Measuring Interest in Science: The Science Curiosity Scale

Asheley R. Landrum[1] (ALandrum@asc.upenn.edu)

Joseph Hilgard[1] (JHilgard@asc.upenn.edu)

Heather Akin[1] (HAkin@asc.upenn.edu)

Nan Li[1] (NLi@asc.upenn.edu)

Dan M. Kahan[1,2] (Dan.Kahan@yale.edu)

[1]Annenberg Public Policy Center
University of Pennsylvania
202 S. 36th Street, Philadelphia, PA 19104

[2]Yale Law School
Yale University
127 Wall Street, New Haven, CT 06511

Abstract

In the current study, we present the methods for creating and validating a science curiosity scale. We find that the scale presented here is unidimensional and highly reliable. Moreover, it predicts engagement with a science documentary clip more accurately than do measures of science intelligence or education. Although more steps are needed, this provides initial evidence for the utility of our measure of science curiosity.

Keywords: curiosity; science curiosity; scale; psychometrics; Item Response Theory

Introduction

Do people differ in their desire to seek out and consume science information for personal satisfaction? Determining the answer to such a question requires having a genuine measurement of science curiosity. Although many scales purport to measure such a construct, performance assessments of these scales reveal that most (if not all) of such attempts are psychometrically weak and often not genuinely predictive of what they are supposed to be assessing (e.g., Blalock et al., 2008; Osborne, Simon, & Collins, 2003). The aim of the current study was to take the first steps toward developing an original, valid, and reliable science curiosity scale.

One problem with existing scales that purport to measure science interest or science curiosity is heavy reliance on self-reported measures. Although asking people directly is often a good way of gaining information, there are potential problems. For instance, asking participants to what extent they agree with statements such as "I'm curious about the world in which we live," and "I find it boring to hear about new ideas," (Fraser, 1978) is likely to provoke socially desirable responding.

Public opinion research, for example, has used numerous "science attitudes" batteries that purport to measure science interest by literally asking people if they "like" science. The National Science Foundation Indicators (2014), for instance, feature an array of "public attitudes toward science" items. These items consistently find that members of the public hold overwhelmingly pro-science attitudes (e.g., 4 out of 5 Americans say they are interested in new scientific discoveries, National Science Board, 2014). As these items are subject to the same problems related to self-report measures generally, it is unclear whether the American public actually *does* hold overwhelmingly pro-science attitudes or if this positivity is a direct result of socially-desirable responding. Therefore, when using self-report measures, it is important to determine whether the scales are measuring the underlying disposition or trait of interest and not simply the motivation for others to perceive them as possessing it.

Current Study

Our strategy for conducting a valid science curiosity scale was to combine a number of self-report measures with behavioral and performance ones. First, we used the behavioral and performance items to validate the self-report items—that is, to confirm that the variance in the self-report items could be treated as originating in difference in science curiosity rather than because of some other reason. Then, we combined the self-report items with the behavior and performance items to form a scale that would reliably discriminate among study participants with varying levels of science curiosity.

To counteract the problem of socially-desirable responding, we disguised our objectives by presenting our scale as a marketing survey. That is, we embedded individual self-report items relating to science interest in modules consisting of multiple items reflecting an array of interests (e.g., sports, entertainment, business, and politics). Thus, there was no particular reason for participants to suspect that we were specifically interested in capturing their motivation to learn about science, and we could avoid inadvertently encouraging participants to express pro-science sentiments or to engage in the form of overstatement that pervades many self-report scales.

In addition to the self-report interest and behavior items, we included an objective performance item. Near the middle of the survey, participants were told that we wanted to get their reactions to a news story "of interest to them." In order to ensure that participants were presented with a news story that matched their interests, they were provided with a discrete list of news story sets (See Figure 1) and asked to choose the set that would be of most interest to them. One set consisted of science stories, while the others consisted of popular entertainment, sports, and financial news. Given that reading an article and answering questions is more cognitively taxing than simply answering questions about one's self, we perceived that a participant's purposeful selection of a science story over the others as one valid indicator of genuine science interest.

