Supplementary Materials

for

The Reasoning through Evidence versus Advice (EvA) Scale: Scale Development and Validation

Hwayong Shin[1], Priti Shah[2], and Stephanie D. Preston[2]

[1] Department of Political Science, University of Michigan

[2] Department of Psychology, University of Michigan

Initial Scale Development and Pilot Study

To ensure item clarity and a fit between intended and inferred constructs, the full set of initial 57 items were tested using individual, in-person interviews with Introductory Psychology Subject Pool participants (n = 5) and adults without a college education that we recruited through flyers and craigslist advertisements (n = 4). The goal of the interviews was to ensure that questions were written in a manner participants could easily understand. In Table S1, the acronyms for the original scale from which the items were adopted are: Schommer Epistemological Questionnaire (SEQ, Schommer, 1998; 63-item version), Epistemic Beliefs Inventory (EBI, Schraw et al., 2004), the Updated Dogmatism Scale (UDS, Shearman & Levine, 2006), and Attitudes Toward Science Scale (ATSS, Francis & Greer, 1999). Hypothesized traits of each item are indicated as: pro-evidence (E+), anti-evidence (E-), pro-advice (A+), anti-advice (A-). Bolded items were included in the final EvA scale.

Table S1

The List of Initial EvA Items

Item	Statement	Expected Trait
1	I pay close attention to what my religious leader tells me I should do.	A
2	**When I have to vote, I see what my politician says and follow their lead.**	A
3	I generally follow my parents' advice.	A
4	**I assume that when my favorite blogger or social media personality gives advice, they know what they are talking about.**	A
5	I respect law enforcement, like police officers.	A
6	**I often make changes to my diet based on what my friends tell me is more healthy.**	A
7	My behavior is usually dictated by my religious values.	A
8	I assume my doctors know what they're talking about, so I follow their recommendations.	A
9	**When I think a politician has a confident, assertive personality, I naturally like them and vote for them.**	A
10	People who challenge authority are overconfident. (adapted from SEQ)	A
11	Sometimes you just have to accept the teacher's answer even though you don't understand it. (adapted from SEQ)	A
12	People should always respect authority. (adapted from UDS)	A
13	People who are in a position of authority have the right to tell others what to do. (adapted from UDS)	A
14	Children should be allowed to question their parents' authority. (adapted from EBI)	A-
15	I wonder how much my teachers really knew. (adapted from SEQ)	A-
16	Even advice from experts should be questioned. (adapted from SEQ)	A-
17	I'm the type of person who questions authority. (adapted from UDS)	A-
18	I am doubtful that my teachers really understood what they were teaching me.	A-
19	**Hosts of major television news shows do not know enough to be reliable sources of information.**	A-
20	Just because people are older or more experienced does not mean their claims are necessarily correct.	A-
21	**Government officials often say things that are untrue in their public statements.**	A-
22	I assume that people in positions of power are corrupt.	A-
23	**People who are telling us how to act don't always have an incentive to tell the truth.**	A-
24	From my perspective, people in positions of authority should generally not be trusted.	A-
25	Scientists' research doesn't matter in the real world.	E-
26	When I hear a news story about health, I wonder if there is really good evidence behind the assertion.	E+

27	**When someone makes a statement that sounds like a fact, I want to know the evidence behind it.**	E+
28	When someone cites a statistic, I want to know where they got it from.	E+
29	**I am concerned that news reports are based on people's opinions rather than actual evidence.**	A-
30	I pay attention to science news and try to follow the latest findings.	E+
31	I am more likely to avoid a risk when I learn about the statistics rather than personal stories and anecdotes.	E+
32	**When I hear a news story reporting research about health, I want to look up the study they are referring to.**	E+
33	Before I vote on an issue in my state or city, I try to look up the ballot items so that I vote correctly.	E+
34	**I carefully examine research on important issues to make sure it is valid and unbiased.**	E+
35	When I hear about new research, I look into who funded it to be sure it is unbiased.	E+
36	I think news reports about science should include more information so that we can evaluate the strength of the evidence.	E+
37	When my doctor tells me about a new treatment, I like to find out about any research on the treatment.	E+
38	Science is very important for the country's development. (adapted from ATSS)	E+
39	Money spent on science is well worth spending. (adapted from ATSS)	E+
40	In general, you should consider whether the information in your textbook is accurate. (adapted from SEQ)	E+
41	**When debating an important issue, I try to fact-check things that people state as statistics.**	E+
42	It is usually wise to seek out evidence and research before making decisions. (adapted from UDS)	E+
43	I believe that things that are natural are always better for you.	E-
44	I am wary of medical procedures that interfere with my body's natural processes.	E-
45	I am hesitant to take prescription medicines because they seem like chemicals I am putting into my body.	E-
46	When scientists change their minds, I stop trusting their research on what we are supposed to eat to be healthy.	E-
47	**When new evidence reverses a previous scientific theory, I just stop paying attention to it and make my own decisions.**	E-
48	**People make too much of scientific studies in the news when I know that the research is biased anyway.**	E-
49	People can talk about data, but I think that my intuitions are a better guide for my decisions.	E-
50	**Even if scientific studies are done carefully and transparently, I still don't really believe them.**	E-
51	I am not interested in looking into the details when I hear the results of a new study.	E-
52	I think scientific data is too hard to understand, so I generally ignore it.	E-
53	When it comes to controversial issues in society, I don't think "the data" can tell us much.	E-
54	I avoid GMOs and pesticides, no matter what the evidence says.	E-
55	I rarely check the nutrition facts or ingredient list on food labels.	E-
56	When choosing between products, I don't spend much time comparing the specifications.	E-
57	**Scientific evidence is overrated; there are often better ways to understand the world.**	E-

Note. Hypothesized traits of each item are indicated as: pro-evidence (E+), anti-evidence (E-), pro-advice (A+), anti-advice (A-). Bolded items were included in the final EvA scale.

The first round of interviews was conducted on undergraduate students (n = 5), using the U-M Intro Psychology Subject Pool (August 11, 2019 ~ August 15, 2019). To assess whether non-student adults without a college degree comprehend the items in accordance with our intentions, non-student participants without college education (n = 4) were invited for a second round of interviews to provide feedback on the revised questions (October 16, 2019 ~ October 28, 2019). In the second round, participants were recruited to interview regarding the scale development by flyers posted in nearby business locations and by an ad posted on Craigslist.

Both forms of advertisement used the same text asking for adults who do not have a college education to participate in our research. Participants were compensated $10 for 45 minutes of participation at the lab at the university.

Each interview was conducted by having participants complete the survey while talking through their reasoning as they answered each question. With the experimenter present, for each item participants assessed: What is the item asking you?; Is there anything confusing or ambiguous about this item?; Is there an answer choice that accurately reflects how you would like to respond? For questions that took a long time to reason through, or otherwise took a lot of effort to understand or a lot of debate back and forth between various interpretations, the participant was asked further questions about semantic definitions (i.e. What do you think of when you think of authority? What kinds of politicians are you thinking about in this question?). Later, participants described to the experimenter how well they believed the scale captured tendencies to seek or avoid evidence or authorities and suggested improvements.

It was found that many people shared similar conceptions of these terms—authority tended to refer to parents, police, and doctors, and politicians were often discussed at a national level, rather than a state or local level. Items were edited when there was non-uniformity in semantic understandings or when the wording of the question was unnecessarily difficult. The items shown in Table S2 had outstanding comments and criticisms. Other items were interpreted by interviewees as we intended, specifically whether each item pertains to individual tendencies to rely on or suspect evidence and/or authorities.

Table S2

Items Revised through Pilot Study

Item	Original item wording	Interview feedback	Revised item wording
20	I don't think the media are knowledgeable enough about the facts to rely on them for answers.	Interviewees felt that 'media' was too vague. Social media sources like facebook, twitter? Cable news? "Mainstream media" or "news media" or "the major media sources" might make this item clearer	Hosts of major television news shows do not know enough to be reliable sources of information.
25	I generally ignore scientists when they are talking about their research.	Some felt that this was less about the scientists than about the research, despite the wording seeming to focus on the scientists. Something that singles out and dismisses "scientists" would avoid this.	Scientists' research doesn't matter in the real world.
33	Before I vote on an issue in my state or city, I try to look up the details so that I vote correctly.	One thought this item was asking about whether you look up details about voting procedures ("When, where, how do I vote correctly?); another disagreed with the item because of lack of interest in politics/voting.	Before I vote on an issue in my state or city, I try to look up the ballot items so that I vote correctly.
40	In general, you should evaluate the accuracy of information in a textbook.	Mixed views on this item. The impression was that participants recognized that this is a pro-evidence item, but most wouldn't agree/strongly agree even if they were strongly pro-evidence people, because they wouldn't take the time to do fact-check textbooks or because they view textbooks as classic/time-honored sources (e.g. a college calculus textbook) not worth checking.	In general, you should consider whether the information in your textbook is accurate.

Study 1

Data Collection and Descriptive Statistics

Study 1 was conducted via the survey platform CloudResearch between March 13, 2021 and March 21, 2021. To collect similar number of respondents with and without college education, we used CouldResearch's prescreening data to recruit half of the participants without college degree, and another half with a college degree. To ensure response quality, approval rate was set at 95% and the number of approved HIT's was set to be greater than or equal to 100.

Table S3

Distribution of Demographic Variables in Study 1 (n = 549)

Variable	Distribution
Gender	Male (1) = 46.4%; Female (2) = 52.6%; Prefer to self-identify (3) = 0.9%
Race	White (1) = 83.1%; Black or African American (2) = 6.9%; American Indian or Alaska Native (3) = 0.5%; Asian (4) = 6.7%; Native Hawaiian or other Pacific Islander (5) = 0.0%; Other (e.g., mixed) (6) = 2.7%
Age	Age 18-24 = 7.3%; Age 25-34 = 24.4%; Age 35-44 = 27.7%; Age 45-54 = 18.0%; Age 55-64 = 14.2%; Age 65 or older = 8.4%
Education	No high school diploma (1) = 1.1%; High school diploma (2) = 42.4%; Some college, no degree (3) = 3.8%; Associate degree (4) = 8.4%; Bachelor's degree (5) = 26.8%; Master's degree (6) = 13.7%; Professional degree (7) = 2.6%; Doctorate degree (8) = 1.3%
Partisan Identity	Strong Democrat (1) = 19.5 %; Weak Democrat (2) = 20.4%; Democratic leaner (3) = 9.5%; Independent (4) = 14.8%; Republican leaner (5) = 7.1%; Weak Republican (6) = 15.3%; Strong Republican (7) = 13.5%
Ideology	Very liberal (1) = 12.4%; Liberal (2) = 18.0 %; Slightly liberal (3) = 13.3% ; Moderate (4) = 25.1%; Slightly conservative (5) = 9.3%; Conservative (6) = 15.5%; Very conservative (7) = 6.4%

Additional Details for Exploratory Factor Analysis

We chose common factors analysis (FA) rather than principal components analysis (PCA), as FA is more suitable to understand the latent constructs that explain shared variance among items in scale development (Fabrigar et al., 1999). In Study 1, for exploratory factor analysis, we used an oblique rotation for factor rotation. Fabrigar et al. (1999) suggest that using oblique rotation reduces the number of cross-factor loadings, producing superior simple structure ("cleaner" solutions). Both Fabrigar et al. (1999) and Marcus et al. (2006) recommend using oblique rotation because using orthogonal rotation will forfeit any knowledge of existing correlations among factors. Brown (2015) says that oblique rotation is preferred in most cases because it provides a more realistic representation of how factors are interrelated. We used maximum Likelihood as the model-fitting procedure, following Fabrigar et al. (1999), Cudeck & O'Dell (1994), and Brown (2015), who suggested using ML as the model-fitting procedure because it provides a wide range of fit indices and allows computation of correlations among factors, thus preferred over principal factors procedure.

