For example, every respondent in the GSS is asked for their specific denominational affiliation, and the many Christian denominations are already conveniently coded with respect to their degree of fundamentalism. Each respondent is asked if they hold a belief in a personal God, about the inerrancy of the Bible, about how often they pray and attend religious services, and other details of their religious beliefs, religious practice, and overall religiosity.

set of public science knowledge indicators. As has already been suggested in the report of the Toumey conceptualization workshop, it would also be possible to test knowledge of evolution by asking about other aspects, such as plant evolution, survival of the fittest, or other elements of evolutionary process that are less directly tied to the hot-button issue of human origins. It might also be of interest to ask whether respondents perceive there to be a consensus among scientists on evolution, the origins of the universe, or on other items of current controversy such as climate change, in a manner similar to questions Scott Keeter and his associates at the Pew Research Center, in conjunction with AAAS, asked in their 2009 national survey (Pew 2009).

If NSF were to keep these items (for the sake of continuity and international comparison) we would not recommend their inclusion in their original form. Continued use of the split-ballot form of the items, in which at least half of respondents are asked whether the item is true or false "according to evolutionary theory" or "according to astronomers," would be preferable. An even better approach was suggested in the workshop and warmly endorsed by the participants: one could construct a contingent or "unfolding" form of the question, in which respondents would first be asked if the item were true "according to…," and then asked directly if they personally share this belief. That approach would have the virtue of clearly separating a respondent's knowledge about what scientists believe from his or her personal beliefs, which would thus also be measured in the survey for every respondent.[10]

We understand that some would argue that a person who disbelieves what science teaches cannot be said to be truly knowledgeable about science. This "either/or" assertion would echo the age-old conflict between religion and science. It is likely, however, that the role of religious doctrine on the one hand and science-based knowledge on the other in shaping personal belief systems is, for many adults, more complex and situational than would be suggested by this "either/or" view of the conflict. It is quite possible for people to hold contradictory beliefs or attitudes that they will apply as needed to different situations. It would certainly be a mistake to equate religious commitment, automatically, with a lack of scientific knowledge, even if survey items from the two domains may be negatively correlated.

The NSF public science knowledge items should be clearly focused on science knowledge. A change in measurement of the concept will not, in itself, resolve the debate over whether or not personal belief in human evolution is an essential part of scientific knowledge, or—more generally—whether science literacy necessarily assumes acceptance of the scientific consensus, as opposed to mere knowledge of that consensus. Our suggestion of an unfolding version of the question will, at least, provide researchers with information that more clearly separates a respondent's knowledge of scientists' views from his or her personal agreement with those views, thus allowing for a more informed investigation of both aspects of public science knowledge.

[10] In the split-ballot experiment, we have an imperfect measure of belief for those who are asked the unmodified version of the question, while for those asked the "according to…" version we have a measure of knowledge or familiarity, separate from belief. With the contingent form suggested here, we would measure each dimension separately for each respondent. Of course, the new item format would need careful development and pretesting.

The workshop participants are in consensus that the topic of the relationship between religion and science deserves additional research. This could be addressed through the periodic inclusion in the NSF science survey series of a topical module that delves into additional facets of the science/religion discussion. Other topics on the public agenda—nuclear energy, climate change, stem cell research, biotechnology, nanotechnology, etc.—similarly deserve to be periodically addressed in greater depth. Such modules should consider both knowledge and attitudes. Inclusion of such modules might also contribute to additional secondary analysis of the data by giving scholars focused on specific substantive topics a reason to use the *Science Indicators* data. The research necessary to develop suitable measures for these areas could also contribute to enhancing the validity of scholarship on public attitudes toward and understanding of science.