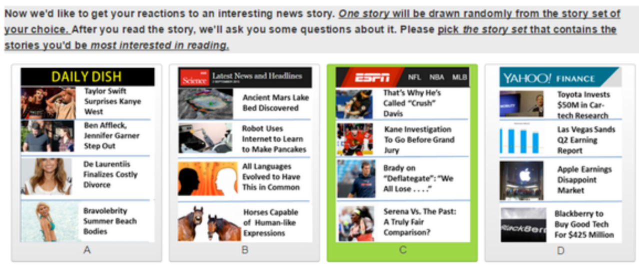

Figure 1. News Story-Set Selection Task. Subjects were instructed to select one set of the four, from which a story would be selected "at random" for them to read and answer two questions on. The task was conceived of as a performance-based measure of interest in science.

With the collected data, we used item response theory (IRT) to combine the items into a composite scale. IRT, unlike simpler alternatives to scale creation and validation, does not assume each item in a questionnaire is equivalent to another item. Rather, it is expected that a pro-curiosity response to some items may indicate higher levels of curiosity than pro-curiosity responses on other items (also known as differences in item "difficulty"). These differences are taken into account by IRT when calculating the scores on the latent variable (Embretson & Reise, 2000).

Moreover, IRT allows one to examine how informative the proposed scale is for each level of the latent variable. While some scales may have high inter-item reliability (e.g., Cronbach's alpha), they may be informative for only a portion of the scores. For example, the cognitive reflection test (i.e., CRT, Frederick, 2005) is very discriminating among people who score above the mean, but provides little to no information among those who score below (Kahan, 2014).

After determining reliability, we were able to behaviorally validate the scale by determining to what extent it predicted engagement with a clip from a science documentary. Indeed, these data are part of a larger study that examines how science curiosity measures might be useful tools for science filmmakers to better target their documentaries to more diverse audiences.

Method

Participants

Participants were 2,267 adults (54.3% female) who are members of a nationally-representative panel of participants from YouGov[1]. Of the participants, 76% reported being white, 9.3% reported being black or African American, 7.8% reported being Hispanic, 2.1% reported being Asian, 0.6% reported being Native American, 2.6% reported being of mixed race, and 1.4% reported being of another race. The participants ranged in age from 18 to 90 years old (Mean = 46.7, Median = 48, SD = 16.23).

Procedure

As previously stated, participants completed the science curiosity items as part of a larger experimental survey examining science curiosity and engagement with science documentary films. Participants completed the survey over the internet, either on computers or mobile devices.

The first part of the survey included 7 modules, in which our science curiosity items (items related to scientific research or discoveries and new technologies) were embedded with items related to other issues, such as crime, education, government or politics, sports, religion, international affairs, business or finance.

The first module was *News Interest*. In this module, participants were asked to rate how closely they follow the news related to each topic: not at all (1), a little, but not closely (2), closely but not very closely (3) or very closely (4).

The second module was *Leisure Activity*. In this module, participants were asked to indicate how many times in the past year they had engaged in several activities including visiting a science or technology museum, attending a live sporting event, visiting an art museum, attending a musical performance or concert, going to a zoo or aquarium, going to a public library, going to a gun show, visiting a theme park or amusement park, attending a political rally or political event, attending a public lecture (on history, science and technology, public affairs or politics, religion, economics, or other).

The third module was *Books*. In this module, participants indicated whether they had read a book in the past year on each of several topics. Topics included crime, science fiction, mystery, education, government or politics, sports, religion (other than Holy Scripture text), international affairs, business or finance, scientific research or discoveries, history.

The fourth module was *Conversation*. In this module, participants indicated what types of topics they discuss with their friends, family members or co-workers. For each topic, participants were asked to say whether they discussed it never, rarely, more than rarely but not often, or often. Topics included crime, education, government or politics, sports, religion, international affairs, business or finance, scientific research or discoveries, new technology, entertainment or celebrities.