Skewness (mean = –0.22, range = –1.35-1.42) and kurtosis per item (mean = 3.13, range = 1.73-5.58) suggested the appropriateness of the maximum likelihood (ML) factor extraction

procedure (Baker et al., 2010; Fabrigar et al., 1999; normality violated if skewness > 2, kurtosis > 7). Worthington and Whittaker (2006) recommend that EFA be followed by examining the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy. The KMO test examines whether the associations among items can be accounted for by a smaller set of factors (Ferguson & Cox, 1993), thus indicating whether the data is adequate for examining meaningful factor structures, rather than chance correlations among a small subset of items (Worthington & Whittaker, 2006). KMO values of .60 and higher are recommended for reliable factor analysis (Beavers et al. 2013; Tabachnick & Fidell, 2001). Our KMO was .91, which suggested an adequate sample size for EFA on 57 items. Because it is advised not to include additional scales at the early stage of scale development, especially when there are a high number of initial items compared to the final scale (Worthington & Whittaker, 2006), this study did not contain any other scales (e.g., convergent or discriminant). No items had to be removed due to highly skewed distributions (> ±2.0; Cassidy et al., 2005).

Table S4 shows the list of questionnaire items that were included in Study 1 for exploratory factor analysis. Hypothesized dimensions behind the construction of each item are indicated as: pro-evidence (E+), anti-evidence (E-), pro-advice (A+), anti-advice (A-). The acronyms for the original scale from which the items were adopted are: Schommer Epistemological Questionnaire (SEQ, Schommer, 1998; 63-item version), Epistemic Beliefs Inventory (EBI, Schraw et al., 2004), the Updated Dogmatism Scale (UDS, Shearman & Levine, 2006), and Attitudes Toward Science Scale (ATSS, Francis & Greer, 1999). Bold-faced items were selected for the final EvA scale.

Table S4

EvA Items and Exploratory Factor Analysis Results: Study 1 (n = 549)

			Factors								
Item	Statement	Expected Trait	1	2	3	4	5	6	Skew	Kurt	Comm
57	**Scientific evidence is overrated; there are often better ways to understand the world.**	**E-**	0.84						0.80	2.91	1.1
25	Scientists' research doesn't matter in the real world.	E-	0.78						1.42	5.06	1.1
50	**Even if scientific studies are done carefully and transparently, I still don't really believe them.**	**E-**	0.76						0.78	3.08	1.1
48	**People make too much of scientific studies in the news when I know that the research is biased anyway.**	**E-**	0.62						0.26	2.13	1.5
47	**When new evidence reverses a previous scientific theory, I just stop paying attention to it and make my own decisions.**	**E-**	0.57						0.36	2.18	1.2
7	My behavior is usually dictated by my religious values.	A	0.53	0.31	0.39				0.45	1.73	2.7
46	When scientists change their minds, I stop trusting their research on what we are supposed to eat to be healthy.	E-	0.51						0.31	2.33	1.6
52	I think scientific data is too hard to understand, so I generally ignore it.	E-	0.49	-0.3					1.03	3.63	2

53	When it comes to controversial issues in society, I don't think "the data" can tell us much.	E-	0.46						0.52	2.55	1.6
49	People can talk about data, but I think that my intuitions are a better guide for my decisions.	E-	0.46						0.28	2.31	1.8
1	I pay close attention to what my religious leader tells me I should do.	A	0.39		0.39				0.50	1.87	3.4
8	I assume my doctors know what they're talking about, so I follow their recommendations.	A	-0.32				0.32	-0.31	-1.11	4.74	4.2
42	It is usually wise to seek out evidence and research before making decisions. (adapted from UDS)	E+	-0.37	0.3		0.31			-1.03	4.63	3.1
39	Money spent on science is well worth spending. (adapted from ATSS)	E+	-0.83						-0.90	3.68	1.3
38	Science is very important for the country's development. (adapted from ATSS)	E+	-0.94				0.31		-1.35	5.26	1.3
32	**When I hear a news story reporting research about health, I want to look up the study they are referring to.**	**E+**		0.85					-0.84	3.45	1.2
34	**I carefully examine research on important issues to make sure it is valid and unbiased.**	**E+**		0.84					-1.09	4.78	1.1
35	When I hear about new research, I look into who funded it to be sure it is unbiased.	E+		0.71					-0.63	2.69	1.3
41	**When debating an important issue, I try to fact-check things that people state as statistics.**	**E+**		0.68					-0.90	4.17	1.1
27	**When someone makes a statement that sounds like a fact, I want to know the evidence behind it.**	**E+**		0.67					-1.02	4.47	1.2
28	When someone cites a statistic, I want to know where they got it from.	E+		0.61					-1.17	4.66	1.5
30	I pay attention to science news and try to follow the latest findings.	E+		0.61					-0.67	3.04	2.5
33	Before I vote on an issue in my state or city, I try to look up the ballot items so that I vote correctly.	E+		0.57					-1.21	4.50	1.1
37	When my doctor tells me about a new treatment, I like to find out about any research on the treatment.	E+		0.57					-1.24	5.58	1.3
26	When I hear a news story about health, I wonder if there is really good evidence behind the assertion.	E+		0.42					-0.87	4.07	1.9
36	I think news reports about science should include more information so that we can evaluate the strength of the evidence.	E+		0.39		0.36			-0.69	3.73	2.6
40	In general, you should consider whether the information in your textbook is accurate. (adapted from SEQ)	E+		0.3					-0.77	3.47	3.4
51	I am not interested in looking into the details when I hear the results of a new study.	E-		-0.57					0.82	3.02	1.5
5	I respect law enforcement, like police officers.	A			0.77				-1.04	3.46	1.1
12	People should always respect authority. (adapted from UDS)	A			0.69				-0.30	2.36	1.3
3	I generally follow my parents' advice.	A			0.42				-0.59	2.87	2
13	People who are in a position of authority have the right to tell others what to do. (adapted from UDS)	A			0.35				-0.26	2.31	2.2

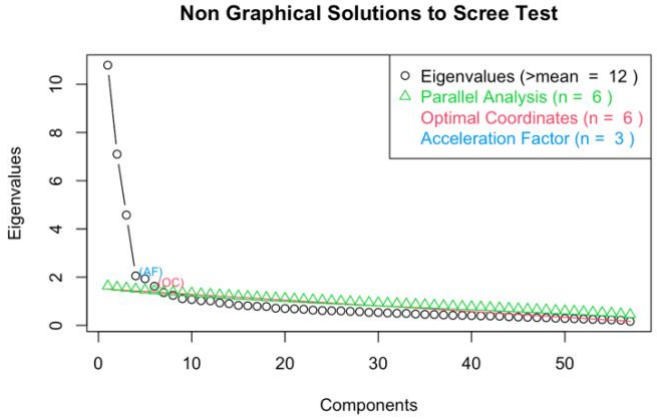
15	I wonder how much my teachers really knew. (adapted from SEQ)	A-	-0.30	0.33			-0.25	2.24	2.4
18	I am doubtful that my teachers really understood what they were teaching me.	A-	-0.37				0.48	2.57	2.8
22	I assume that people in positions of power are corrupt.	A-	-0.37	0.43			-0.07	2.44	2.3
24	From my perspective, people in positions of authority should generally not be trusted.	A-	-0.54				0.36	2.85	1.8
14	Children should be allowed to question their parents' authority. (adapted from EBI)	A-	-0.56				-0.11	2.23	1.6
17	I'm the type of person who questions authority. (adapted from UDS)	A-	-0.61				-0.25	2.40	1.5
29	**I am concerned that news reports are based on people's opinions rather than actual evidence.**	**A-**		0.66			-0.73	2.79	1.3
21	**Government officials often say things that are untrue in their public statements.**	**A-**		0.66			-0.69	3.42	1.1
19	**Hosts of major television news shows do not know enough to be reliable sources of information.**	**A-**		0.61			-0.35	2.47	1.1
23	**People who are telling us how to act don't always have an incentive to tell the truth.**	**A-**		0.49			-0.56	2.96	1.4
16	Even advice from experts should be questioned. (adapted from SEQ)	A-		0.41			-0.75	3.63	2.1
20	Just because people are older or more experienced does not mean their claims are necessarily correct.	A-		0.32			-1.16	4.88	2.7
2	**When I have to vote, I see what my politician says and follow their lead.**	**A**			0.64		0.24	2.23	1.2
6	**I often make changes to my diet based on what my friends tell me is more healthy.**	**A**			0.55		0.80	3.02	1.5
9	**When I think a politician has a confident, assertive personality, I naturally like them and vote for them.**	**A**			0.54		0.18	2.40	1.3
4	**I assume that when my favorite blogger or social media personality gives advice, they know what they are talking about.**	**A**			0.51		0.46	2.43	1.4
11	Sometimes you just have to accept the teacher's answer even though you don't understand it. (adapted from SEQ)	A			0.35		-0.23	2.17	2.2
10	People who challenge authority are overconfident. (adapted from SEQ)	A			0.31		0.25	2.44	2.4
44	I am wary of medical procedures that interfere with my body's natural processes.	E-				0.79	-0.64	2.63	1.2
45	I am hesitant to take prescription medicines because they seem like chemicals I am putting into my body.	E-				0.76	0.01	1.78	1.1
43	I believe that things that are natural are always better for you.	E-				0.63	-0.57	2.64	1.6
54	I avoid GMOs and pesticides, no matter what the evidence says.	E-				0.53	0.11	2.01	1.7
56	When choosing between products, I don't spend much time comparing the specifications.	E-					0.94	3.67	2.9
55	I rarely check the nutrition facts or ingredient list on food labels.	E-					0.71	2.35	3.8
31	I am more likely to avoid a risk when I learn about the statistics rather than personal stories and anecdotes.	E+					-0.55	3.17	2.9

Note. Factor loadings smaller than .3 are not displayed. The bolded items were retained for the EvA scale. *Skew* = skewness; *Kurt* = Kurtosis; *Comm* = Communality. Hypothesized traits of each item are indicated as: pro-evidence (E+), anti-evidence (E-), pro-advice (A+), anti-advice (A-).

