

Operationalizing science literacy: an experimental analysis of measurement

Meaghan McKasy, Michael Cacciatore, Leona Yi-Fan Su, Sara K. Yeo and Liane O'Neill

Abstract

Inequalities in scientific knowledge are the subject of increasing attention, so how factual science knowledge is measured, and any inconsistencies in said measurement, is extremely relevant to the field of science communication. Different operationalizations of factual science knowledge are used interchangeably in research, potentially resulting in artificially comparable knowledge levels among respondents. Here, we present data from an experiment embedded in an online survey conducted in the United States (N = 1,530) that examined the distribution of factual science knowledge responses on a 3- vs. 5-point response scale. Though the scale did not impact a summative knowledge index, significant differences emerged when knowledge items were analyzed individually or grouped based on whether the correct response was “true” or “false.” Our findings emphasize the necessity for communicators to consider the goals of knowledge assessment when making operationalization decisions.

Keywords

Scholarly communication; Science writing

DOI

<https://doi.org/10.22323/2.19040203>

Submitted: 17th April 2020

Accepted: 10th August 2020

Published: 7th September 2020

Context

While scholars agree that factual science knowledge plays an important role in people’s attitudes toward science and technology, exactly *how* and *how much* knowledge affects attitudes remains a topic of considerable debate [e.g. Brossard and Nisbet, 2007; Simis et al., 2016; Sturgis and Allum, 2004]. Advocates of the knowledge deficit model — prominent in the fields of science, technology, engineering, and math (STEM) [Simis et al., 2016] — have persistently argued that filling gaps in scientific literacy among general audiences enables rational scientific decision-making, thus encouraging favorable scientific attitudes among publics [Allum, Sturgis et al., 2008; Kahlor and Rosenthal, 2009; Miller, 2004; Lee and Scheufele, 2006]. However, empirical research on factual knowledge as a predictor of public perceptions of science show mixed results. Some studies report a positive but weak link between scientific literacy and support for science [e.g. Brossard and

Nisbet, 2007; Priest, 2001], while others find no such linkage or only an indirect one [e.g. Brossard, Scheufele et al., 2009; Scheufele and Lewenstein, 2005]. Some work has even uncovered negative relationships between knowledge and attitudes of support for topics like biofuels [Cacciatore, Scheufele, Binder et al., 2012].

Although scholars have critiqued the deficit model as simplistic and called for more nuanced approaches [Simis et al., 2016; Sturgis and Allum, 2004], they have stopped short of disputing the importance of knowledge as a key factor both in public attitudes toward science and democratic processes of technologically-demanding modern societies [see also Brossard and Shanahan, 2006]. Knowledge as a construct is complex since it is more than just a matter of accurate understanding [Mondak and Davis, 2001]. Individuals may have varying levels of information or misinformation, the latter of which is not necessarily reflected in most typical knowledge assessments. Inequalities in scientific knowledge, particularly between different socio-economic groups [Tichenor, Donohue and Olien, 1970], are the subject of increasing attention [Cacciatore, Scheufele and Corley, 2014; Gustafson and Rice, 2016; Su et al., 2014; World Economic Forum, 2011]. As such, a better understanding of how factual science knowledge has been conceptualized and operationalized is overdue.

Objective

Determining consistent measurement scales is one of the most important research challenges in knowledge scholarship [Rich, 1991], including within the discipline of science communication. Recognizing this importance, scholars have conducted comparisons of factual knowledge indicators — typically objective true/false items — with measures of self-reported understanding, or subjective knowledge. Such work has revealed that, despite some studies using these measures interchangeably, each are unique constructs, that while weakly correlated, impact public attitudes differently [Ladwig et al., 2012; Su et al., 2014]. These findings fit with scholarship that suggests individuals tend to overestimate their abilities both socially and intellectually [Kruger and Dunning, 1999], leading to artificially high levels of perceived understanding. This divergence between objective and subjective knowledge is robust and has been studied in multiple scientific contexts including genetically modified organisms (GMOs) [Rose et al., 2019], nanotechnology [Su et al., 2014], and energy technology risks [Stoutenborough and Vedlitz, 2016].

Beyond comparisons between factual and subjective knowledge, some scholars have pointed attention toward subtler differences that may be impacting knowledge assessment [e.g. Taddicken, Reif and Hoppe, 2018]. Most public opinion scholarship that incorporates a measurement of factual knowledge asks participants' responses to a series of true/false indicators. It has been noted that publications often fail to adequately describe the measurement and recoding efforts associated with such knowledge items [Taddicken, Reif and Hoppe, 2018]. Various operationalizations of science knowledge exist, and each can be useful depending on the situational context and research goals. Some researchers may be interested in measuring confidence in understanding, while others may be focused solely on accuracy. As such, depending on the research goals, some measurements may be beneficial, while others may be limiting. For example, the National Science Board (NSB) conducts a biennial survey which contains a set of items to test American's science understanding using a 5-point true/false scale (Definitely True, Likely True,

Likely False, Definitely False, or Don't Know) [National Science Board, 2018] that can be found in many public opinion studies [e.g. Allum, Sibley et al., 2014; Yang et al., 2017]. Similarly, the General Social Survey (GSS), conducted by NORC at the University of Chicago, asks respondents similar questions using a 3-point true/false scale (True, False, or Don't Know) [Smith et al., 2019]. This scale is also popular among social scientists [e.g. Bauer, Petkova and Boyadjieva, 2000; Ho et al., 2017; Rose et al., 2019].

In an attempt to address the variation in knowledge scales, Taddicken, Reif and Hoppe [2018] tested different operationalizations of factual science knowledge by recoding survey responses into scales that are used interchangeably in studies. They found that the 5-point scale — the one employed by the NSB — measured an additional integration of confidence in one's knowledge relative to the GSS's 3-point scale, or a binary scale (True or False), the latter two of which primarily indicated accuracy. Yet, they acknowledged that recoding the variables after data collection limited their conclusions and called for future experimental research to expand upon their findings.