The fifth module was *Social Media*. In this module, participants indicated whether (and if so, how often) they share news stories on social media. Participants who said that they did so were asked to rank the topics in order of how likely they were to share. Topics included crime, education, government or politics, sports, religion, international affairs, business or finance, scientific research

[1] For information about YouGov's panel of participants see: https://today.yougov.com/about/about-the-yougov-panel/

or discoveries, new technologies, entertainment or celebrities.

The sixth module was the *Reading Selection Task*. As previously stated, participants were told that we wanted to get their reactions to an interesting news story drawn from a story set of his or her choice. Thus, participants were asked to pick the story set that contained the stories that they would be most interested in reading from entertainment, science, sports, and business or finance (see Figure 1). Following their selection, participants were shown one story from the set and were asked to read it and answer two factual questions about that story.

The seventh module was *Self-Reported Interests*. In this module, participants were told that they would see several topics and for each topic they would be asked to indicate how interested they were in that topic: not at all interested (1), slightly interested (2), more than slightly—but not very—interested (3), or very interested (4). Topics included government or politics, sports, religion, foreign travel, scientific research or discoveries, new technologies, entertainment or celebrities, nature, and music.

Following these modules, participants watched a clip from a science documentary about the evolution of color vision and were asked questions about their interest in the clip, factual questions about the clip, and agreement questions (e.g., whether they thought the documentary supplied convincing evidence of how color vision came about). Moreover, several behavioral variables were collected such as how long participants watched the clip before turning it off. In addition, participants answered a battery of items related to their beliefs about policies and risks, and their cultural worldviews (Kahan, Braman, Slovic, Gastil, & Cohen, 2008) and a questionnaire measuring ordinary science intelligence (e.g., Kahan, in press).

Results

Self-Report Science Interest

First, because we aimed to use the Reading Selection Task and the behavior items as validators of the self-report items (News Interest, Conversation, and Self-Reported Interests modules), we started by forming a scale that aggregated these self-report interest items. The resulting self-report science interest (SRSI) scale displayed a high degree of measurement precision (α = 0.85). More importantly, the scale's properties suggest valid measurement of science curiosity.

First, SRSI was positively correlated with the subjects' science comprehension as measured by the Ordinary Science Intelligence assessment (i.e., Kahan, in press; Kahan et al., 2012), $r = .26$, $p < .001$, and education, $r = .21$, $p < .001$. See Figure 2. Although science curiosity and science comprehension are not the same constructs, one would suppose that people who are proficient in science comprehension would also be more likely to like science and that those who were genuinely interested in science would at least be modestly more proficient in comprehending it.

Second, SSRI predicted variance in the responses to the behavioral measures. For instance, it predicted which subjects would select the science set in the Reading Selection Task, $X^2(1) = 181.17$, $p < .001$, which subjects attended a science lecture in the last year $X^2(1) = 230.86$, $p < .001$, and which subjects had read a book about scientific research and discoveries $X^2(1) = 1197.56$, $p <. 001$.

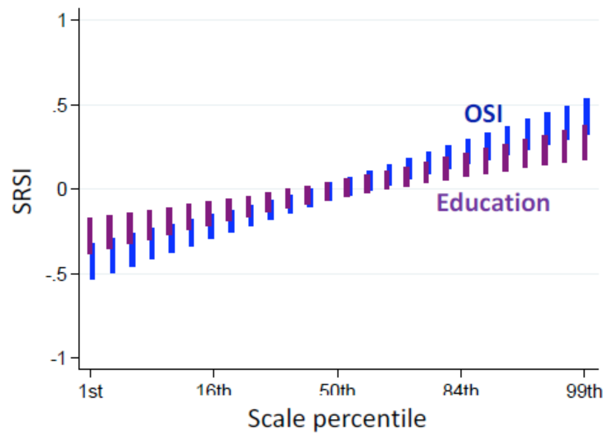

Figure 2. Relationship of Self-Report Science Interest (SRSI) to science comprehension (OSI) and education. Results based on multivariate linear regression (OSI: b=0.21, p<0.01; Education: b=0.14, p<0.01; R^2=0.08). Colored bars are 95% confidence intervals. Scales are normalized with the mean equal to zero and units expressed in standard deviations.