Regarding assessing the number of factors, while the Kaiser test (Kaiser 1960) is a widely used criteria (finding number of factors through the number of eigenvalues greater than 1), we considered a scree plot and a parallel analysis because the eigenvalue rule can be too generous as a basis of retaining factors (DeVellis, 2006). The scree plot and parallel analysis both suggested six factors (Figure S1; Cattell, 1966; Hayton et al., 2004).

Figure S1

Parallel Analysis of Study 1



Following recommended item deletion criteria (Baker et al., 2010; Haws et al., 2012; Svedholm-Häkkinen & Lindeman, 2017; Worthington & Whittaker, 2006), eight items that cross-loaded on more than one factor > .30 (Items 1, 7, 8, 22, 36, 38, 42, 52) were dropped along with ten items whose highest loading was < .40 (Items 10, 11, 13, 15, 18, 20, 31, 40, 55, 56). No factors needed to be removed due to having fewer than three items (Baker et al., 2010; Brown 2015; Tabachnick & Fidell, 2001), with at least four items loading > .40 per factor. In terms of item communality, items with low communalities (< .40) are not highly correlated with one or more of the factors in the solution and should be dropped (Tabachnick & Fidell, 2001). Table S4 presents the full EFA results, including factor loadings, kurtosis, skewness, and communalities. Communalities of remaining items ranged between 1.0 to 2.5, thus none were dropped on the basis of communalities.

Item Retention Decisions

Because some of the factors contained a relatively large number of items, the last step of EFA was to shorten the scale, where researchers often aim for a balanced scale with similar number of items per factor (Baker et al., 2010; Worthington & Whittaker, 2006). Optimizing the scale length is recommended for the efficiency of the scale (e.g., respondent fatigue) despite its tradeoff of sacrificing a certain degree of internal consistency (Worthington & Whittaker, 2006) Since EFA is "a combination of empirical and subjective approaches to data analysis" with the goal of arriving at a solution that makes sense (Worthington & Whittaker, 2006), we actively employed both empirical and substantive rationales in item selection decisions. To have a

balanced number of items across factors, because Factors 5 and 6 contained four items, we aimed to select four items for four other dimensions that contained more than four items. Following the literature (Brown, 2015; Worthington & Whittaker, 2006), we deleted items that (a) have the lowest factor loadings, (b) have the highest cross-loadings, (c) contribute to internal consistency the least, and (d) have low conceptual consistency with other items. Among the four criteria, there was no item to drop based on cross-loadings because cross-loading items were dropped in the previous step. Employing a mix of criteria (a), (c), and (d), we made the following decisions.

Table S5

Change in Reliability after Removal of Each Item

Factor 1	Alpha if item deleted	**Factor 2**	Alpha if item deleted	**Factor 3**	Alpha if item deleted	Factor 4	Alpha if item deleted
Item 57	.83	Item 32	.79	Item 5	.70	Item 29	.68
Item 25 (-)	**.85**	Item 34	.78	Item 12	.69	Item 21	.68
Item 50	.83	**Item 35**	**.82**	**Item 24 (-)**	**.78**	Item 19	.70
Item 48	.84	Item 41	.81	Item14 (-)	.76	Item23	.73
Item 47	.85	Item 27	.80	Item17 (-)	.73	**Item 16**	**.74**
5-item Alpha	*.87*	*5-item Alpha*	*.83*	*5-item Alpha*	*.78*	*5-item Alpha*	*.75*

For Factor 1 (Anti-evidence), the four lowest loading items (Items 46, 49, 53, 39) were removed. Compared to the retained items that tap onto individuals' general tendency of how they approach "scientific evidence" (Items 50, 25, 57, 48, 47), the removed items were relevant to "data," "healthy eating," or "spending on science," thus our decision to remove them was supported in terms of conceptual consistency. Although Item 39 strongly loaded on Factor 1 with a negative factor loading, we did not retain it because its focus (spending on science) was quite distant from the overall substantive meaning of this factor. Among the five retained items, we employed the criterion (c), removing item that contributes the least to internal consistency, which is indicated by the expected increase in Cronbach's alpha for the subscale if the item is removed (Gliem & Gliem, 2003; Raubenheimer, 2004). The item analysis in Table S5 suggested that the deletion of either Item 25 or Item 47 would deteriorate internal reliability the least. Given the concerns that the item deletion decision solely based on alpha could be imperfect (Raykov, 2008), we additionally considered (d) conceptual consistency, and decided to drop Item 25, because it refers to "scientists" whereas all other remaining items are about "scientific evidence," "scientific studies," or "evidence on scientific theory." Another conceptual reason behind this decision was that Item 25 was the only reverse-coded item that we originally expected to load on the Anti-advice factor because it was intended to capture individuals' resistance toward "scientists" as advisory figures. Based on these considerations, Items 50, 57, 48, and 47 were chosen for Factor 1.

For Factor 2 (Pro-evidence), the six lowest loading items (Items 28, 30, 33, 37, 26, 51) were removed. The removed items were conceptually distinct from retained items that were closely related to the general attitude toward evidence ("examine research for validity and unbiasedness," "check evidence behind claims," "look up study behind a news story"), whereas three of the deleted items (Items 33, 37, 51) were relevant to specific topics such as "medical treatment" or "voting," and one of them (Item 51) was a reverse-coded item originally intended to load onto as the Anti-evidence dimension. Because it is recommended to avoid redundancy among the retained items (Aluja et al., 2006), dropping Item 26 was reasonable because its content overlapped with one of the retained items, Item 32 (both items were about the tendency

to check evidence behind a health news story). We also took notice of the redundancy between Item 28 and Item 41, which were conceptually identical, both asking about the tendency to check evidence behind other people's use of statistics. Between the two items, we removed Item 28 that had smaller loading. We also removed Item 30 not only because of its relatively smaller loading but also because it pertained to habitual reading of science news, whereas the other retained items were relevant to general tendencies to seek out evidence as they encounter new information or claims. Among the five remaining items (Items 32, 34, 35, 41, 27), we employed (c) reliability criterion, which suggested the deletion of Item 35 would lower Cronbach's alpha the least. The deletion of Item 35 was conceptually reasonable as well, because it captured individuals' interest in "funding" source of information, whereas other retained items involved tendencies to examine or seek further evidence to check the validity of information or claims at hand. Following these decisions, Factor 2 consisted of Items 34, 32, 41, and 27.

For Factor 3 (Pro-advice 1), there were six items loaded onto the factor, where Items 3 and 14 conceptually overlapped – both items were about "parents" as advisory figures. Between the two, we dropped Item 3 because its factor loading had smaller magnitude than that of Item 14. Among the remaining five items, to decide which one of the two items should be removed, we employed (c) reliability criterion, which suggested that dropping Item 24 would lower internal reliability of the scale the least. Among the items that loaded on Factor 3, Items 17 and 14 were originally intended to capture the Anti-advice reasoning tendency, thus their factor loadings suggested they be reverse-coded on this factor. It is possible in the exploratory stage of scale development that researchers may encounter some items that load onto a factor different from their original expectations, which requires considerations of whether the set of items that load together share a conceptual meaning that reasonably constitutes a single construct (e.g., Newton et al., 2021; Svedholm-Häkkinen & Lindeman, 2017). We reasoned that this factor would make a conceptual sense as a scale that captures individuals' tendency to rely on and follow advisers if Items 17 and 14 are reverse-coded. However, relatively complex nature of Factor 3 led us to consider Factor 5, on which the items we originally designed to capture Pro-advice tendency loaded together, as a better candidate for Pro-advice dimension of the EvA scale. Thus, we tentatively named Factor 3 as Pro-advice 1, to compare with Factor 5, which we named as Pro-advice 2.

For Factor 4 (Anti-advice), five items were loaded on this factor. The (c) reliability criterion suggested that dropping Item 16, which had the smallest factor loading, would lower internal reliability of the scale the least. The remaining four items conceptually had a shared meaning – a tendency to resist relying on others' opinions. One of the items, Item 29, was originally designed to capture Pro-evidence tendency, but our reassessment of the item in light of other items on this factor suggested that this item was closely related to Anti-advice tendency, tapping onto the tendency to resist opinion-based news reports. Based on these considerations, Items 29, 21, 19, 23 consisted Factor 4.

Factor 5 (Pro-advice 2) and Factor 6 (Anti-medicine, Anti-evidence 2) contained four items that meaningfully loaded on each factor. Four items that loaded on Factor 5 (Items 2, 6, 9, 4) were all relevant to individuals' tendency to rely on and follow preferred advisers, such as politician, friends, and celebrities. Four items on Factor 6 (Items 44, 45, 43, 54), although they

were originally developed to capture Anti-evidence tendency, were specifically related to individuals' aversion to medicine and chemical.

Table S6

Exploratory Factor Analysis: Factor Loadings of Four Items Selected per Factor (Subset of Table S4)

Item	Statement	Expected Trait	Factors					
			1	2	3	4	5	6
57	**Scientific evidence is overrated; there are often better ways to understand the world. (E-)**	**Anti-evidence 1**	.84					
50	**Even if scientific studies are done carefully and transparently, I still don't really believe them. (E-)**		.76					
48	**People make too much of scientific studies in the news when I know that the research is biased anyway. (E-)**		.62					
47	**When new evidence reverses a previous scientific theory, I just stop paying attention to it and make my own decisions. (E-)**		.57					
32	**When I hear a news story reporting research about health, I want to look up the study they are referring to. (E+)**	**Pro-evidence**		.85				
34	**I carefully examine research on important issues to make sure it is valid and unbiased. (E+)**			.84				
41	**When debating an important issue, I try to fact-check things that people state as statistics. (E+)**			.68				
27	**When someone makes a statement that sounds like a fact, I want to know the evidence behind it. (E+)**			.67				
5	I respect law enforcement, like police officers. (A)	Pro-advice 1			.77			
12	People should always respect authority. (A, adapted from UDS)				.69			
14	Children should be allowed to question their parents' authority. (A-, adapted from EBI)				-.56			
17	I'm the type of person who questions authority. (A-, adapted from UDS)				-.61			
29	**I am concerned that news reports are based on people's opinions rather than actual evidence. (A-)**	**Anti-advice**				.66		
21	**Government officials often say things that are untrue in their public statements. (A-)**					.66		
19	**Hosts of major television news shows do not know enough to be reliable sources of information. (A-)**					.61		
23	**People who are telling us how to act don't always have an incentive to tell the truth. (A-)**					.49		
2	**When I have to vote, I see what my politician says and follow their lead. (A)**	**Pro-advice 2**					.64	
6	**I often make changes to my diet based on what my friends tell me is more healthy. (A)**						.55	
9	**When I think a politician has a confident, assertive personality, I naturally like them and vote for them. (A)**						.54	
4	**I assume that when my favorite blogger or social media personality gives advice, they know what they are talking about. (A)**						.51	
44	I am wary of medical procedures that interfere with my body's natural processes. (E-)	Anti-evidence 2						.79
45	I am hesitant to take prescription medicines because they seem like chemicals I am putting into my body. (E-)							.76
43	I believe that things that are natural are always better for you. (E-)							.63
54	I avoid GMOs and pesticides, no matter what the evidence says. (E-)							.53
		Alpha	.85	.82	.78	.74	.68	.76

Note. Entries are standardized factor loadings. Factor loadings smaller than .3 are not displayed. Bolded items were retained; Letters in parentheses indicate *a priori* dimensions from item development: E+ (Pro-evidence), E- (Anti-evidence), A+ (Pro-advice), A- (Anti-advice).