In response to that call, we experimentally test American responses to factual science knowledge questions using the commonly employed 3- and 5-point scales.¹ With this approach, we focus not only on whether one response scale results in greater or fewer correct answers than the other on a battery of seven factual science knowledge questions, but also whether the distinct response scales differentially influence a participant's propensity to guess at or admit ignorance of the correct responses. There are reasons to expect a different distribution of correct and incorrect responses between the two scales, although, to the best of our knowledge, such an analysis has not been undertaken. Most notably, we anticipate that the leaning options ("Likely true" and "Likely false") of the 5-point scale present a more inviting opportunity for respondents to guess at the correct response when they are unsure. Lacking the same opportunity to cautiously choose a response, we expect those exposed to the 3-point scale to opt toward the "Don't know" option when they are not confident in the correct answer. Despite this thinking, the lack of empirical evidence to support our expectations leads us to offer the following research questions:

RQ1: Will the distribution of correct responses to seven factual knowledge items differ based on whether respondents answered on a 3- or 5-point response scale?

RQ2: Are there patterns in responses to individual items based on assignment to the 3- or 5-point response scale?

Most public audiences today acquire information about science from online media [National Science Board, 2018]. Specifically, individuals rely on news media for information and analysis of science topics which are considered complex, such as nanotechnology, fracking, and synthetic biology [Anderson et al., 2014; Brossard, 2013; Yeo et al., 2017]. Research in new online technologies analyzes a theorized

¹We chose not to include a third option — the binary true/false scale that lacks the 'Don't know' option — as it does not offer respondents the opportunity to admit ignorance, thereby increasing the likelihood of a false positive due to guessing.

digital divide, or a disconnect between those individuals who have access to the internet and online resources, with those who do not [van Dijk, 2006; Warschauer, 2004]. The impact of this digital access can create unequal technology skills and social benefits [van Dijk, 2017]. Research finds that media consumption is an important factor in numerous variables such as civil participation [Shah, Rojas and Cho, 2009] or pro-environmental behaviors [Holbert, Kwak and Shah, 2003]. Less clear, however, is whether one's science media attention might interact with our knowledge response scale manipulation, resulting in higher or lower knowledge scores. Therefore, we offer a final research question:

RQ3: Will an individual's attention to science media result in different distributions of correct responses based on their assignment to the 3- or 5-point response scale?

Correlates of knowledge

Americans differ substantially in terms of their levels of factual science knowledge and many studies find differences based on demographics. For example, survey results consistently reveal gaps in factual science knowledge between men and women [e.g. Brossard and Nisbet, 2007]. Mondak and Anderson [2004] specifically looked at this gap in political knowledge, where men typically are found to have higher levels of political knowledge and discovered that this finding is partially an artifact of men's propensity to guess on questions. Similarly, the Pew Research Center found that men, on average, scored higher on science knowledge questions than women, even after controlling for education levels [Funk and Goo, 2015]. Meanwhile, the Science and Engineering Indicators from National Science Board [2018] suggest these differences depend on the specific questions selected; men performed better than women on physical science questions, but the gap narrowed for questions about biological sciences.

In addition to sex, other demographic characteristics (education, race and ethnicity, political ideology, and religious beliefs) have been associated with factual science knowledge. Unsurprisingly, the level of formal schooling and the number of science and math courses taken is positively related to factual knowledge of science [Funk and Goo, 2015]. Further, the combined results of the biennial NSB survey data from 2006 to 2016 revealed systematic differences in factual science knowledge scores by race and ethnicity. Specifically, White respondents scored the highest compared with Hispanic, Black, and other minorities across all education levels [National Science Board, 2018]. Additionally, Nisbet [2005] found that while science awareness was positively related to research support, religious beliefs were negatively associated with that same dependent variable. Finally, political ideology has played a well-established role in support of, or belief in, contentious partisan-divergent scientific topics such as climate change, evolution, and stem cell research [Mooney, 2007]. As the literature noted above suggests, demographics and individual value predispositions correlate with science knowledge. Since the focus of our analysis is how response scales impact knowledge scores, we control for these demographics and value predispositions that have been found to correlate with knowledge levels.

Data were obtained through an experiment embedded in an approximately 15-minute online survey. The survey was fielded in October 2018 with participants enrolled in Qualtrics opt-in online panels. Qualtrics partners with online market research panel providers to select survey respondents [The American Association for Public Opinion Research, 2016]. Participants were selected to meet a quota representative of the 2013 U.S. Census American Community Survey for age, sex, and geographic region variables. People with relevant backgrounds who had already agreed to participate in the research were contacted through email, text, or Qualtrics panel real-time software and offered incentives to participate in the survey.

The survey had a completion rate of 99.2%; 1,543 individuals started the survey and 1,530 completed it. Since Qualtrics invites participants through multiple avenues, including their real-time software, we do not know how many individuals received an invitation to participate. Therefore, a response rate cannot be calculated.

Experimental design and measures

This study employed a between-subjects experimental design that assigned respondents to one of two conditions. In both conditions, respondents answered seven general textbook scientific knowledge questions from the Oxford Scale of Science Literacy [Allum, Sturgis et al., 2008]. The Oxford Scale is designed to generally distinguish an individual's relevant science knowledge. In one condition ($N=755$), responses were recorded on a 5-point response scale (appearing as: Definitely True, Likely True, Likely False, Definitely False, Don't Know); in the other condition ($N=775$), responses were recorded on a 3-point scale (appearing as: True, False, Don't Know). Thus, our experimental manipulation variable (*Response Scale*) refers to whether a respondent was randomly assigned to the 3-point scale (coded as '0') or the 5-point scale (coded as '1'). To ensure proper comparisons in our analyses, we recoded the 5-point scale by combining the two true categories and the two false categories.

Distribution of Correct Responses was measured using the following seven Oxford Scale items from the 2018 NSB report: "The center of the Earth is very hot (True)," "The continents have been moving their location for millions of years and will continue to move (True)," "All radioactivity is man-made (False)," "Electrons are smaller than atoms (True)," "Lasers work by focusing sound waves (False)," "It is the father's gene that decides whether the baby is a boy or a girl (True)," and "Antibiotics kill viruses as well as bacteria (False)." These items were summed for an overall measure of knowledge in which the number of correct responses ranged from zero to seven for both the 3-point ($M = 3.85$, $SD = 1.80$) and the 5-point scale ($M = 3.86$, $SD = 1.75$). These two overall measures were also combined to create a measure of complete knowledge regardless of scale ($M = 3.85$, $SD = 1.78$). It is worth noting that the 2018 NSB report relied on 10 total knowledge items. However, we did not ask respondents the item "Does the Earth go around the Sun, or does the Sun go around the Earth (Earth around Sun)" because it was not posed using the same true/false scale and we did not ask the items "The universe began with a huge explosion (True)" and "Human beings, as we know them today,

developed from earlier species of animals (True)" as these items are often tied to one's religious beliefs [Pasek, 2018].