Using Item Response Theory to Create a Science Curiosity Scale

The power of individual items to contribute to measurement precision at different levels of a latent variable can be incorporated into a scale using Item Response Theory modeling (Embretson & Reise, 2000). We used IRT to form a composite scale that combines responses to the self-report interest items (SRSI), the self-report behavioral items (book item and public lecture item), and the Story Selection Task item. The ranking scores for the science news stories in the social media module were also included for participants who indicated that they did indeed share material via social media platforms (social media module).

The resulting scale—the "Science Curiosity Scale" (SCS)—displayed desirable psychometric properties. It was unidimensional (Figure 3, image C), supporting the inference that it measured a single, unitary latent disposition. Moreover, the scale reflected a high level of reliability—at or above α = 0.80— across the entire range of latent science interest disposition (Figure 3, image D).

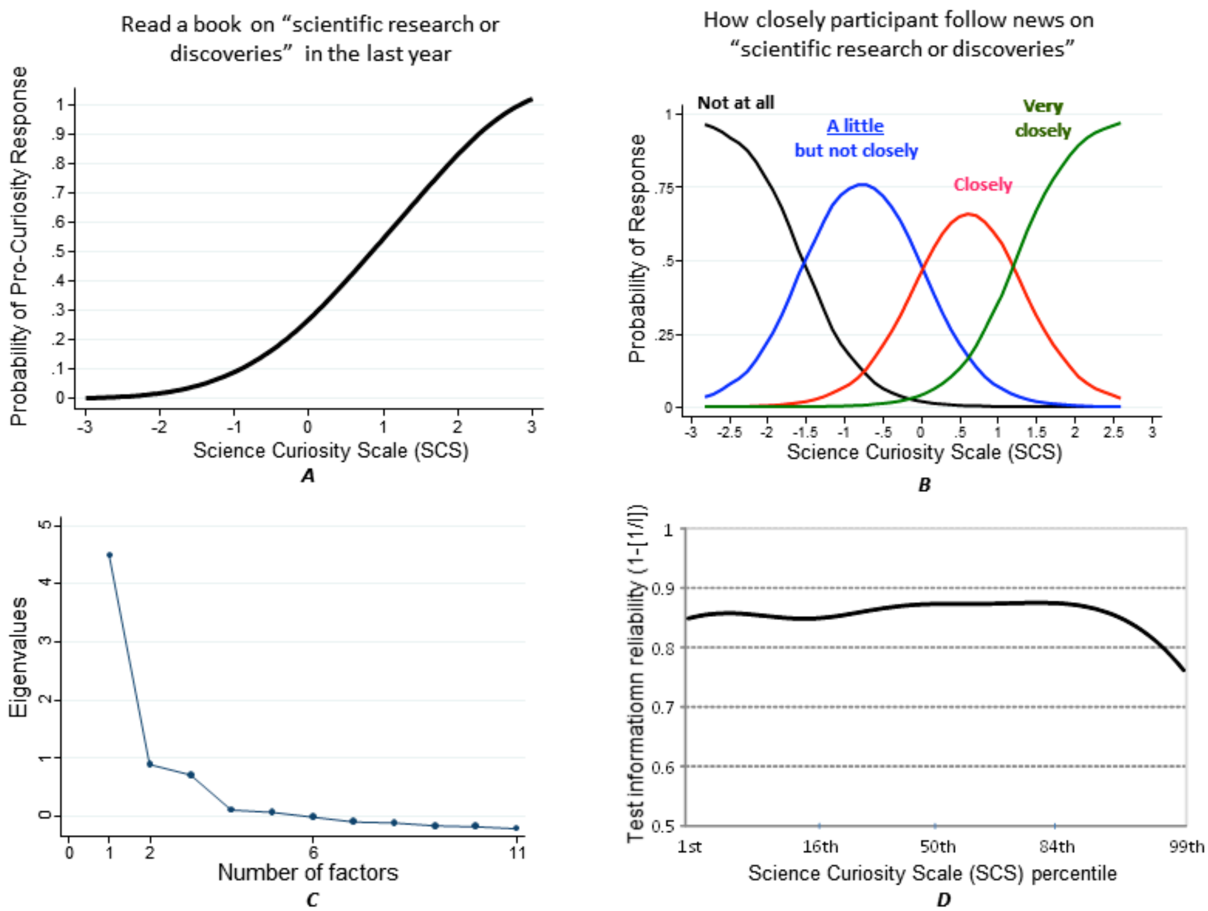

Figure 3. The Science Curiosity Scale (SCS) based on a 2pl Item Response Theory Model. SCS scores are standardized with the mean centered at 0 and units measured in standard deviations. Images A and B reflect representative "item response profiles": the relative probability of the indicated response conditional on a specified level of the latent science curiosity disposition, which is used to estimate subjects' SCS scores. Image C reflects the unidimensionality of the scale. Image D illustrates the measurement precision (test information reliability, similar to Cronbach's alpha) at various levels of science curiosity.

Science Curiosity Predicts Engagement with a Science Documentary Clip