Each subscale had acceptable internal reliability (Cronbach's alpha): Factor 1 (Anti-evidence): .85; Factor 2 (Pro-evidence): .82; Factor 3 (Pro-advice 1): .78; Factor 4 (Anti-advice): .74, Factor 5 (Pro-advice 2): .68; Factor 6 (Anti-medicine): .76 (DeVellis, 2017; Tavakol & Dennick, 2011). Based on the rationales for factor retention that we explain in the main text, we retained Factors 1, 2, 4 and 5. For validation in Study 2, we retained four items per factor chosen based on the considerations above, in order to create a smaller, more efficient 16-item scale that was balanced by factor.

In the main text, we also use AIC to assess the model fit. AIC adjusts χ^2 for the number of estimated parameters, allowing us to compare non-nested competing models, with lower AIC suggesting a better model fit (Schermelleh-Engel et al., 2003). The conceptual distinctness of our scale compared to other similar prior ones is bolstered by the fact that all of items taken from existing scales in our initial, larger set were not retained in the final scale.

Confirmatory Factor Analysis: Studies 1-3

For Study 1, skewness per item (mean = –0.22, range = –1.35 ~ 1.42) and kurtosis per item (mean = 3.13, range = 1.73-5.58) suggested the appropriateness of the maximum likelihood (ML) factor extraction procedure (Baker et al., 2010; Fabrigar et al., 1999; normality violated if skewness > 2, kurtosis > 7). Study 2 (skewness per item was mean = –0.25, range = –1.32 ~1.35, kurtosis per item: mean = 3.02, range = 1.76 ~ 5.36), Study 3 (skewness per item was mean = –0.30, range = –1.43 ~0.87, kurtosis per item: mean = 3.07, range = 1.91 ~ 6.36). ML procedure was appropriate for Studies 2 and 3, because none of the items had an exceedingly skewed distribution to suggest removal (e.g., skewness > ±2; Cassidy et al. 2005).

Table S7

Confirmatory Factor Analysis and Item-total Correlations: Studies 1-3

	Study 1		Study 2		Study 3	
	Factor loadings	Item-total correlation	Factor loadings	Item-total correlation	Factor loadings	Item-total correlation
Pro-evidence						
When I hear a news story reporting research about health, I want to look up the study they are referring to.	.70	.63	.67	.74	.65	.54
I carefully examine research on important issues to make sure it is valid and unbiased.	.76	.68	.78	.66	.66	.56
When debating an important issue, I try to fact-check things that people state as statistics	.72	.61	.75	.64	.70	.59
When someone makes a statement that sounds like a fact, I want to know the evidence behind it.	.78	.67	.83	.57	.65	.53
Anti-evidence						
Scientific evidence is overrated; there are often better ways to understand the world.	.77	.70	.85	.61	.83	.73
Even if scientific studies are done carefully and transparently, I still don't really believe them.	.81	.73	.76	.72	.73	.66
People make too much of scientific studies in the news when I know that the research is biased anyway.	.78	.68	.70	.64	.75	.67

When new evidence reverses a previous scientific theory, I just stop paying attention to it and make my own decisions.	.72	.64	.63	.71	.70	.61
Pro-advice						
When I have to vote, I see what my politician says and follow their lead.	.65	.53	.68	.44	.70	.60
I often make changes to my diet based on what my friends tell me is more healthy.	.53	.42	.61	.52	.60	.53
When I think a politician has a confident, assertive personality, I naturally like them and vote for them.	.60	.45	.69	.50	.81	.65
I assume that when my favorite blogger or social media personality gives advice, they know what they are talking about.	.59	.44	.64	.40	.61	.54
Anti-advice						
I am concerned that news reports are based on people's opinions rather than actual evidence.	.72	.59	.55	.57	.59	.48
Government officials often say things that are untrue in their public statements.	.66	.57	.67	.51	.61	.47
Hosts of major television news shows do not know enough to be reliable sources of information.	.69	.55	.64	.57	.56	.46
People who are telling us how to act don't always have an incentive to tell the truth.	.50	.42	.52	.52	.60	.46
CFA fit statistics						
CFI	.954		.928		0.928	
TLI	.943		.911		0.912	
SRMR	.046		.069		0.071	
RMSEA	.049		.063		0.060	
$\chi^2(df)$	228.89 (98)		172.35 (98)		209.99 (98)	
N	547		189		316	

Note. Entries for factor loadings are standardized and all were statistically significant ($p < .01$).

For scale homogeneity, items are retained with item-total correlations above .3 (Streiner et al., 2015; e.g., Duckworth et al., 2007; Lipkus et al., 2001) and dropped for being below .2 (Ames et al., 2005; Morof et al., 2012). Across studies, all items contributed to the homogeneity of relevant underlying constructs.

Dynamic Fit Index Cutoffs: Studies 1-3

Following McNeish & Wolf (2023), we additionally used the dynamic fit index (DFI) cutoffs that are calculated based on simulations and tailored to the specific model and data at hand. The dynamic cutoffs are calculated using the package "dynamic" in R. Overall, the model fit statistics of three studies (CFI, SRMR, RMSEA in Table S7) were acceptable at least one or two levels of misspecificatitableons (CFI ≥ DFI cutoffs; SRMR, RMSEA ≤ DFI cutoffs).

Table S8

Dynamic Fit Index Cutoffs: Studies 1-3

	Study 1			Study 2			Study 3		
	SRMR	RMSEA	CFI	SRMR	RMSEA	CFI	SRMR	RMSEA	CFI
Level 1: 95/5	.057	.050	.956	NONE	NONE	NONE	.070	.490	.954
Level 2: 95/5	.077	.065	.933	.093	.073	.914	.089	.067	.932
Level 3: 95/5	.106	.088	.881	.120	.098	.855	.114	.098	.871

Note. Levels 1, 2, and 3 indicates cutoffs for a small, medium, and large misspecification respectively (McNeish & Wolf, 2023, p. 82). NONE indicates there are no dynamic fit index cutoff values that distinguish between well-fitting and ill-fitting models for that specific level of misspecification (McNeish & Wolf, 2023, p. 74).

Factor Correlations: Studies 1-3

Table S9

Correlations among the EvA Subscales: Studies 1-3

	1	2	3	4
Study 1 (N = 547)				
1. Pro-evidence	(.82, .82)			
2. Anti-evidence	–.19***	(.85, .85)		
3. Pro-advice	–.12**	.36***	(.68, .68)	
4. Anti-advice	.27***	.24***	–.22***	(.74, .75)
Study 2 (N = 189)				
1. Pro-evidence	(.84, .84)			
2. Anti-evidence	–.19***	(.83, .83)		
3. Pro-advice	–.06	.32***	(.75, .75)	
4. Anti-advice	.36***	.24***	–.28***	(.68, .69)
Study 3 (N = 316)				
1. Pro-evidence	(.75, .76)			
2. Anti-evidence	–.08	(.84, .84)		
3. Pro-advice	.04	.52***	(.78, .78)	
4. Anti-advice	.26***	.12**	–.20***	(.68, .68)

Note. Entries are bivariate correlations among EvA subscales with Cronbach's alpha (first entry, calculated using ltm package in R) and McDonald's omega total (second entry, calculated using MBESS package in R) on the diagonal in parentheses. ***$p < .01$, **$p < .05$, *$p < .10$.

Study 2

Data Collection and Descriptive Statistics

Study 2 was conducted via the survey platform CloudResearch between November 26, 2019 and November 27, 2019. Participants were conducted with the same set of targeting and quality controls as Study 1 (targeting by education, approval rate, number of approved HITs). Data collection for Study 2 preceded (November 2019) that of Study 1 (March 2021) for the following reason. Our earlier interpretation of EFA on Study 2 data (n=189) suggested a three-factor EvA construct, so we proceeded with follow-up data collection to validate 12 candidate items in 2020. However, through a review process, we later learned that the sample size of 189 was too small relative to the number of our initial items (58 items). To ensure reliable EFA, we decided to conduct the survey for Study 1 on a larger sample. This time, we determined the sample size on the basis of the guidelines on the minimum ratios of participants to items (5:1 or 10:1) for exploratory factor analysis (Gorsuch, 1983; Worthington & Wittaker, 2006). Given that we were at the early stage of scale development, we targeted a participant-item ratio of 10:1 and aimed to recruit 580 participants in Study 1, conducted in March 2021. Because the data

collected in November 2019 contained all 58 items, the dataset was still useful as an independent sample to confirm the reliability of the CFA results, thus was included in our paper as Study 2.