How we reach respondents and gather their answers

For several decades, starting after 1979[11], the NSF surveys of public understanding of science were conducted by telephone, first by the Northern Illinois University Public Opinion Laboratory and then by the University of Michigan Survey Research Center.[12] As noted in the RTI report, response rates for this survey were typically at 70 percent or higher. However, for 1999 and 2001, response rates declined sharply to 66 percent and 39 percent, respectively. Additionally, NSF found an overrepresentation of the highly educated and an underrepresentation of those with little education (Bann and Schwerin, p. 1). These considerations, along with the degradation of the landline telephone sampling frame, led NSF to move the vehicle for the biennial surveys to NORC's General Social Survey, an NSF-funded biennial survey of the national adult population that uses (primarily) face-to-face interviews with a sample drawn through area-probability sampling, the gold standard for survey samples. This key decision has put the NSF survey on a very solid footing for full representativeness of the sample and high quality in data collection. It also carries the benefit that the GSS includes a very detailed set of demographic background questions and (as noted) quite detailed questions on religion and religiosity, allowing researchers to correlate science knowledge items with other variables that were not available on the prior telephone surveys. Because the GSS is conducted face-to-face, it has also been possible to design science process questions that include pictures, hand cards, and other visual cues that could not be used in a phone interview; these are important in the "science process" questions that were introduced in the 2008 GSS science knowledge module.

However, the use of the GSS as the NSF's primary survey vehicle for measuring public science knowledge has several costs and disadvantages that must be acknowledged. First, the GSS is a very expensive survey when evaluated on a per-question basis. Second, it is fielded only once every two years. The lead time for getting new items onto the GSS is long and requires several levels of rigorous review. Third, as an omnibus survey with many continuing items and new topical modules each year, the GSS can devote only a small portion of interview time to the science knowledge topic. The short-form items developed

[11] The original 1979 surveys were in-person, conducted by the Institute for Survey Research at Temple University.
[12] The 2001 telephone survey was fielded by ORC Macro.

in the RTI report work reasonably well, but it is very difficult to make room on the GSS for additional items or expanded experiments with variations on the current items.

It is also the case that, in recent years, innovations in telephone sampling have been achieved that might greatly alleviate NSF's concerns about the representativeness of telephone samples. In particular, the introduction of dual-frame telephone surveys that combine random-digit dialing of landline telephones with randomly dialed cell phone numbers has greatly enhanced the ability of telephone samples to reach a representative cross-section of the population (see AAPOR Task Force, 2010.) By directly calling persons with cell phones, researchers are able to fill the main gap in coverage that now exists in the landline RDD frame (the absence from that frame of the "cell-phone-only" respondents). Moreover, the introduction of the cell phone component in these samples helps researchers to overcome the limitations of "at-homeness" that have caused consistent and familiar biases in household surveys by phone (for example, the over-representation of women and married persons.) Using cell phones in combination with traditional telephone sampling methods, researchers can now reach, much more efficiently, precisely those groups that were previously under-represented: those who are young, poor, unmarried, transient, male, urban, or members of minority groups. Reaching all these groups is essential if the level of public understanding of science is to be accurately gauged.

The potential advantages of telephone surveys need to be considered by NSF as it considers expanding research on public understanding of science. It is true that response rates to telephone surveys are lower than they once were, and rates are lower than those achieved in door-to-door surveys. But the lower response rates do not appear to have diminished the accuracy of telephone surveys. Compared to face-to-face surveys using area-probability samples, telephone surveys can be completed at a fraction of the cost and in much shorter turn-around times. With the help of computer-assisted interviewing systems now standard in the industry, complex experiments are easily fielded by telephone (although the same is true when a face-to-face interview is controlled by a lap-top computer.) Since the original NSF surveys were conducted by phone, the original public science knowledge questions were written for oral administration without visual aids. And there is thus no threat of "mode effects" when we compare the results of a new telephone survey to results achieved by telephone in the NSF surveys prior to 2006.

We are ***not*** suggesting that the biennial survey data collected for *Science Indicators* should be moved from their current home at the GSS and shifted back to a telephone survey. However, we do not see how the needed testing, experimentation, and item development tasks can be accomplished in timely fashion if the science knowledge items cannot be deployed in telephone surveys.

If items require visual presentation, as some of the current GSS items do, they could be fielded at fairly low cost via the Internet, using a probability-based web panel such as that offered by Knowledge Networks. However, any such effort will run up against issues of representativeness in web panels (Baker 2010), which are known to under-represent those with lesser education, economic resources, and access to technology. Administration by mail, using the newly developed Address-Based Sampling [ABS] methods, might also be considered, but here again literacy bias will be significant. And in both the Internet and postal modes, one would not be able to prevent respondents from asking for help from others, looking up answers, or (in a paper format) going back to change their answers on some items. Given the objective of properly gauging scientific knowledge across the population, these biases and threats to accuracy would seriously jeopardize the results. But telephone surveys based on a dual-frame design truly reach across the entire spectrum of the population, can fairly easily accommodate language differences as well, can test the un-prompted knowledge of the respondent, and yet remain reasonably affordable. It is also possible to field surveys in which the main mode of response is oral (by telephone), while the sampling frame is based on ABS, and other modes of response might be offered. Such mixed-mode designs offer the possibility of addressing both coverage and non-response issues that plague the single mode approaches.[13]