Sex was a nominal variable (male = 48.4%; female = 51.6%). We assessed respondents' race as a nominal variable then we recoded it into a *White* (71.4%) versus other (28.6%) variable. *Education* was measured by asking the number of years of formal education the respondent had completed ($M = 14.0$, $SD = 4.57$). *Age* was measured as a ratio variable ($M = 47.6$, $SD = 16.76$).

Religiosity was measured by asking respondents to report "how much guidance religion provides in your everyday life" on a 7-point scale from 1 = "No guidance at all" to 7 = "A great deal of guidance" ($M = 4.80$, $SD = 2.13$). *Conservative ideology* was assessed by asking respondents to report their political ideology on both social and economic issues. The two items were measured on 7-point scales, with "1" indicating "Very liberal" and "7" indicating "Very conservative" and were then averaged together into a single variable ($M = 3.77$, $SD = 1.76$, Pearson's $r = .83$).

Attention to science media was measured with a total of six items that probed a respondent's attention to science and technology information in newspapers, on television, and online. Using scales ranging from "1" ("None") to "7" ("A lot"), respondents were asked how much attention they pay to (a) "Stories related to science and technology" and (b) "Stories about scientific studies in new areas of research" when they used each of the following three mediums: newspapers (either in print or online), television (either traditional television or online television sources such as Hulu or websites of television networks), and online sources (blogs and websites, excluding social networking sites). Responses to the six items were averaged together into a single index ($M = 4.51$, $SD = 1.73$, Cronbach's $\alpha = .95$).

Analysis

The analysis was conducted using IBM SPSS Statistics 26. We used cross-tabulations to test the distribution of correct responses for both the overall distribution of knowledge responses index and the individual knowledge response measures. Additionally, we relied on hierarchical ordinary least squares (OLS) regression models to test the effects of the response scale manipulation on the distribution of correct responses after controlling for a host of other factors that have been known to correlate with the distribution of correct responses (sex, white, age, religiosity, and conservative ideology). Three regressions — one predicting the distribution of accurate answers on the full battery of response items, one predicting just the items having "true" as the correct response, and one predicting those items having "false" as the correct response — were run. Variables were entered in blocks (demographics, value predispositions, media use, knowledge scale manipulation) to determine their relative explanatory power. Interactions were created by multiplying standardized versions of the main effects variables to prevent multicollinearity between the interaction term and its component parts [Cohen et al., 2003].

Results

First, we examined whether the overall distribution of correct responses differed based on assignment to the 3- or 5-point knowledge response scale (RQ1). The results showed very similar distributions of correct responses and the chi-square

test for independence indicated no significant difference between the two scales ($\chi^2(7, 1530) = 10.32, p = .171$), although the Bonferroni test comparing column proportions — a post hoc test — did note a significant difference in the percentage of people getting 4 items correct between the two scales (see Table 1).

Table 1. Distribution of correct response scores based on assigned condition (N = 1,530).

Knowledge Score	3-point scale	5-point scale
0	4.8%	6.0%
1	6.2%	4.6%
2	11.1%	9.4%
3	18.2%	15.9%
4	24.1%	29.7%
5	15.7%	16.0%
6	12.8%	12.6%
7	7.1%	5.8%
Total	100.0%	100.0%
N	775	755

Note: significant differences are indicated by bolded text.

Next, we tested the effects of the response scale manipulation on knowledge while controlling for a host of known predictors of it. The results of the regression model predicting the full knowledge battery can be found in the first reported data column of Table 2. As the table shows, White and older respondents and those with higher reported levels of education all performed better on our 7-item knowledge scale. Also, religiosity was negatively related to knowledge, while science media attention was positively related to the dependent variable. Most importantly, the response scale provided to respondents did not impact knowledge levels. Also, of note, the interaction between the response scale and science media attention did not achieve statistical significance.

Next, we turned attention to each of the individual knowledge items rather than the summative scale. We were curious to see if there were any significant differences in response distribution across individual items based on the response scale offered, and if so, whether there might be patterns in which items had significant differences. While we did not find any significant chi-square differences in the distribution of correct or “Don’t know” responses across any of the individual items (RQ2), we found statistically significant differences in distributions of incorrect responses across two of the seven items (see Table 3).² Specifically, a significant chi-square was found for the items “All radioactivity is man-made” (false; $\chi^2(2, n=1530) = 11.69, V = .087, p = .003$) and “Antibiotics kill viruses and bacteria” (false; $\chi^2(2, n=1530) = 7.34, V = .069, p = .026$), with respondents in the 5-point scale condition providing significantly more incorrect responses (26.2% compared to 19.2% and 52.3% compared to 45.4%, respectively).

²It is worth noting that although the overall chi-square test was not significant, the post hoc Bonferroni test comparing column proportions indicated a significant difference in the distribution of “Don’t know” responses between the 3- and 5-point scale conditions for the item “*The continents have been moving their location for millions of years and will continue to move (true).*” Similarly, there was a non-significant overall chi-square test, but a significant post hoc Bonferroni test comparing scale conditions between the incorrect responses for the item “*Lasers work by focusing sound waves (false).*” These findings are reflected in Table 3.

Table 2. Regressions predicting scores on (a) the full 7-item knowledge battery (“All Items”), (b) the four knowledge items for which “true” is the correct response (“True Items”), and (c) the three knowledge items for which “false” is the correct response (“False Items”) (N = 1,530).