Table S10

Distribution of Demographic Variables in Study 2 (n = 189)

Variable	Distribution
Gender	Male (1) = 51.9%; Female (2) = 48.1%
Race/Ethnicity	White/Caucasian (1) = 75.1%; African American (2) = 9.0%; Latino (3) = 8.5%; East Asian (4) = 3.70%; Native/Alaskan (5) = 1.1%; Other (6) = 2.6%
Age	Age 18-24 = 9.0%; Age 25-34 = 45.0%; Age 35-44 = 23.8%; Age 45-54 = 10.1%; Age 55-64 = 10.6%; Age 65 or older = 1.6%
Education	No high school diploma (1) = 2.1%; High school diploma (2) = 41.8%; Some college (3) = 9.5%; College degree (4) = 36.5%; Some post-graduate work (5) = 1.6%; Post-graduate degree (6) = 8.5%
Income	Less than $5,000 (1) = 5.8%; $5,000 to $6,999 (2) = 1.1%; $7,000 to $7,499 (3) = 1.1%; $7,500 to $9,999 (4) = 1.6%; $10,000 to $12,499 (5) = 2.1%; $12,500 to $14,999 (6) = 2.6%;$15,000 to $19,999 (7) = 5.8%; $20,000 to $24,999 (8) = 9.5%; $25,000 to $29,999 (9) = 7.4%; $30,000 to $34,999 (10) = 6.3%; $35,000 to $39,999 (11) = 6.3%; $40,000 to $49,999 (12) = 6.9%; $50,000 to $59,999 (13) = 13.2%; $60,000 to $74,999 (14) = 6.3%; $75,000 to $84,999 (15) = 6.3%; $85,000 to $99,999 (16) = 4.8%; $100,000 to $124,999 (17) = 5.8%; $125,000 to $149,999 (18) = 2.1%; $150,000 to $174,999 (19) = 1.6%; $175,000 or more (20) = 3.2%
Social Ideology	Very Liberal (1) = 16.4%; Liberal (2) = 27.5%; Moderate (3) = 28.0%; Conservative (4) = 21.2%; Very conservative (5) = 6.9%
Economic Ideology	Very Liberal (1) = 13.3%; Liberal (2) = 21.8%; Moderate (3) = 27.1%; Conservative (4) = 31.4%; Very conservative (5) = 6.4%

Model Comparisons: Studies 1-3

Table S11

Fit Statistics for the Proposed and Alternative Models: Studies 1-3

	$\chi^2(df)$	$\chi^2_{diff}(\Delta df)$	RMSEA	SRMR	CFI	TLI
Study 1 (N = 547)						
Proposed model	228.9 (98)		.049	.046	.954	.943
Alternative model A	477.4 (104)	248.5 (6)***	.081	.132	.867	.847
Alternative model B	1725.6 (104)	1496.8 (6)***	.169	.174	.424	.336
Alternative model C	930.7 (103)	701.8 (5)***	.121	.123	.706	.658
Alternative model D	1495.8 (103)	1267.0 (5)***	.157	.163	.506	.424
Alternative model E	367.1 (103)	138.2 (5)***	.068	.095	.906	.891
Alternative model F	461.9 (103)	233.0 (5)***	.080	.135	.873	.852
Study 2 (N = 189)						
Proposed model	172.4 (98)		.063	.069	.928	.911
Alternative model A	293.7 (104)	121.3 (6)***	.098	.149	.815	.787
Alternative model B	763.4 (104)	591.0 (6)***	.183	.194	.358	.259
Alternative model C	428.9 (103)	256.5 (5)***	.129	.134	.683	.630
Alternative model D	585.1 (103)	412.8 (5)***	.157	.162	.530	.453
Alternative model E	243.9 (103)	71.5 (5)***	.085	.106	.863	.840
Alternative model F	269.4 (103)	97.0 (5)***	.092	.137	.838	.811

Study 3 (N = 316)						
Proposed model	210.0 (98)		.060	.071	.928	.912
Alternative model A	384.3 (104)	174.3 (6)***	.092	.150	.819	.791
Alternative model B	868.7 (104)	658.7 (6)***	.153	.152	.507	.431
Alternative model C	538.2 (103)	328.2 (5)***	.116	.109	.719	.673
Alternative model D	697.3 (103)	487.4 (5)***	.135	.144	.617	.553
Alternative model E	272.4 (103)	62.4 (5)***	.072	.092	.891	.873
Alternative model F	269.4 (103)	59.4 (5)***	.092	.137	.838	.811

Note. *Proposed model:* Items load on four factors (Pro-evidence, Anti-evidence, Pro-advice, Anti-advice); *A:* No relationships between factors; *B:* All items load on one factor; *C:* Items load on two factors (Evidence-oriented, Advice-oriented); *D:* Items load on two factors (Pro/anti-evidence, Pro/anti-advice, where Anti-items are reverse-coded); *E:* Items load on four first-order factors, with two second-order factors as specified in model C; *F:* Items load on four first-order factors, with two second-order factors as specified in model D. **p < .1; **p < .05; ***p < .01.

Another alternative is a bifactor model, where a general factor reflecting accepting or rejecting information sources with specific residual factors reflecting independent tendencies to accept or reject information coming from scientific evidence or evidence (Flora, 2020). This alternative model produces the following model fits: Study 1 (χ^2(df) = 329.93 (88), RMSEA = .071, SRMR = .094, CFI = .914, TLI = .883), Study 2 (χ^2(df) =211.18 (88), RMSEA = .086, SRMR = .116, CFI = .880, TLI = .836), Study 3 (χ^2(df) = 206.75 (88), RMSEA = .065, SRMR = .082, CFI = .923, TLI = .896). Overall, the bifactor model produces worse model fits compared to the proposed model and its RMSEA, SRMR, and TLI do not meet the recommended criteria (Bentler, 1990, McDonald & Ho, 2002). Another concern with this alternative is that, for identification purposes, it forces the inter-factor correlations to 0 (Flora, 2020). Substantively, this assumption is incompatibles with how we conceptualize the EvA tendencies, where we assume the tendencies on can be correlated to each other in how people reason through evidence or advice (e.g., it is possible that more pro-advice individuals tend to be more pro-evidence, as Table S9 suggest). Thus, we propose the EvA tendencies as a four-dimensional construct that consists of Pro-evidence, Anti-evidence, Pro-advice, and Anti-advice.

Study 3

Data Collection and Descriptive Statistics

Study 3 was conducted via the survey platform Prolific between June 18, 2021 and June 19, 2021. Participants were conducted via the online crowdsourcing platform Prolific, with a target to recruit an equal number of individuals with and without a college degree.

Because we planned to examine factor structure invariance between two demographic subgroups (female, male), we needed twice the sample size required for reasonable factor analysis (respondent-item ratio of 10:1). Because the EvA scale has 16 items, we needed 160 respondents per subgroup, so we preregistered to collect 320 respondents.

Table S12

Distribution of Demographic Variables in Study 3 (n = 316)

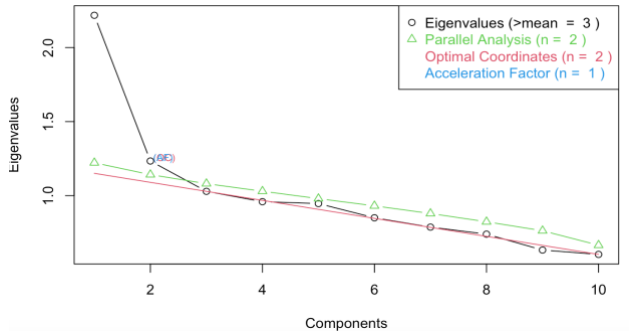
Variable	Distribution
Gender	Male (1) = 51.3%; Female (2) = 45.9%; Prefer to self-identify (3) = 2.8%
Race	White (1) =65.5%; Black or African American (2) = 14.2%; American Indian or Alaska Native (3) = 1.6%; Asian (4) = 12.0%; Native Hawaiian or other Pacific Islander (5) = 0.6%; Other (e.g., mixed) (6) = 6.0%
Age	Age 18-24 = 37.0%; Age 25-34 = 38.9%; Age 35-44 = 14.6%; Age 45-54 = 5.7%; Age 55-64 = 2.5%; Age 65 or older = 1.3%
Education	No high school diploma (1) = 2.5%; High school diploma (2) = 19.6%; Some college, no degree (3) = 26.9%; Associate degree (4) = 1.6%; Bachelor's degree (5) = 28.2%; Master's degree (6) = 18.0%; Professional degree (7) = 1.3%; Doctorate degree (8) = 1.9%
Religion	Protestant (1) = 13.9%; Roman Catholic (2) = 24.1%; Orthodox Christian (3) = 3.8%; Mormon (4) = 0.9%; Jewish (5) = 0.9%; Muslim (6) = 1.9%; Buddhist (7) = 1.9%; Hindu (8) = 1.6%; Atheist (9) = 15.5%; Agnostic (10) = 14.6%; Other (11) = 7.3%; Nothing in particular (12) = 13.6%
Partisan Identity	Strong Democrat (1) = 29.1 %; Weak Democrat (2) = 24.1%; Democratic leaner (3) = 13.6%; Independent (4) = 14.2%; Republican leaner (5) = 3.8%; Weak Republican (6) = 7.3%; Strong Republican (7) = 7.9%
Ideology	Very liberal (1) = 20.6%; Liberal (2) = 27.6%; Slightly liberal (3) = 12.4% ; Moderate (4) = 18.4%; Slightly conservative (5) = 5.4%; Conservative (6) = 11.7%; Very conservative (7) = 3.8%

Measurement Issue with Social Desirability

In Study 3, we used a 10-item scale that was developed by Strahan & Gerbasi (1972), as suggested by Fischer & Fick (1993) (items are displayed in Table S13). Only after we conducted Study 3, we noticed that Ramanaiah & Martin (1980) suggested two components of social desirability: 1) Attribution: The tendency to attribute socially desirable characteristics; 2) Denial: The tendency to deny socially undesirable characteristics. Given the controversy on the dimensionality of social desirability in the literature (e.g., Hart et al., 2015; Helmes & Holden, 2003), we used a parallel analysis to assess how the items are related to the underlying construct, which suggested that there exists two underlying factors.

Figure S2

Parallel Analysis of the Social Desirability Items



We then used exploratory factor analysis with an oblique rotation to assess how the items loaded on two different factors. It suggested two sets of items that loaded on Attribution (Factor 1) and Denial (Factor 2) respectively, similar to what Ramanaiah & Martin (1980) suggested. Although this two-factor solution could be due to the method effect (i.e., the reverse-coded nature of half of the items), it was still tricky to figure out how we should construct the scale, particularly because one of the items on Factor 1 had relatively weak loading (< .4), and three of the items did not meaningfully load on either factor. While Fischer & Fick (1993) recommended the Strahan & Gerbasi (1972)'s scale, other studies have argued that Strahan & Gerbasi's scale might be less reliable than other measures of social desirability (e.g., Reynolds, 1982). Some have suggested that older measures of social desriability may not suit contemporary society, and have thus suggested using alternative set of items (Stöber, 2001, Hart et al., 2015).

Table S13

EFA on social desirability items in Study 3

Item	Statement	Factor 1	Factor 2
1	I'm always willing to admit it when I make a mistake.		0.42
2	I always try to practice what I preach.		
3	I never resent being asked to return a favor.		
4	I have never been irked when people expressed ideas very different from my own.		0.69
5	I have never deliberately said something that hurt someone's feelings.		0.42
6	I like to gossip at times. (R)		
7	There have been occasions when I took advantage of someone. (R)	0.64	
8	I sometimes try to get even rather than forgive and forget. (R)	0.36	
9	At times I have really insisted on having things my own way. (R)	0.43	
10	There have been occasions when I felt like smashing things. (R)	0.51	

Note: Factor loadings smaller than .3 are not displayed.