If the NSF science items are to be tested by phone, then an immediate research task is to develop equivalents for the GSS "science process" questions that will work without the assistance of visual aids. An Internet experiment that offered some respondents the existing, visually based items and other respondents the newly devised, text-only items could serve as an initial, low cost test of scale equivalence. A telephone-based test of the two approaches could also be devised, if respondents in the "visual" arm of the trial were pre-mailed the visual aids in a "lock-box" format (that is, a packet that is opened by the respondent only at the time of the interview, upon instructions from the interviewer).

Development of an oral version of the science knowledge items would also be important in encouraging and facilitating research on science and the public across the US and internationally. Those interested in particular science issues could easily add some or all of the NSF public science knowledge questions to their tailored, issue-specific telephone questionnaires. When new science issues or science policy legislation enter public discussion, nationwide or local polls could incorporate the public science knowledge questions as a point of reference. There is much research to be done in the area of public attitudes toward and understanding of science, and the instrumentation must once again become telephone-friendly if NSF wants to see that research undertaken by the research community.

[13] Looking to the future, it may become practical to incorporate visual materials into telephone surveys by making use of the ability of 'smart' phones to receive and display images. Recent experiments with surveys using handhelds have only been conducted with limited samples of suitably equipped volunteers, but in the future the technologies could conceivably become widely enough available to allow their use in surveys of the general population, without undue income bias.

Recommendations

1. We endorse all of the recommendations of the conceptualization workshop (Toumey 2010). In particular, the new SSCC conceptual framework that they recommend would be fairly straightforward to implement and we are confident that proper instrumentation can be developed to realize this framework in practice. We join them in saluting the prior work of Jon Miller and the continued attention to measurement quality that has come from SRS's leadership and the many who have contributed to the development of the current instrumentation.
   1.1. We also wish to emphasize, as did the conceptualization workshop, the need for continuity and international comparability as the measures are modified and augmented.
2. A direct test of the validity of the public science knowledge items should be undertaken, to see if scores on these items are closely correlated with scores on accepted, standardized tests of science knowledge such as those used in PISA, TIMSS, or NAEP.
3. Researchers should explore the efficacy of alternative question formats for the knowledge items (both content and process items).
   3.1. Research should consider the effect of offering or not offering a "don't know" response to respondents. This is a ripe area for experimentation in preparation for the 2012 GSS. Effects on accuracy and usability should both be considered.
   3.2. Research should consider whether the questions about evolution and the "big bang" might be more effectively asked in a contingent format that separates knowledge of what scientists say from what the respondent personally believes.
   3.3. Research should consider the effect of increasing the number of offered alternatives and explore its effect on guessing behavior by respondents. Again, effects on both accuracy and usability should be considered, including usability in a telephone interview.
4. The science knowledge items should be tested further for comprehension and other cognitive issues by using cognitive interviewing and behavioral coding techniques. These tests can be performed in a laboratory setting but should be applied to respondents recruited from the general population by probability sampling methods.
5. Implementation of the SSCC framework will require the development of new items that tap the "institutional knowledge" dimension. These items will ask respondents about how science research is organized and carried out, attempting to tap information people would need to know to gauge the credibility of publicly reported scientific 'evidence' and also to understand how scientific research is supported.
   5.1. The existing items give adequate coverage to the 'content' and 'process' categories of the SSCC framework, but new items could also be considered for these categories and some of the existing items, as noted, could be improved.
6. New items should be developed to measure how respondents use science knowledge. Do they typically use scientific or objective data to inform their purchase decisions and consumer choices? Do they refer to scientific or technical sources when they participate as citizens in debates of public issues, especially those that involve science issues? Do

they have a passion for science learning "for its own sake," that is, as a tool for self-expression, personal fulfillment, or as an important source of meaning? Without these items, the SSCC framework will not be fully measured.

6.1. Once these "purpose scales" are developed, researchers can learn which science knowledge items tend to be used for each purpose, and by whom.