	<i>Knowledge Scale</i>		
	<i>All Items</i>	<i>True Items</i>	<i>False Items</i>
<i>Block 1: Demographics</i>			
Sex (female coded high)	-.01	-.02	.01
Race (White = 1; non-White = 0)	.13***	.15***	.03
Education	.13***	.10***	.10***
Age	.16***	.03	.22***
<i>Inc. R² (%)</i>	8.7***	5.9***	7.2***
<i>Block 2: Value Predispositions</i>			
Religiosity	-.19***	-.09**	-.23***
Conservative Political Ideology	.03	.00	.05*
<i>Inc. R² (%)</i>	2.3***	0.3	4.4***
<i>Block 3: Media Attention</i>			
Science Media Attention	.18***	.20***	.06*
<i>Inc. R² (%)</i>	2.7***	3.6***	0.3*
<i>Block 4: Experimental Condition</i>			
Response Scale (5-point coded high)	-.01	.03	-.05*
<i>Inc. R² (%)</i>	0.0	0.1	0.3*
<i>Block 5: Interactions</i>			
Scale X Science Media Attention	.01	.05*	-.04
<i>Inc. R² (%)</i>	0.0	0.3*	0.1
<i>Total R² (%)</i>	13.8	10.1	12.3

Notes: (1) *p < .05, **p < .01, ***p < .001;

(2) Cell entries are standardized regression coefficients, except for those in Block 5, which are before-entry standardized regression coefficients.

It should be noted that the effect size of each significant result was small. Overall, this reveals that while there were no significant overall chi-square differences in response distributions across any of the four items where “true” was the correct response, there were significant chi-square differences for two of the three items where “false” was the correct response.

We speculated that audiences might be impacted by the response scale offered, but that it primarily reveals itself for specific knowledge items (i.e., those items for which “false” is the correct response). We thus divided our summative knowledge scale based on the correct responses of the individual knowledge items. This left us with two new scales. Since there were four knowledge items for which “true” was the correct response, they were summed to create an overall measure, the “true scale,” in which the number of correct responses ranged from zero to four ($M = 2.55$, $SD = 1.24$). The second knowledge scale consisting of the three knowledge items for which “false” was the correct response (the “false scale”), measured on a scale from zero to three ($M = 1.30$, $SD = 1.06$). We then replicated the regression we ran earlier on the entire knowledge battery for both the “true” and the “false” scales.

Table 3. Distribution of responses on the individual knowledge items (N = 1,530).

Knowledge Item	3-Point Scale	5-Point Scale			Overall Significance
	Overall	Leaning Responses	Definite Responses	Overall	Pearson Chi-Square <i>p</i> -value
<i>The center of the earth is very hot (true)</i>					
Incorrect	11.1%	8.1%	4.2%	12.3%	.368
Don't know	13.3%	—	—	11.1%	
Correct	75.6%	27.0%	49.5%	76.6%	
<i>The continents have been moving their location for millions of years and will continue to move (true)</i>					
Incorrect	13.7%	7.3%	7.7%	15.0%	.105
Don't know	17.7%	—	—	13.8%	
Correct	68.6%	31.5%	39.7%	71.3%	
<i>It is the father's gene that decides whether the baby is a boy or a girl (true)</i>					
Incorrect	26.5%	10.9%	12.7%	23.6%	.313
Don't know	19.4%	—	—	18.5%	
Correct	54.2%	20.9%	37.0%	57.9%	
<i>All radioactivity is man-made (false)</i>					
Incorrect	19.2%	15.0%	11.3%	26.2%	.003
Don't know	22.8%	—	—	18.9%	
Correct	57.9%	21.1%	33.8%	54.8%	
<i>Electrons are smaller than atoms (true)</i>					
Incorrect	21.4%	11.8%	9.5%	21.3%	.303
Don't know	27.2%	—	—	24.0%	
Correct	51.4%	25.0%	29.7%	54.7%	
<i>Antibiotics kill viruses and bacteria (false)</i>					
Incorrect	45.4%	29.9%	22.4%	52.3%	.026
Don't know	13.4%	—	—	11.4%	
Correct	41.2%	12.2%	24.0%	36.3%	
<i>Lasers work by focusing sound waves (false)</i>					
Incorrect	32.1%	24.9%	12.1%	37.0%	.100
Don't know	32.3%	—	—	28.3%	
Correct	35.6%	13.0%	21.7%	34.7%	
N	775			755	

Notes:

(1) Significant differences in the overall distribution of responses between the 3- and 5-point scale are indicated by the last column, the "Pearson Chi-Square *p*-value." Significant differences in the Bonferroni test of incorrect, don't know, and correct responses between the 3- and 5-point scales are indicated by bolded text.

(2) For the 5-point scale, the information in the "Leaning Responses" column can be added to the information in the "Definite Responses" column to arrive at the values in the "Overall" column.

(3) As an example to better interpret the table above, a respondent would be in the "Incorrect" row and the "Leaning Responses" column if they chose "Likely True" for an item where false was the correct response (or if they chose "Likely False" for an item where true was the correct response), while a respondent would be in the "Incorrect" row and the "Definite Responses" column if they chose "Definitely True" for an item where false was the correct response (or if they chose "Definitely False" for an item where true was the correct response).

The results of the regressions on the “true” and “false” scales can be found in the second and third data columns of Table 2. First, it is worth noting that the results presented for the “true” and “false” scales generally replicated those present in the overall knowledge scale. For example, across all three regressions, education and science media attention positively predicted knowledge scores, while religiosity negatively predicted knowledge. Most importantly, however, we find a significant impact of the response scale on knowledge for our “false” items battery. Those exposed to the 5-point scale performed significantly worse than those exposed to the 3-point scale on the three “false” items.

In addition, the interaction between response scale and science media attention emerged as significant when examining the “true” scale knowledge measures. This finding is plotted using an ANOVA with a dichotomized attention to media variable (see Figure 1). Among low science media attention respondents, the response scales had no impact on knowledge levels. However, those who pay greater attention to science media performed significantly better when they were provided the 5-point rather than 3-point response scale (RQ3).