In our study, we aimed to externally validate SCS with engagement with a documentary clip focused on the evolution of color vision. To measure engagement with the video clip we had a combination of self-report measures and behavioral measures that we combined in a manner similar to SCS using IRT. One item asked participants how interesting they found the documentary (M = 3.22, SD = 1.69, Range = 0 to 5). Another item was part of an experimental condition in which we provided half of the sample the ability to turn off the clip whenever they felt they watched enough. This allowed us to measure the number of minutes of the clip watched (out of 10 minutes; M = 6.42 minutes, SD = 4.12). We hypothesized that people who were more science curious would watch the clip for a longer period of time than people who were less science curious. We also offered the sample the option of requesting the full episode of the documentary. If participants were interested in watching the full episode (and selected "Yes", 51%), we would email them a link to the full episode (no payment required). We had hypothesized that participants who were more science curious would be more likely to request the full episode of the documentary. Moreover, we combined these items using IRT to create an index of engagement with the science documentary clip using IRT. See Figure 4.

Indeed, subjects' SCS scores were a strong predictor of their level of engagement with the documentary clip. The practical significance of the predictive power of SCS can be gauged by examining its relationship to various components of the engagement index. Subjects who scored one standard deviation (84^{th} percentile) or above on the SCS were disproportionately likely to rate the clip as "very interesting" and to watch the entire clip. Subjects who scored +1 or higher on SCS were also far more likely than others to request access to the full episode from the documentary.