Furthermore, we also noticed that the social desirability scale we used in Study 3 contained a number of argument-related items (Items 1, 4, and 9), which could misleadingly inflate its correlation with some of our EvA scale. Therefore, we decided to administer an alternative measure of social desirability—based on a more recent study that identifies the two-factor structure of social desirability, in the context outside the realm of reasoning and argumentation (Impression Management dimension in Hart et al. 2015)— in Study 4 to more reliably test discriminant validity that the EvA traits relative to social desirability.

Correlations among the EvA subscales and Convergent/Discriminant Scales

Construct validity refers to the extent that an operationalization measures the construct it purports to measure (Campbell & Fiske, 1959). It can be assessed by whether the given measure is associated with other indicators in a way that conforms to theoretical expectations, through convergent and divergent validity. Convergent validity is established through a strong association with measures that are theoretically similar or overlapping whereas discriminant validity is achieved when theoretically distinct constructs are less associated (Adcock & Collier, 2001).

Table S14

Correlations among the EvA Subscales and Convergent/Discriminant Scales

		EvA scale			
		Pro-evidence	Anti-evidence	Pro-advice	Anti-advice
Convergent	Need for Cognition	.39***	–.11*	–.03	.06
	Distrust in Science	–.13**	.78***	.36***	.08
	Respect for Convention	–.01	.56***	.45***	–.11*
	Defiance to Authority	–.02	–.43***	–.53***	.31***
Divergent	Numeracy	.05	–.31***	–.32***	.10*
	Pessimism	.02	–.02	–.08	.22***
	Dispositional Trust	–.02	–.09	.04	–.23***
	Dispositional Distrust	.02	.09	–.04	.23***

Note. *$p < .1$; **$p < .05$; ***$p < .01$. The cells that evaluate convergent and discriminant validity of each EvA subscale are shaded.

Measurement Invariance and Internal Reliability of EvA Subscales among Subgroups

Table S15

Confirmatory factor analysis using the entire sample (N= 316), non-college sample (n=160), college sample (n=156)

	Factor											
	Pro-evidence			Anti-evidence			Pro-advice			Anti-advice		
Item	All	No college	College	All	No college	College	All	No college	College	All	No college	College
1	.65	.67	.59									
2	.66	.70	.61									
3	.70	.69	.70									
4	.65	.67	.67									
5				.83	.79	.82						
6				.73	.81	.68						
7				.75	.67	.76						
8				.70	.62	.70						
9							.70	.61	.72			
10							.60	.43	.64			
11							.81	.78	.81			
12							.61	.42	.68			
13										.59	.59	.60
14										.61	.59	.59
15										.56	.47	.67
16										.60	.61	.60
Alpha	.75	.77	.73	.84	.81	.83	.78	.65	.80	.68	.65	.71
M	.75	.73	.77	.33	.26	.41	.43	.35	.50	.68	.69	.67
SD	.15	.16	.14	.23	.19	.24	.22	.18	.23	.17	.16	.18

Note. All of the standardized factor loadings are statistically significant ($p < 0.01$). *M* = mean; *SD* = standard deviation. Means and standard deviations are based on coarse factor scores (i.e., composite score as the average of four items, scaled to range from 0 to 1).

Table S16

Tests for measurement invariance between education groups

Model	χ^2 (df)	χ^2_{diff} (Δdf)	CFI	ΔCFI	RMSEA	SRMR	TLI
Configural invariance	322.56 (196)		0.913		0.064	0.071	0.893
Metric invariance	335.48 (208)	12.92 (12)	0.912	-0.001	0.062	0.075	0.898
Scalar invariance	352.06 (220)	16.58 (12)	0.909	-0.004	0.062	0.076	0.900
Residual invariance	389.74 (236)	37.68 (16)***	0.894	-0.019	0.064	0.080	0.892

Note. *p < .1; **p < .05; ***p < .01; non-college = 160, college = 156. ΔCFI compared to the base model (Configural invariance).

Table S17

Confirmatory factor analysis using the entire sample (N= 316), female sample (n=145), male sample (n=162).

	Factor											
	Pro-evidence			Anti-evidence			Pro-advice			Anti-advice		
Item	All	Female	Male	All	Female	Male	All	Female	Male	All	Female	Male
1	.65	.67	.64									
2	.66	.62	.71									
3	.70	.70	.69									
4	.65	.63	.70									
5				.83	.86	.83						
6				.73	.66	.76						
7				.75	.65	.82						
8				.70	.75	.67						
9							.70	.72	.69			
10							.60	.55	.63			
11							.81	.76	.82			
12							.61	.60	.60			
13										.59	.57	.63
14										.61	.59	.62
15										.56	.54	.61
16										.60	.68	.52
Alpha	.75	.74	.77	.84	.82	.85	.78	.76	.78	.68	.69	.68
M	.75	.74	.76	0.33	0.31	0.36	0.43	0.38	0.47	0.68	0.68	0.69
SD	.15	.15	.14	0.23	0.21	0.24	0.22	0.20	0.22	0.17	0.17	0.17

Note. All of the standardized factor loadings are statistically significant (p < 0.01). *M* = mean; *SD* = standard deviation. Means and standard deviations are based on coarse factor scores (i.e., composite score as the average of four items, scaled to range from 0 to 1).

Table S18

Tests for Measurement Invariance between Gender Groups

Model	χ^2 (df)	χ^2_{diff} (Δdf)	CFI	ΔCFI	RMSEA	SRMR	TLI
Configural invariance	321.26 (196)		0.918		0.065	0.077	0.900
Metric invariance	336.51 (208)	15.24 (12)	0.916	-0.002	0.063	0.081	0.903
Scalar invariance	347.53 (220)	11.02 (12)	0.917	-0.001	0.061	0.083	0.909
Residual invariance	371.67 (236)	24.14 (16)*	0.912	-0.006	0.061	0.085	0.910

Note. *p < .1; **p < .05; ***p < .01; female = 145, male = 162. ΔCFI compared to the base model (Configural invariance).

Cronbach's α of some of the EvA subscales for certain subgroups is slightly lower than the recommended value of .7. However, this can be evaluated as acceptable given that the scale consists of the relatively low number of items (4 items per scale) and that the 'unacceptable level' is defined as Cronbach's alpha 'below .60.' (Peterson, 1994; Price & Mueller, 1986). Cronbach (1951) suggested that a high value of alpha is 'desirable', but also made the point that the key point should be that scores obtained when using an instrument had to be interpretable—and this was often possible without needing very high values of alpha. Although adding more items into an instrument can increase the value of alpha, he pointed out that adding additional items that measure the same thing may result in a redundancy that is inefficient (Taber 2018, p.1288). Schmitt (1996) also suggested that "there is no general level where alpha becomes acceptable, but rather that instruments with quite a low value of alpha can still prove useful in some circumstances."

Study 4

Data Collection and Descriptive Statistics

Study 4 was conducted via the survey platform Prolific between January 14, 2022 and January 15, 2022. Participants were conducted via the online crowdsourcing platform Prolific, with a target to recruit an equal number of individuals with and without a college degree. To address the potential issue with gender imbalance at Prolific (Charalambides, 2021), the recruitment was balanced across gender groups.

We had a number of rationales for power analysis, following Gelman & Hill (2007). In prior studies that examined our criterion measures for scale development, the magnitudes of correlations between the target construct and behavioral measures were: 1) Science curiosity reduced scales RS, RF (Motta et al. 2021): Study 1 (r = .25~ .31, with OSI); Study 2 (r = .20~.24, with OSI) (average r = .26); 2) Perceived accuracy/influence of misinformation (Sherer et al., 2020, Table E in supplement): $|r|$ = .17~.26 (with science reasoning, health literacy, education, healthcare system trust, reflective thinking) (average r = .22); 3) Adherence to COVID behavioral guidelines (Graupensperger et al. 2020): r = .12 ~ .49 (average r = .28). Using a moderately conservative approach, we assumed .17 (smallest significant correlation in Sherer et al. 2020) as the expected effect size in power analysis. Power analysis using 'pwr' package in R with the effect size of .17, .05 significance level, and 0.8 power suggested n = 270. To ensure enough power for two subpopulations of our interest (i.e., college, non-college), we preregistered our plan to collect a total of 540 respondents.

Table S19

Distribution of Demographic Variables in Study 4 (n = 529)

Variable	Distribution
Gender	Male (1) = 48.2%; Female (2) = 50.9%; Prefer to self-identify (3) = 0.9%
Race	White (1) = 82.2%; Black or African American (2) = 7.2%; American Indian or Alaska Native (3) = 0.4%; Asian (4) = 5.9%; Native Hawaiian or other Pacific Islander (5) = 0.2%; Other (e.g., mixed) (6) = 4.2%
Age	Age 18-24 = 19.5%; Age 25-34 = 30.8%; Age 35-44 = 20.0%; Age 45-54 = 12.5%; Age 55-64 = 10.8%; Age 65 or older = 6.4%

Education	No high school diploma (1) = 0.6%; High school diploma (2) = 13.2%; Some college, no degree (3) = 26.1%; Associate degree (4) = 9.8%; Bachelor's degree (5) = 35.0%; Master's degree (6) = 12.1%; Professional degree (7) = 2.3%; Doctorate degree (8) = 0.9%
Income	Less than $10,000 (1) = 4.2%; $10,000 to $19,999 (2) = 6.8%; $20,000 to $29,999 (3) = 10.4%; $30,000 to $39,999 (4) = 9.1%; $40,000 to $49,999 (5) = 7.0%; $50,000 to $59,999 (6) = 10.8%; $60,000 to $69,999 (7) = 7.6%; $70,000 to $79,999 (8) = 9.6%; $80,000 to $89,999 (9) = 5.5%; $90,000 to $99,999 (10) = 4.5%; $10,0,000 to $149,999 (11) = 14.6%; More than $150,000 (12) = 10.0%
Religion	Protestant (1) = 16.8%; Roman Catholic (2) = 14.4%; Orthodox Christian (3) = 2.1%; Mormon (4) = 0.2%; Jewish (5) = 3.4%; Muslim (6) = 1.1%; Buddhist (7) = 1.1%; Hindu (8) = 0.6%; Atheist (9) = 16.3%; Agnostic (10) = 20.0%; Other (11) = 7.8%; Nothing in particular (12) = 16.3%
Partisan Identity	Strong Democrat (1) = 27.7 %; Weak Democrat (2) = 22.5%; Democratic leaner (3) = 15.7%; Independent (4) = 12.5%; Republican leaner (5) = 5.7%; Weak Republican (6) 9.8%; Strong Republican (7) = 6.1%
Ideology	Very liberal (1) = 19.0%; Liberal (2) = 26.2%; Slightly liberal (3) = 15.4% ; Moderate (4) = 17.8%; Slightly conservative (5) = 8.0%; Conservative (6) = 9.9%; Very conservative (7) = 3.8%
Social media usage	Facebook (1) = 68.0%; Twitter (2) = 53.5%; Reddit (3) = 29.1%; Instagram (4) = 62.7%; YouTube (5) = 77.0%; Do not use any kind of social media (9) = 2.4% *Note:* Sum of percentages exceed 100%, because multiple choices were allowed (i.e., "check all that apply")

Robustness Check for the Criterion Validity Test

Following the preregistration, we additionally run OLS regression models on the relationships between the EvA scale and criterion variables (susceptibility to misinformation, adherence to social distancing, confidence in COVID vaccine, science curiosity, and religiosity) while controlling for demographic variables. The direction and statistical significance of the relationships largely remained the same even after controlling for age, gender, education, and income.