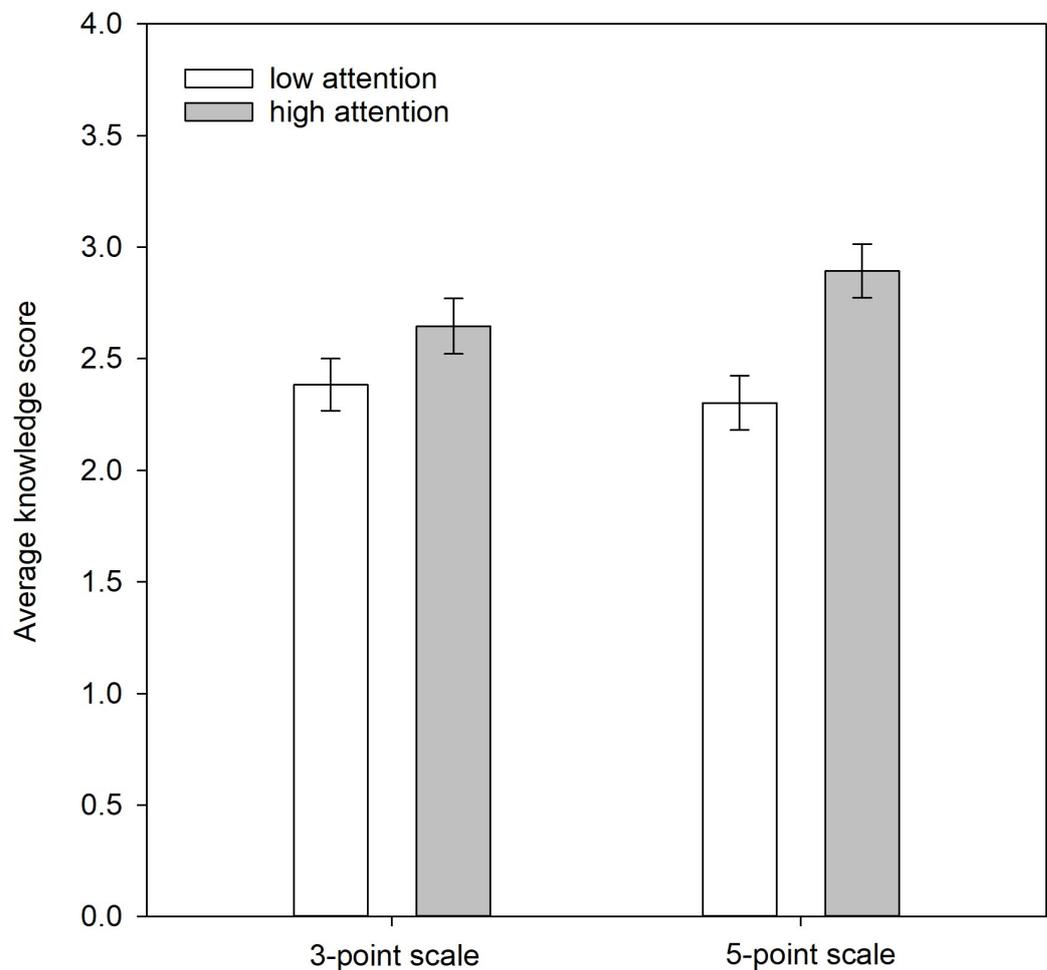


Figure 1. Interaction between Knowledge Response Scale Manipulation and Science Media Attention on “True” Knowledge Scale ($N = 1,530$). The error bars represent the 95% confidence interval.

Discussion

The present study examined potential differences in factual science knowledge levels by comparing two different knowledge response scales. We found that the use of 3- or 5-point response scales did not matter once all seven knowledge items of the Oxford Scale were summed to create an index. While this suggests that response scale choice failed to impact the distribution of correct responses — a positive consideration for social scientists — a different story emerged when (a) examining the distribution of responses of the individual knowledge items, (b) investigating indices of items that have been grouped based on having the same “true” or “false” answer, and (c) exploring response scale’s impacts on unique sub-groups in the population.

Beginning with the first of these factors, we found significant differences in the distribution of incorrect responses based on the scale for two of our seven individual questions (“All radioactivity is man-made” and “Antibiotics kill viruses as well as bacteria”). Notably, the pattern was the same between both these items with the 5-point scale resulting in more respondents getting the items incorrect. Further, these findings occurred for two of the three items for which “false” was the correct response. Based on this pattern, it may be that individuals were more likely to guess “true” when they were unsure about an item and the 5-point scale provided a *safer* guess by offering the hedging position (“Likely True”), a pattern that emerged most often when the correct response was “false.” In other words, the additional option of “Likely True” appeared to offer a comfortable alternative to “Don’t Know” when one was unsure about the correct response to a question.

Indeed, there was evidence supporting this hypothesis when examining the overall distribution of responses across the individual knowledge items. Across all seven of the items, the percentage of people who selected “Don’t know” was higher among those in the 3-point than the 5-point scale condition. Thus, respondents appeared more likely to admit that they “Don’t know” when they were presented the 3-point rather than the 5-point knowledge response scale. For example, for the statement “Lasers work by focusing sound waves,” for which the correct answer was “false,” 32.3% of individuals selected the “Don’t know” response on the 3-point scale, compared to 28.3% on the 5-point scale. The differences never quite achieved statistical significance as the 3-point scale generally resulted in only about 2–4% more “Don’t know” responses, but the pattern was consistent across all seven of the knowledge items. Further, when looking at the percentage of correct responses for the items where “true” was the right answer, we found that the 5-point scale outperformed the 3-point scale in all four cases. Conversely, if looking at the percentage of correct responses for the items where “false” was the right answer, the 3-point scale outperformed the 5-point scale across all three of those items. These patterns suggest that a small group of people were more willing to hazard a guess when they were presented the 5-point response scale, and in those situations, they tended to guess that the answer was “true.” Once again, the differences were rather small, typically no more than 2–4%, but they were consistent across all seven of the individual knowledge items.

While these small, and statistically non-significant patterns may seem unimportant, they became impactful when individual items were combined in the form of summative knowledge indices. This is best demonstrated in the regression analysis we ran on the grouped set of “false” knowledge items. Here we found a main effect of response scale such that those exposed to the 5-point scale performed

significantly worse than those exposed to a 3-point scale for the items where “false” was the correct response. In other words, the small differences across individual items became impactful when those items were grouped together to create an index. As such, the question of *how* knowledge is measured becomes an important area for future research.