6.2. If those with greater science knowledge also score higher on the three "purpose scales," then some key but untested assumptions of the SSCC framework will be empirically validated.

7. The current true/false items on evolution and the "big bang" are flawed. They display measurement properties that set them apart from other public science knowledge items and mar their usefulness for international comparison. They should be modified in form, or replaced with items on related topics (e.g., plant evolution) that are not as directly affected by respondents' religious beliefs.

7.1. Experiments should be conducted that test a contingent form of these items in which knowledge and belief could be separately measured. On the human evolution item, for example, respondents would first be asked if the item is true or false "according to the theory of evolution." They would then be asked a separate, follow-up question asking if they personally hold the statement to be true.

8. Future rounds of the NSF science surveys should include topical modules that allow for expanded exploration of relevant topics. (Simply repeating the core items from year to year will not allow research in this field to progress at the rate that it should.)

8.1. Alternatively, NSF should fund topically focused research that uses the NSF public science knowledge items on survey platforms outside of the GSS.

9. The science knowledge indicators should be the subject of recurrent review (at least every two years), to take into account important changes in science and in science institutions and to ensure continued improvement in measurement quality.

10. Much could be learned from secondary analysis of the existing NSF public science knowledge data. Not enough such analysis is currently being undertaken.

11. We see a need for considerable research effort in order to accomplish the steps recommended above. NSF should increase its funding for research on public attitudes toward and understanding of science.

11.1. Some of the more immediate research steps should be funded directly by SRS using contract mechanisms. Other projects would lend themselves to funding through competitive grant mechanisms in programs housed in other Directorates such as Social and Behavioral Sciences or Education and Human Resources.

11.2. NSF should establish a dedicated research fund that could be used to augment funding for research in public attitudes toward and understanding of science that is favorably reviewed by existing grant programs. The availability of these funds would increase the chances such proposals would have for being funded and thereby generate a larger number of such proposals from the research community.

Conclusion

Our review of the current instrumentation and the proposed Science in the Service of Citizens and Consumers [SSCC] framework is generally supportive of the present measures as well as the proposal to expand them. There is clearly a need to augment the present measures to fully encompass the several dimensions of the newly proposed SSCC framework. It will be essential to develop items to measure 'institutional knowledge' of science. We also think it will be important to measure the extent to which people use science knowledge for the several identified purposes: consumer decision-making, democratic participation, and self-improvement or 'science for its own sake'.

We have looked at the controversial true-false items on evolution and "the big bang" and conclude that these are indeed problematic. We have recommended that they be asked in a different format, or revised to ask about other aspects of evolution.

We see the need for experimentation in several areas, including whether or not respondents should be offered "don't know" as a response to the questions, and suggest that the NSF surveys be augmented with periodic topical modules that expand the information gathered on selected topics.

We urge that a version of the NSF items be developed that is suitable for oral administration via telephone. This would involve further shortening of some of the existing scales and elimination of the need for visual material as part of the 'science process' question series. Facilitating use of these measures in telephone surveys will enable NSF to move forward with the speed necessary in order to develop an improved set of questions, fully covering the SSCC conceptual framework, in time for their inclusion in the 2012 GSS and the 2014 *Science Indicators* report.

All this research will require a substantial funding commitment from NSF, appropriate incentives to university researchers to carry out the research in a way that serves NSF's measurement needs, and speedy response both from contractors and from university-affiliated scholars working in the public understanding of science field. We are confident that such ways can be found and that the instrumentation for measuring the public's attitudes toward and understanding of science will be substantially improved.

All of the participants in the instrumentation workshop are grateful to NSF for supporting our efforts and for continuing to move this important area of research forward.

Participants

The November 12, 2010 Instrumentation Workshop meeting was moderated by Thomas M. Guterbock, Director of the Center for Survey Research, Professor of Sociology, and Research Professor of Public Health Sciences at the University of Virginia.