Table S20

Relationships between the EvA tendencies and Criterion Behaviors/Attitudes (Robustness check with control variables)

EvA Reasoning Tendencies	Perceived accuracy of health misinformation		Adherence to the CDC guide on social distancing		Confidence in COVID vaccine	
	b (SE)	*t*	*b (SE)*	*t*	*b (SE)*	*t*
Pro-evidence	0.02 (.05)	0.4	0.25 (0.08)	3.0***	0.19 (0.07)	2.5**
Anti-evidence	0.40 (.04)	9.8***	–0.25 (0.06)	–4.1***	–0.78 (0.06)	–14.0***
Pro-advice	0.16 (.05)	3.5***	0.03 (0.07)	0.4	0.37 (0.0)	5.4***
Anti-advice	–0.06 (.05)	–1.2	–0.12 (0.07)	–1.8*	–0.20 (0.07)	–2.5***
Age	–0.00 (.001)	–0.1	0.003 (0.001)	3.6***	–0.0002 (0.001)	–0.2
Female	0.02 (.01)	1.1	0.01 (0.02)	0.6	0.01 (0.02)	0.6
Education	–0.004 (.01)	–0.8	0.004 (0.01)	0.4	0.01 (0.01)	1.7*
Income	0.002 (0.002)	1.1	–0.01 (0.003)	–1.6*	0.01 (0.003)	2.3**
Constant	0.09 (0.07)	1.3	0.60 (0.10)	6.1***	0.68 (0.10)	6.6***
N	524		524		524	
Adjusted R^2	.27		.10		.34	

EvA Reasoning Tendencies	Science Curiosity		Religiosity	
	b (SE)	*t*	*b (SE)*	*t*
Pro-evidence	0.63 (0.05)	11.8***	0.08 (0.08)	1.0
Anti-evidence	–0.14 (0.05)	–3.0***	0.41 (0.07)	5.9***
Pro-advice	0.08 (0.06)	1.4	0.24 (0.08)	2.9***
Anti-advice	–0.01 (0.06)	-0.1	–0.15 (0.08)	–1.9*
Age	0.001 (0.001)	0.9	0.002 (0.001)	2.8***
Female	–0.05 (0.02)	–2.8***	0.04 (0.02)	1.7*
Education	0.02 (0.01)	2.9***	0.01 (0.01)	0.9
Income	0.01 (0.003)	2.2***	0.01 (0.004)	1.6
Constant	–0.05 (0.07)	–0.7	-0.03 (0.10)	-0.3
N	524		524	
Adjusted R^2	.26		.15	

Note. *b* = unstandardized regression coefficient robust with standard errors in parentheses. *t* = t-value for regression coefficient. To facilitate comparisons of coefficients, the four EvA tendencies and dependent variables were scaled to range from 0 to 1. *Age* ranged between 18 and 78. *Female* was a binary variable, 1 if female, 0 if male. *Education* and *Income* were coded as shown in Table S19. *p < .10; **p < .05; ***p < .01.

Given that the four EvA tendencies are correlated to each other, the analysis above provides the association between each EvA tendency with the criterion variables while controlling for the other three EvA tendencies (Table 4, Table S20). We also provide the bivariate correlations between the EvA scales and the criterion validity measures in Table S21.

Table S21

Bivariate Correlations between the EvA Subscales and Criterion Scales

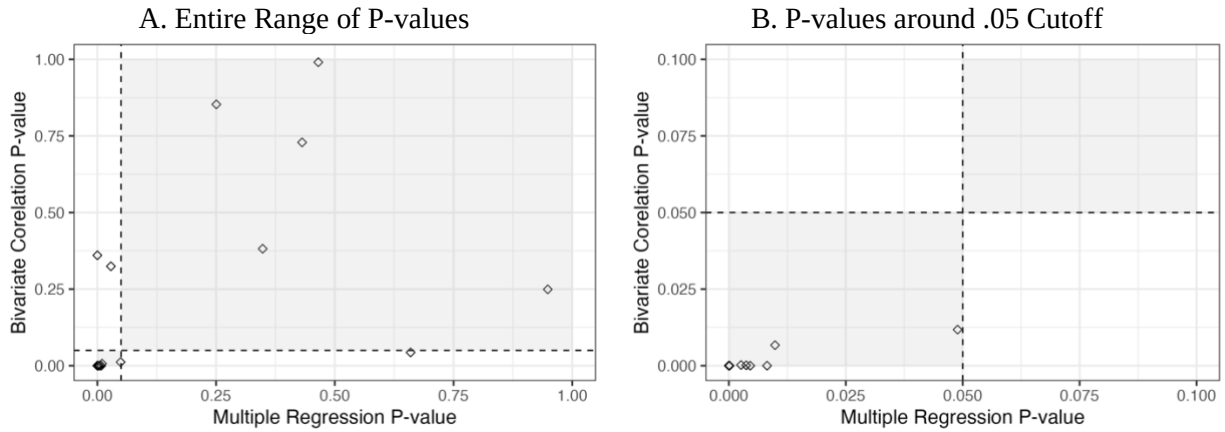
	Perceived accuracy of health misinformation	Adherence to the CDC guide on social distancing	Confidence in COVID vaccine	Science Curiosity	Religiosity
Pro-evidence	–.05	.16***	.12***	.45***	–.02
Anti-evidence	.50***	–.25***	–.51***	–.17***	.32***
Pro-advice	.34***	–.09**	.04	–.01	.26***
Anti-advice	.04	–.11**	–.27***	–.001	–.04

Note. *p < .1; **p < .05; ***p < .01.

Overall, bivariate correlations (Table S21) are consistent with the relationships found from multiple regression (Table 4). The signs of the coefficients are the same between the two approaches. Figure S3 plots p-values of the correlations between the EvA tendencies and criteria variables estimated from each approach, and the shaded area implies that the statistical inference is consistent, with most observations inside of or close to the shaded area. There are three exceptions: statistically significant relationship ($p < .05$) in multiple regressions but not significant in bivariate correlations (Pro-advice & vaccine confidence, Anti-advice & religiosity); one observation lies near the borderline that is statistically not significant in multiple regressions but significant in bivariate correlation (Pro-advice & CDC social distancing). Even for these outlier cases, the sign of correlations (positive or negative) is consistent between the two approaches, and two of these outlier bivariate correlation estimates are within the 95% CI of the multiple regression coefficient estimates.

Figure S3

Comparison of P-values between Multiple Regression Coefficients and Bivariate Correlations



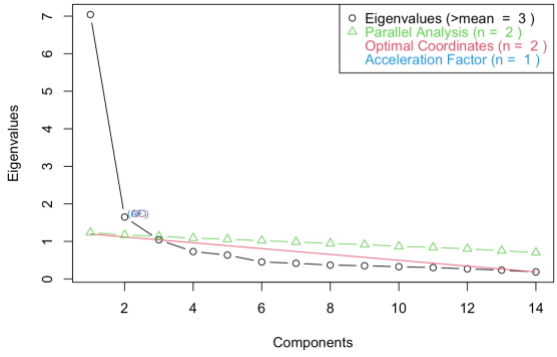
Note. Each observation represents the p-value of the correlation between the EvA tendencies and criteria variables, using multiple regression (Table 4 in the main text) and bivariate correlations (Table S21). Dashed lines are drawn at $p = .05$ for horizontal and vertical axes. Shaded area indicates observations where statistical inference is consistent between the two approaches (either "$p < .05$ for both approaches" or "$p > .05$ for both approaches").

Exploratory Factor Analysis on Adherence to CDC Guidelines on COVID-19

In Study 4, we used fourteen items on CDC guideline adherence on COVID-19 by Graupensperger et al. (2021). Parallel analysis suggested that there exist two underlying factors. As shown in Table S22, the first factor is about distancing behaviors, whereas the second factor pertains to sanitizing behaviors. The necessity of sanitizing behaviors was less uniform across individuals than distancing behaviors. For instance, if individuals abide by CDC guidelines, they were expected to regularly engage with distancing behaviors. However, less frequency of sanitizing behaviors did not necessarily mean less compliance. For instance, individuals who mostly stay indoors had less need to engage with sanitizing behaviors, which did not necessarily indicate less compliance to CDC guidelines. Thus, we used the composite score of ten distancing behavior items as the main measure of adherence to CDC guidelines on COVID-19.

Figure S4

Parallel Analysis of the CDC Guideline Adherence on COVID-19



Exploratory factor analysis revealed that these fourteen items loaded onto two factors, distancing behaviors (ten items: e.g., six-feet distancing, wearing mask) and sanitizing behaviors (three items: e.g., hand washing, disinfecting surfaces); only one item, "getting tested when feeling sick," did not meaningfully load on either (Table S22).

Table S22

EFA on CDC Guideline Adherence on COVID-19 Items

Item	Statement	Factor 1	Factor 2
1	Hand washing with soap and water for 20 seconds when available		0.65
2	Using hand sanitizer in between activities		0.87
3	Wearing a face mask when in indoor public spaces (e.g., shopping)	0.58	
4	Staying 6 feet away from other people you don't live with	0.67	
5	Doing things at home rather than in public, when possible (e.g., work)	0.79	
6	Clean and disinfect frequently touched surfaces (e.g., tables, doorknobs)		0.84
7	Avoiding dining in restaurants by cooking meals at home and using takeout/delivery options	0.85	
8	Avoiding crowded indoor hang-out spots (e.g., bars, pubs, lounges)	0.96	
9	Avoiding large indoor gatherings such as weddings, shows, or parties	0.98	
10	Avoiding indoor social gatherings (e.g., friends' houses)	0.88	
11	Staying home and getting tested when feeling sick		
12	Avoiding contact with at-risk individuals (e.g., older people)	0.52	
13	Avoiding physical contact with others you do not live with (e.g., handshakes, hugs)	0.71	
14	Wearing a face mask while using public transportation (e.g., buses, trains, planes)	0.41	

Note: Factor loadings smaller than .3 are not displayed.