But again, why should we care if these patterns were present only among items with a “false” response since they appeared to be canceled out in the overall summative scale? Although the overall summative knowledge index was not significantly impacted by the response scale manipulation, the fact that two of the individual items were significantly different depending on if the respondent received a 3- or 5-point scale, and the general pattern concerning when people were willing to either admit ignorance or guess at an answer suggest that researchers must be careful when deciding which knowledge items to include and which scale to use when building surveys. The lower percentage of “Don’t know” responses in the 5-point condition than the 3-point condition indicated that the 5-point scale may encourage guessing when respondents are uncertain of the correct response. Meanwhile, the 3-point scale seemed to encourage a more honest “Don’t know” response. These differences are critical since oftentimes research does not include as many knowledge items as the present research’s seven. Often, batteries of 5, 4, or even 3 items are utilized to assess knowledge [Ahern, Connolly-Ahern and Hoewe, 2016; Allum, Sturgis et al., 2008], due to their convenience for survey participants and concerns about the length of data collection. Therefore, the selection of specific questions, a particular response scale, and whether an item has a “true” or “false” answer, carry heavy consequences. As such, we argue that the selection of items for measurement should be a strategic act and consider the goals of the researchers.

Also, of note, we found evidence that sub-groups in the population might be differently impacted by the response scale when responding to specific knowledge questions. Our finding of the significant, positive interaction between response scale and science media attention for the “true” knowledge scale suggests that those who frequently consume science media were more likely to accurately guess for those items where “true” was the correct response and when the 5-point scale was provided. This tendency might be linked to heightened confidence among high science media consumers concerning their understanding of science. That is, their frequent consumption of science content through newspapers, television, and the internet may leave them feeling like they can figure out the correct answer when they are provided an inviting 5-point response scale that allows for leaning responses. This finding is interesting considering our earlier review of the knowledge deficit model, which has been criticized for being overly simplistic in its emphasis that more information enables rational decision-making. Yet, here we find that consumption of science content, presumably with the intent of increasing science knowledge, was related to correct knowledge item responses when true was the accurate answer.

This work is not without its limitations. First, the data for this study came from a non-probability quota sample. While a probability sample would have been preferable for generalizing about the U.S. population at large, a non-probability quota sample is still helpful to address the research questions found in this scholarship. It is also worth noting that the quota sample aligned with current U.S. Census Bureau data, giving us further confidence in its representativeness. A

second limitation concerns our choice of knowledge indicators. There is, of course, no shortage of proposed measures for tapping into factual knowledge. From this universe of options, we decided upon questions from the Oxford Scale. And, even within that battery, we narrowed our selection down to seven items that shared the same true/false response options and were less likely to be influenced by a respondent's religious beliefs. While this limits our ability to generalize to other knowledge indicators, the Oxford Scale is arguably the most common factual knowledge scale used in the social sciences [Allum, Sturgis et al., 2008]. As such, we believe these questions are reflective of typical research in this domain. Third, while we did have significant findings for two of the seven indicators, we should mention that the effect size for these findings was small. As such, we must be careful about placing too much weight on their statistical significance. Additionally, we recognize that this study may have been overpowered with 1,530 participants split among two experimental conditions, although we also believe that the consistency of the findings speaks to very real patterns in how audiences respond to the two knowledge scales.

One path for future research would be to add more conditions to this analysis. We chose not to include a third option — the binary true/false scale that lacks the 'Don't know' option — as it does not offer respondents the opportunity to admit ignorance. Previous scholarship has advised social scientists to avoid restricted-choice questions for measurements of factual knowledge [Sturgis, Allum et al., 2005]. This analysis was primarily interested in the role that the leaning options ("Likely true" and "Likely false") play in one's propensity to guess at a correct response or admit ignorance. As the binary true/false scale lacks both leaning options and an opportunity to admit ignorance, it was deemed outside the scope of our primary research interests. Yet, with that choice we also lost additional comparisons. Future research should explore how true/false choice items impact knowledge distributions, particularly considering some of the patterns uncovered in the present analysis. Finally, we recognize that the order in which the items appeared on the scales may have influenced responses. Across both of our conditions, our knowledge questions presented the true option(s) first, followed by the false option(s), with the "Don't know" option presented last. It is possible that presenting the false option(s) before the true option(s) could impact the distribution of responses, as might moving the location of the "Don't know" category.

This research built on previous scholarship by Taddicken, Reif and Hoppe [2018], which identified that knowledge is a nuanced construct for which determining appropriate measurements is challenging. Our findings show there can be advantages of using one scale over the other. While the 5-point scale allowed individuals to express certainty in their response, it also appeared to encourage guessing, if only among a small subset of respondents. Conversely, the 3-point scale may offer more of an understanding of levels of ignorance, but without the nuanced understanding of how confident respondents are in their knowledge. Since these are different ideas, we offer that scholars must question what they are trying to learn with their measurement of knowledge to help determine which scale may offer the most representative and valid data.

References

- Ahern, L., Connolly-Ahern, C. and Hoewe, J. (2016). 'Worldviews, issue knowledge and the pollution of a local science information environment'. *Science Communication* 38 (2), pp. 228–250.
<https://doi.org/10.1177/1075547016636388>.
- Allum, N., Sibley, E., Sturgis, P. and Stoneman, P. (2014). 'Religious beliefs, knowledge about science and attitudes towards medical genetics'. *Public Understanding of Science* 23 (7), pp. 833–849.
<https://doi.org/10.1177/0963662513492485>.
- Allum, N., Sturgis, P., Tabourazi, D. and Brunton-Smith, I. (2008). 'Science knowledge and attitudes across cultures: a meta-analysis'. *Public Understanding of Science* 17 (1), pp. 35–54. <https://doi.org/10.1177/0963662506070159>.
- Anderson, A. A., Brossard, D., Scheufele, D. A., Xenos, M. A. and Ladwig, P. (2014). 'The "Nasty Effect": Online Incivility and Risk Perceptions of Emerging Technologies'. *Journal of Computer-Mediated Communication* 19 (3), pp. 373–387.
<https://doi.org/10.1111/jcc4.12009>.
- Bauer, M. W., Petkova, K. and Boyadjieva, P. (2000). 'Public knowledge of and attitudes to science: alternative measures that may end the "science war"'. *Science, Technology, & Human Values* 25 (1), pp. 30–51.
<https://doi.org/10.1177/016224390002500102>.
- Brossard, D. (2013). 'New media landscapes and the science information consumer'. *Proceedings of the National Academy of Sciences* 110 (Supplement 3), pp. 14096–14101. <https://doi.org/10.1073/pnas.1212744110>. PMID: 23940316.
- Brossard, D. and Nisbet, M. C. (2007). 'Deference to Scientific Authority Among a Low Information Public: Understanding U.S. Opinion on Agricultural Biotechnology'. *International Journal of Public Opinion Research* 19 (1), pp. 24–52.
<https://doi.org/10.1093/ijpor/edl003>.
- Brossard, D., Scheufele, D. A., Kim, E. and Lewenstein, B. V. (2009). 'Religiosity as a perceptual filter: examining processes of opinion formation about nanotechnology'. *Public Understanding of Science* 18 (5), pp. 546–558.
<https://doi.org/10.1177/0963662507087304>.
- Brossard, D. and Shanahan, J. (2006). 'Do they know what they read? Building a scientific literacy measurement instrument based on science media coverage'. *Science Communication* 28 (1), pp. 47–63.
<https://doi.org/10.1177/1075547006291345>.
- Cacciatore, M. A., Scheufele, D. A. and Corley, E. A. (2014). 'Another (methodological) look at knowledge gaps and the Internet's potential for closing them'. *Public Understanding of Science* 23 (4), pp. 376–394.
<https://doi.org/10.1177/0963662512447606>.
- Cacciatore, M. A., Scheufele, D. A., Binder, A. R. and Shaw, B. R. (2012). 'Public attitudes toward biofuels: effects of knowledge, political partisanship and media use'. *Politics and the Life Sciences* 31 (1-2), pp. 36–51.
https://doi.org/10.2990/31_1-2_36.
- Cohen, J., Cohen, P., West, S. G. and Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. 3rd ed. New York, NY, U.S.A.: Routledge. <https://doi.org/10.4324/9780203774441>.