The remaining participants were:

Nick Allum
Department of Sociology
University of Essex

John Besley
School of Journalism and Mass Communications
University of South Carolina

Frederick Conrad
Institute for social Research
University of Michigan

Allyson Holbrook
Survey Research Laboratory
University of Illinois

Scott Keeter
Director of Survey Research
Pew Research Center

Susan Losh
Department of Educational Psychology
Florida State University

Jeff Mondak
Department of Political Science
University of Illinois

Bryce Reeve
Department of Health Policy
University of North Carolina

David Sikkink
Associate Professor in Sociology
University of Notre Dame

Sally Stares
Methodology Institute
London School of Economics

Roger Tourangeau
Survey Research Center
University of Michigan

Chris Toumey
Nano Center
University of South Carolina

Chris Toumey was the convener of the Conceptualization Workshop held in October and John Besley participated in that workshop as well.

In addition to these attendees, Robert Bell and Brett Pelham participated in the Instrumentation Workshop as representatives of the National Science Foundation. Carolyn Funk, who is an Associate Professor at Virginia Commonwealth University, attended the workshop in her role as the author of Chapter 7 of the 2012 Science and Engineering Indicators. Deborah Rexrode, who is a Project Manager at the Center for Survey Research at the University of Virginia assisted in the organization and recording of the workshop.

We express our appreciation to Dr. Myron Gutmann, Assistant director, NSF Directorate for Social, Behavior, and Economic Sciences for his support of these two workshops and his welcome to the group.

References

AAPOR Cell Phone Task Force. 2010. *New Considerations for Survey Researchers When Planning and Conducting RDD Telephone Surveys in the U.S. With Respondents Reached via Cell Phone Numbers.* American Association for Public Opinion Research. Available online at www.aapor.org.

Baker, Reg, Stephen L. Blumberg, J. Michael Brick, Mick P. Couper, Melanie Courtright, J. Michael Dennis, Don Dillman, Martin R. Frankel, Philip Garland, Robert M. Groves, Courtney Kennedy, Jon Krosnick, Paul J. Lavrakas, Sunghee Lee, Michael Link, Linda Piekarski, Kumar Rao, Randall K. Thomas, and Dan Zahs. 2010. "AAPOR Report on Online Panels." *Public Opinion Quarterly* 74(4), Winter: 711-781.

Bann, Carla M. and Michael J. Schwerin. 2004. *Public Knowledge and Attitudes Scale Construction: Analysis of Short Forms.* Prepared for NSF Science Resource Studies. RTI International (May). Contract No. GS-10F-0097L ; RTI Project No. 8796.

Bauer, M.W., N. Allum, and S. Miller. 2007. "What can we learn from 25 years of PUS survey research? Liberating and expanding the agenda." *Public Understanding of Science*, 16:1, pp. 79-95.

Bhattacharjee, Y. 2010. "NSF Board Draws Flak for Dropping Evolution from Indicators." *Science*, 328:154.

Delli Carpini, M. and S. Keeter. 1996. *What Americans Know About Politics and Why It Matters.* MA: Yale University.

Krosnick, Jon A. 2002. "The Causes of No-Opinion Responses to Attitude Items in Surveys: They Are Rarely What They Appear to Be." Pp. 87-100 in Robert M. Groves, Don Dillman, *et al.*, eds., *Survey Non-Response.* New York: Wiley.

Laugksch, R.C. and P. E. Spargo. 1996. "Construction of a paper-and-pencil test of basic scientific literacy based on selected literacy goals recommended by the American Association for the Advancement of Science." *Public Understanding of Science*, 5:4, pp. 331-359.

Losh, Susan C. (In press). "Age, Generational, and Educational Effects on American Adult Public Understanding of Science." Chapter in Martin W. Bauer, Rajesh Shukla and Nick Allum (eds.), *The Culture of Science - How does the Public relate to Science across the Globe?* NY: Routledge.

Pardo, R. and F. Calvo. 2004. "The cognitive dimension of public perceptions of science: Methodological issues." *Public Understanding of Science*, 13:3, pp. 2103-227.

Pew Research Center for the People and the Press. 2009. *Scientific Achievements Less Prominent Than a Decade Ago; Public Praises Science; Scientists Fault Public, Media.* A survey conducted in collaboration with the American Academy for the Advancement of Science. http://people-press.org/reports/pdf/528.pdf.

Presser, Stanley. (2004) *Methods for Testing and Evaluating Survey Questionnaires*. NJ: Wiley and Sons, Inc.

Toumey, Chris, et al. 2010. *Science in the Service of Citizens & Consumers: The NSF Workshop on Public Knowledge of Science, 22-23 October 2010*.