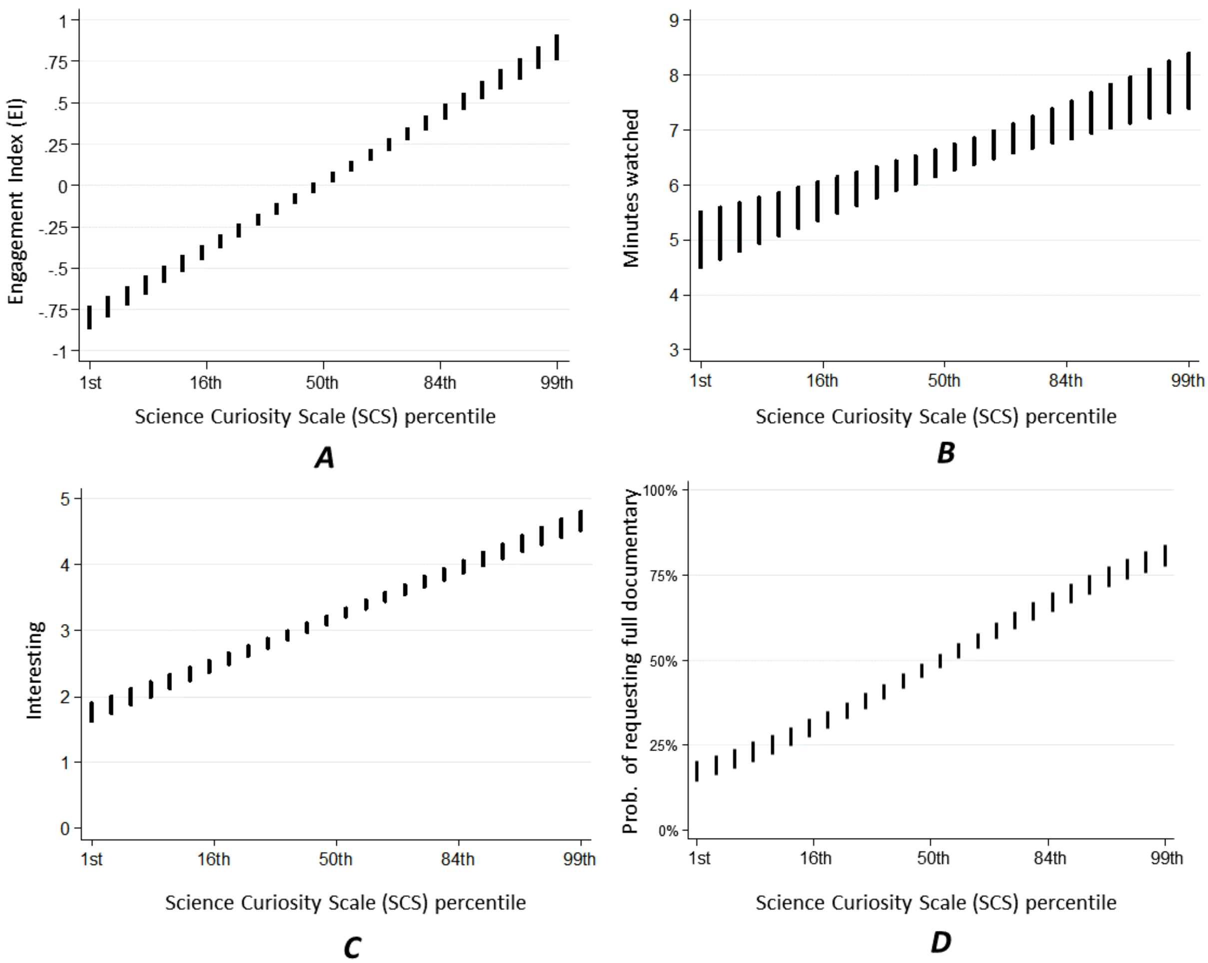

Figure 4. Engagement with the documentary clip as a function of science curiosity. N=2500 for images A, C, & D, and N=1250 for image B. Images A, B, and C are based on linear regression analyses and Image D is based on logistic regression. Bars represent 95% confidence intervals.

SCS also appeared to be a stronger predictor of subjects' engagement with the clip than did their ordinary science intelligence (OSI) scores. See Figure 5. One would expect science comprehension to predict engagement with a science documentary. However, because taking pleasure in contemplating scientific discovery and the capacity to recognize and make use of scientific evidence are distinct dispositions, one would also expect a valid science interest measure to be more discerning of engagement.

Although it is uncommon for researchers to present evidence behaviorally validating curiosity scales, investigations of such scales typically find that the disposition being measured reduces to reasoning proficiency (Loewenstein, 1994). The power of SCS to predict engagement with the clip independent of, and more powerfully than, Ordinary Science Intelligence is thus a highly desirable property of the Science Curiosity Scale.

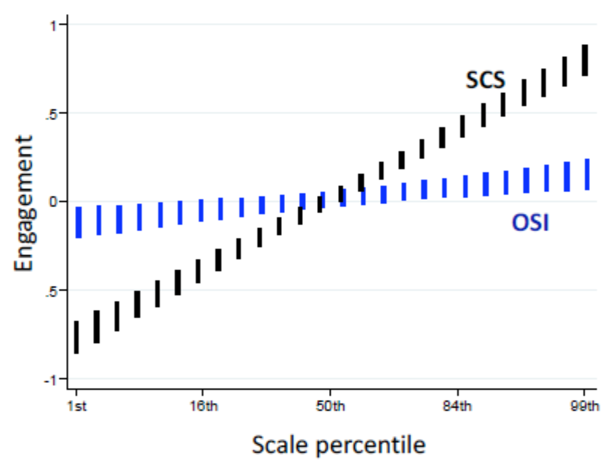

Figure 5. Relative impact of OSI and SCS on engagement with the clip. Results based on multivariate linerar regression (including SCS, OSI, and cross-product interaction predictors. Bars are 0.95 CIs.