- Funk, C. and Goo, S. K. (10th September 2015). 'A look at what the public knows and does not know about science'. *Pew Research Center Report*.
URL: <https://www.pewresearch.org/science/2015/09/10/what-the-public-knows-and-does-not-know-about-science/>.
- Gustafson, A. and Rice, R. E. (2016). 'Cumulative advantage in sustainability communication: unintended implications of the knowledge deficit model'. *Science Communication* 38 (6), pp. 800–811.
<https://doi.org/10.1177/1075547016674320>.
- Ho, S. S., Yang, X., Thanwarani, A. and Chan, J. M. (2017). 'Examining public acquisition of science knowledge from social media in Singapore: an extension of the cognitive mediation model'. *Asian Journal of Communication* 27 (2), pp. 193–212. <https://doi.org/10.1080/01292986.2016.1240819>.
- Holbert, R. L., Kwak, N. and Shah, D. V. (2003). 'Environmental concern, patterns of television viewing and pro-environmental behaviors: integrating models of media consumption and effects'. *Journal of Broadcasting & Electronic Media* 47 (2), pp. 177–196. https://doi.org/10.1207/s15506878jobem4702_2.
- Kahlor, L. and Rosenthal, S. (2009). 'If we seek, do we learn?: predicting knowledge of global warming'. *Science Communication* 30 (3), pp. 380–414.
<https://doi.org/10.1177/1075547008328798>.
- Kruger, J. and Dunning, D. (1999). 'Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments'. *Journal of Personality and Social Psychology* 77 (6), pp. 1121–1134.
<https://doi.org/10.1037/0022-3514.77.6.1121>.
- Ladwig, P., Dalrymple, K. E., Brossard, D., Scheufele, D. A. and Corley, E. A. (2012). 'Perceived familiarity or factual knowledge? Comparing operationalizations of scientific understanding'. *Science and Public Policy* 39 (6), pp. 761–774.
<https://doi.org/10.1093/scipol/scs048>.
- Lee, C.-J. and Scheufele, D. A. (2006). 'The influence of knowledge and deference toward scientific authority: a media effects model for public attitudes toward nanotechnology'. *Journalism & Mass Communication Quarterly* 83 (4), pp. 819–834. <https://doi.org/10.1177/107769900608300406>.
- Miller, J. D. (2004). 'Public understanding of, and attitudes toward, scientific research: what we know and what we need to know'. *Public Understanding of Science* 13 (3), pp. 273–294. <https://doi.org/10.1177/0963662504044908>.
- Mondak, J. J. and Davis, B. C. (2001). 'Asked and answered: knowledge levels when we won't take "don't know" for an answer'. *Political Behavior* 23 (3), pp. 199–224. <https://doi.org/10.1023/A:1015015227594>.
- Mondak, J. J. and Anderson, M. R. (2004). 'The knowledge gap: a reexamination of gender-based differences in political knowledge'. *The Journal of Politics* 66 (2), pp. 492–512. <https://doi.org/10.1111/j.1468-2508.2004.00161.x>.
- Mooney, C. (2007). *The republican war on science*. New York, NY, U.S.A.: Basic Books.
- National Science Board (2018). *Science and engineering indicators 2018*. Alexandria, VA, U.S.A. URL: <https://www.nsf.gov/statistics/2018/nsb20181/>.
- Nisbet, M. C. (2005). 'The Competition for Worldviews: Values, Information, and Public Support for Stem Cell Research'. *International Journal of Public Opinion Research* 17 (1), pp. 90–112. <https://doi.org/10.1093/ijpor/edh058>.

- Pasek, J. (2018). 'It's not my consensus: motivated reasoning and the sources of scientific illiteracy'. *Public Understanding of Science* 27 (7), pp. 787–806.
<https://doi.org/10.1177/0963662517733681>.
- Priest, S. H. (2001). 'Misplaced faith: communication variables as predictors of encouragement for biotechnology development'. *Science Communication* 23 (2), pp. 97–110. <https://doi.org/10.1177/1075547001023002002>.
- Rich, R. F. (1991). 'Knowledge creation, diffusion and utilization: perspectives of the founding editor of knowledge'. *Knowledge* 12 (3), pp. 319–337.
<https://doi.org/10.1177/107554709101200308>.
- Rose, K. M., Howell, E. L., Su, L. Y.-F., Xenos, M. A., Brossard, D. and Scheufele, D. A. (2019). 'Distinguishing scientific knowledge: the impact of different measures of knowledge on genetically modified food attitudes'. *Public Understanding of Science* 28 (4), pp. 449–467.
<https://doi.org/10.1177/0963662518824837>.
- Scheufele, D. A. and Lewenstein, B. V. (2005). 'The public and nanotechnology: how citizens make sense of emerging technologies'. *Journal of Nanoparticle Research* 7 (6), pp. 659–667. <https://doi.org/10.1007/s11051-005-7526-2>.
- Shah, D. V., Rojas, H. and Cho, J. (2009). 'Media and civic participation: on understanding and misunderstanding communication effects'. In: *Media effects*. Ed. by J. Bryant and M. B. Oliver. New York, NY, U.S.A.: Routledge, pp. 223–243. <https://doi.org/10.4324/9780203877111-16>.
- Simis, M. J., Madden, H., Cacciatore, M. A. and Yeo, S. K. (2016). 'The lure of rationality: why does the deficit model persist in science communication?' *Public Understanding of Science* 25 (4), pp. 400–414.
<https://doi.org/10.1177/0963662516629749>.
- Smith, T. W., Davern, M., Freese, J. and Morgan, S. L. (2019). *General social surveys, 1972–2018*. [machine-readable data file]. U.S.A.
 URL: http://gss.norc.org/documents/codebook/gss_codebook.pdf.
- Stoutenborough, J. W. and Vedlitz, A. (2016). 'The role of scientific knowledge in the public's perceptions of energy technology risks'. *Energy Policy* 96, pp. 206–216. <https://doi.org/10.1016/j.enpol.2016.05.031>.
- Sturgis, P., Allum, N., Smith, P. and Woods, A. (2005). 'The measurement of factual knowledge in surveys'. In: *General conference of the European consortium for political research* (Budapest, Hungary).
- Sturgis, P. and Allum, N. (2004). 'Science in Society: Re-Evaluating the Deficit Model of Public Attitudes'. *Public Understanding of Science* 13 (1), pp. 55–74.
<https://doi.org/10.1177/0963662504042690>.
- Su, L. Y.-F., Cacciatore, M. A., Scheufele, D. A., Brossard, D. and Xenos, M. A. (2014). 'Inequalities in scientific understanding: differentiating between factual and perceived knowledge gaps'. *Science Communication* 36 (3), pp. 352–378.
<https://doi.org/10.1177/1075547014529093>.
- Taddicken, M., Reif, A. and Hoppe, I. (2018). 'What do people know about climate change — and how confident are they? On measurements and analyses of science related knowledge'. *JCOM* 17 (03), A01.
<https://doi.org/10.22323/2.17030201>.

- The American Association for Public Opinion Research (2016). Standard definitions: final dispositions of case codes and outcome rates for surveys. 9th ed. U.S.A.: AAPOR. URL: https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf.
- Tichenor, P. J., Donohue, G. A. and Olien, C. N. (1970). 'Mass media flow and differential growth in knowledge'. *Public Opinion Quarterly* 34 (2), p. 159. <https://doi.org/10.1086/267786>.
- van Dijk, J. A. G. M. (2006). 'Digital divide research, achievements and shortcomings'. *Poetics* 34 (4-5), pp. 221–235. <https://doi.org/10.1016/j.poetic.2006.05.004>.
- (2017). 'Digital divide: impact of access'. In: *The international encyclopedia of media effects*. U.S.A.: John Wiley & Sons Inc., pp. 1–11. <https://doi.org/10.1002/9781118783764.wbieme0043>.
- Warschauer, M. (2004). *Technology and social inclusion: rethinking the digital divide*. Cambridge, MA, U.S.A.: MIT press.
- World Economic Forum (2011). *Shared norms for the new reality*. URL: <https://www.weforum.org/events/world-economic-forum-annual-meeting-2011/>.
- Yang, X., Chuah, A. S. F., Lee, E. W. J. and Ho, S. S. (2017). 'Extending the cognitive mediation model: examining factors associated with perceived familiarity and factual knowledge of nanotechnology'. *Mass Communication and Society* 20 (3), pp. 403–426. <https://doi.org/10.1080/15205436.2016.1271436>.
- Yeo, S. K., Su, L. Y.-F., Scheufele, D. A., Brossard, D., Xenos, M. A. and Corley, E. A. (2017). 'The effect of comment moderation on perceived bias in science news'. *Information, Communication & Society* 22 (1), pp. 129–146. <https://doi.org/10.1080/1369118x.2017.1356861>.

Authors

Meaghan McKasy (Ph.D., University of Utah) is an Assistant Professor of Public Relations in the Department of Communication at Utah Valley University. She studies information processing and attitude formation in science and environmental communication. She is particularly interested in analyzing variables that influence motivation and ability in the processing of strategic messaging. E-mail: meaghan.mckasy@uvu.edu.

Michael A. Cacciatore (Ph.D., University of Wisconsin-Madison) is an Associate Professor in the Department of Advertising & Public Relations at the University of Georgia. His research examines the communication of science topics ranging from nanotechnology to vaccinations. A significant portion of this research has tracked media depictions of science issues with an additional focus on the interplay between media, values and risk in the opinion formation process. E-mail: mcacciat@uga.edu.

Leona Yi-Fan Su (Ph.D., University of Wisconsin-Madison) is an Assistant Professor in the Department of Advertising and the Institute of Communications Research at the University of Illinois at Urbana-Champaign. Her research focuses on examining the interplay between media and society, particularly how social media and new technologies influence human communication and social behaviors in the context of scientific and health topics. E-mail: lyfsu@illinois.edu.

Sara K. Yeo (Ph.D., University of Wisconsin-Madison) is an Assistant Professor in the Department of Communication at the University of Utah. Her research interests include science and risk communication with a focus on information seeking and processing. She is also trained as a bench and field scientist and holds a M.S. in Oceanography from the University of Hawai'i at Manoa.
E-mail: sara.yeo@utah.edu.

Liane O'Neill (M.S. University of Utah) is a communications officer for the state of Oregon. Her research interests are in science and environmental conservation communication with a focus on the role of emotional and temporal appeals in persuasion. She completed undergraduate degrees in marketing and journalism at the University of Nevada. E-mail: liane.oneill@utah.edu.

How to cite

McKasy, M., Cacciatore, M., Su, L. Y.-F., Yeo, S. K. and O'Neill, L. (2020). 'Operationalizing science literacy: an experimental analysis of measurement'. *JCOM* 19 (04), A03. <https://doi.org/10.22323/2.19040203>.



© The Author(s). This article is licensed under the terms of the Creative Commons Attribution — NonCommercial — NoDerivativeWorks 4.0 License.
ISSN 1824-2049. Published by SISSA Medialab. jcom.sissa.it