Our study design, of course, demanded that a science curiosity measure be developed and validated independently of engagement with the documentary segment itself. Nevertheless, the power of SCS to predict how interesting subjects found the show, how much of the segment they chose to view, and how likely they were to request access to the full episode supplies additional reason to be confident that SCS does indeed measure a general science-interest disposition.

Discussion

The current study demonstrates the feasibility of constructing a valid science curiosity measure that can be used, for example, to evaluate how well science films engage those individuals most interested in contemplating the insights of scientific discovery. However, more work needs to be done, including follow-up studies that aim to continue to validate the Science Curiosity Scale.

One limitation of the scale as it is currently constructed is the amount of items required. In order to avoid some of the problems commonly associated with self-report items, we had to bury indicators of science interest in an entire battery of distractor items. Not all researchers will be able to spare the expense for this measurement. While we anticipate simplifying the measure in future studies, we also will aim to figure out how to put the distractor items to best use—for example, using items that negatively correlate with science curiosity as part of the scale.

All in all, this study demonstrates that by combining appropriately subtle self-report items with behavioral and performance items, it is possible to construct a scale that measures individuals' desire to seek out and consume scientific information for personal satisfaction. Such a measure would likely provide many contributions to the advancement of knowledge. For instance, a science curiosity measure may help improve science education by facilitating investigation of the forms of pedagogy most likely to promote learning (Blalock et al., 2008). In addition, those who study the science of science communication (Fischhoff & Scheufele, 2013; Kahan, 2015) could also use a science curiosity measure to deepen their understanding of how public interest in science shapes the responsiveness of democratically accountable institutions to policy-relevant evidence.

Acknowledgments

The authors would like to thank HHMI and Tangled Bank Studios for their assistance in developing the project as well as for allowing us to use clips from their documentary *Your Inner Fish*. They would also like to thank Kathleen Hall Jamieson and the Annenberg Public Policy Center for their support of the project.

References

Blalock, C. L., Lichtenstein, M. J., Owen, S., Pruski, L., Marshall, C., & Toepperwein, M. (2008). In pursuit of validity: A comprehensive review of science attitude instruments 1935–2005. *International Journal of Science Education, 30*(7), 961-977.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.

Fischhoff, B., & Scheufele, D. A. (2013). The science of science communication. *Proceedings of the National Academy of Sciences, 110*(Supplement_3), 14031. doi:10.1073/pnas.1312080110

Fraser, B. J. (1978). Development of a test of science-related attitudes. *Science Education, 62*(4), 509-515.

Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives, 19*(4), 25-42.

Kahan, D. M. (2014). We need a crt 2.0! And irt shoudl be used to develop it.

Kahan, D. M. (2015). What is the 'science of science communication'? *Journal of Science Communication, 14*(3), 1-10. doi:10.2139/ssrn.2562025

Kahan, D. M. (in press). 'Ordinary science intelligence': A science comprehension measure for use in the study of risk perception and science communication. *Journal of Risk Research*. doi:10.2139/ssrn.2466715

Kahan, D. M., Braman, D., Slovic, P., Gastil, J., & Cohen, G. L. (2008). Cultural cognition of the risks and benefits of nanotechnology. *Nature Nanotechnology, 4*(2), 87.

Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L. L., Braman, D., & Mandel, G. N. (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature Climate Change, 2*(10), 732-735. doi:10.1038/NCLIMATE1547

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin, 116*(1), 75.

National Science Board. (2014). *Science and technology: Public attitudes and understanding*. Arlington, VA: National Sceince Foundation.

Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education, 25*(9), 1049-1079.