Relationships between the EvA Scale and Additional Criterion Variables

We examined the relationships between the EvA scale and additional variables available in our study. Some of these variables were preregistered as exploratory analyses (e.g., trust in various sources of COVID-related information). The questionnaires designed to measure susceptibility to misinformation (Scherer et al. 2021) included not only social media posts with false health information, but also those with true health information to prevent subjects from making inferences about the research purpose. While our main analysis focused on perceived accuracy of *false* health information (susceptibility to health misinformation), as a post-hoc analysis, we additionally examined the relationships between the EvA traits and other available behaviors regarding health information. The relationships between the EvA traits and COVID-related behaviors and attitudes (e.g., vaccine intake, trust in various sources of COVID-related information) were preregistered as exploratory analysis. Overall, these additional results indicate the four EvA reasoning tendencies identify individual differences in attitudes and decision making on the basis of evidence versus advice.

Table S23

Relationships between the EvA tendencies and Criterion Behaviors/Attitudes: Perceived Accuracy and Decision Influence of Health Information, COVID vaccine intake

EvA Reasoning Tendencies	Perceived accuracy of *true* info	Decision influence of *true* info	Decision influence of *false* info	COVID vaccine intake[1]
Pro-evidence	0.11 (0.04)***	0.14 (0.06)**	0.006 (0.05)	–0.06 (0.27)
Anti-evidence	–0.29 (0.04)***	–0.18 (0.06)***	0.035 (0.04)***	–2.09 (0.21)***
Pro-advice	–0.02 (0.04)	0.23 (0.06)***	0.22 (0.05)***	0.18 (0.25)
Anti-advice	0.08 (0.05)	–0.07 (0.06)	–0.06 (0.05)	–0.11 (0.28)
Constant	0.72 (0.05)***	0.51 (0.06)***	0.04 (0.06)	2.68 (0.31)***
N	529	529	529	529
Adjusted R^2	0.17	0.06	0.24	0.16

Note. Unstandardized regression coefficient with robust standard errors in parentheses. All variables were scaled to range from 0 to 1. *p < .10; **p < .05; ***p < .01. *Perceived accuracy* of true health information was measured as the average of perceived accuracy ratings on the four true headlines. *Decision influence* refers to the degree to which individuals rated the given information would influence their own cancer treatment decisions. Decision influence variables were constructed as the average of four *true* and *four* false social media posts respectively. *COVID vaccine intake* refers to the number of COVID vaccine doses, ranging from 0 to 3.

Table S24

Relationships between the EvA tendencies and Criterion Behaviors/Attitudes: Trust in Sources of COVID-related Information

EvA Reasoning Tendencies	Trust in Centers for Disease Control and Prevention (CDC)	Trust in Food and Drug Administration (FDA)	Trust in religious leaders	Trust in social media
Pro-evidence	0.08 (0.07)	0.08 (0.07)	–0.05 (0.07)	0.004 (0.06)
Anti-evidence	–0.69 (0.06)***	–0.51 (0.06)***	0.32 (0.06)***	0.13 (0.05)***
Pro-advice	0.41 (0.07)***	0.42 (0.06)***	0.44 (0.07)***	0.55 (0.05)***
Anti-advice	–0.29 (0.07)***	–0.37 (0.08)***	–0.25 (0.07)***	–0.29 (0.06)***
Constant	0.89 (0.07)***	0.85 (0.07)***	0.23 (0.07)***	0.20 (0.07)***
N	529	529	529	529
Adjusted R^2	0.32	0.27	0.24	

Note. Unstandardized regression coefficient with robust standard errors in parentheses. All variables were scaled to range from 0 to 1. *p < .10; **p < .05; ***p < .01. *Trust in [CDC/FDA/religious leaders/social media]* refers to the degree of trust in [CDC/FDA/religious leaders/social media] as the source of information about COVID-19 vaccines on a five-point scale ranging from "strongly distrust" to "strongly trust."

[1] The direction and statistical significance of the relationships between the EvA traits and COVID vaccine intake remains the same when we run a logistic regression by using a binary variable that indicates COVID vaccine intake (0 = did not take any, 1 = took at least one COVID vaccine dose).

Survey Instruments

Note: The study materials, raw data, and R code for this study are available at: https://osf.io/qeav5/?view_only=cc20ea01c47b48adbde31f7e1e7ec52d.

Study 1

- **Initial 57 items for the scale development**

Instruction:

> We'd like to ask you questions about your typical preferences when obtaining or processing information. Please indicate how much you disagree or agree with each statement.

Response options:

> Strongly disagree (1) – Disagree (2) – Somewhat disagree (3) – Neither disagree nor agree (4) – Somewhat agree (5) – Agree (6) – Strongly Agree (7)

Note: The order of items was randomized.

Item	Statement
1	I pay close attention to what my religious leader tells me I should do.
2	When I have to vote, I see what my politician says and follow their lead.
3	I generally follow my parents' advice.
4	I assume that when my favorite blogger or social media personality gives advice, they know what they are talking about.
5	I respect law enforcement, like police officers.
6	I often make changes to my diet based on what my friends tell me is more healthy.
7	My behavior is usually dictated by my religious values.
8	I assume my doctors know what they're talking about, so I follow their recommendations.
9	When I think a politician has a confident, assertive personality, I naturally like them and vote for them.
10	People who challenge authority are overconfident.
11	Sometimes you just have to accept the teacher's answer even though you don't understand it.
12	People should always respect authority.
13	People who are in a position of authority have the right to tell others what to do.
14	Children should be allowed to question their parents' authority.
15	I wonder how much my teachers really knew.
16	Even advice from experts should be questioned.
17	I'm the type of person who questions authority.
18	I am doubtful that my teachers really understood what they were teaching me.
19	Hosts of major television news shows do not know enough to be reliable sources of information.
20	Just because people are older or more experienced does not mean their claims are necessarily correct.
21	Government officials often say things that are untrue in their public statements.
22	I assume that people in positions of power are corrupt.

23	People who are telling us how to act don't always have an incentive to tell the truth.
24	From my perspective, people in positions of authority should generally not be trusted.
25	Scientists' research doesn't matter in the real world.
26	When I hear a news story about health, I wonder if there is really good evidence behind the assertion.
27	When someone makes a statement that sounds like a fact, I want to know the evidence behind it.
28	When someone cites a statistic, I want to know where they got it from.
29	I am concerned that news reports are based on people's opinions rather than actual evidence.
30	I pay attention to science news and try to follow the latest findings.
31	I am more likely to avoid a risk when I learn about the statistics rather than personal stories and anecdotes.
32	When I hear a news story reporting research about health, I want to look up the study they are referring to.
33	Before I vote on an issue in my state or city, I try to look up the ballot items so that I vote correctly.
34	I carefully examine research on important issues to make sure it is valid and unbiased.
35	When I hear about new research, I look into who funded it to be sure it is unbiased.
36	I think news reports about science should include more information so that we can evaluate the strength of the evidence.
37	When my doctor tells me about a new treatment, I like to find out about any research on the treatment.
38	Science is very important for the country's development.
39	Money spent on science is well worth spending.
40	In general, you should consider whether the information in your textbook is accurate.
41	When debating an important issue, I try to fact-check things that people state as statistics.
42	It is usually wise to seek out evidence and research before making decisions
43	I believe that things that are natural are always better for you.
44	I am wary of medical procedures that interfere with my body's natural processes.
45	I am hesitant to take prescription medicines because they seem like chemicals I am putting into my body.
46	When scientists change their minds, I stop trusting their research on what we are supposed to eat to be healthy.
47	When new evidence reverses a previous scientific theory, I just stop paying attention to it and make my own decisions.
48	People make too much of scientific studies in the news when I know that the research is biased anyway.
49	People can talk about data, but I think that my intuitions are a better guide for my decisions.
50	Even if scientific studies are done carefully and transparently, I still don't really believe them.
51	I am not interested in looking into the details when I hear the results of a new study.
52	I think scientific data is too hard to understand, so I generally ignore it.
53	When it comes to controversial issues in society, I don't think "the data" can tell us much.
54	I avoid GMOs and pesticides, no matter what the evidence says.
55	I rarely check the nutrition facts or ingredient list on food labels.
56	When choosing between products, I don't spend much time comparing the specifications.
57	Scientific evidence is overrated; there are often better ways to understand the world.

Study 2

- **The 16-item Evidence versus Advice (EvA) scale**

Instruction:

> We'd like to ask you questions about your typical preferences when obtaining or processing information. Please indicate how much you disagree or agree with each statement.

Response options:

> Strongly disagree (1) – Disagree (2) – Somewhat disagree (3) – Neither disagree nor agree (4) – Somewhat agree (5) – Agree (6) – Strongly Agree (7)

Note: The order of items was randomized.

Pro-evidence	
1	When I hear a news story reporting research about health, I want to look up the study they are referring to.
2	I carefully examine research on important issues to make sure it is valid and unbiased.
3	When debating an important issue, I try to fact-check things that people state as statistics
4	When someone makes a statement that sounds like a fact, I want to know the evidence behind it.
Anti-evidence	
5	Scientific evidence is overrated; there are often better ways to understand the world.
6	Even if scientific studies are done carefully and transparently, I still don't really believe them.
7	People make too much of scientific studies in the news when I know that the research is biased anyway.
8	When new evidence reverses a previous scientific theory, I just stop paying attention to it and make my own decisions.
Pro-advice	
9	When I have to vote, I see what my politician says and follow their lead.
10	I often make changes to my diet based on what my friends tell me is more healthy.
11	When I think a politician has a confident, assertive personality, I naturally like them and vote for them.
12	I assume that when my favorite blogger or social media personality gives advice, they know what they are talking about.
Anti-advice	
13	I am concerned that news reports are based on people's opinions rather than actual evidence.
14	Government officials often say things that are untrue in their public statements.
15	Hosts of major television news shows do not know enough to be reliable sources of information.
16	People who are telling us how to act don't always have an incentive to tell the truth.

Study 3

Note: For each scale, the order of items was randomized.

- **Need for Cognition Scale** (NCS; Coelho et al., 2018)

Response options:

Strongly disagree (1) – Disagree (2) – Somewhat disagree (3) – Neither disagree nor agree (4) – Somewhat agree (5) – Agree (6) – Strongly Agree (7)

Statements:

- I would prefer complex to simple problems.
- I like to have the responsibility of handling a situation that requires a lot of thinking.
- Thinking is not my idea of fun. (R)
- I would rather do something that requires little thought than something that is sure to challenge my thinking abilities. (R)
- I really enjoy a task that involves coming up with new solutions to problems.
- I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought.

- **Distrust in Science** (Nadelson et al., 2020; 12 distrust items)

Response options:

Strongly disagree (1) – Disagree (2) – Neutral (3) – Agree (4) – Strongly Agree (5)

Statements:

- When scientists change their mind about a scientific idea it diminishes my trust in their work.
- Scientists ignore evidence that contradicts their work.
- Scientific theories are weak explanations.
- Scientists intentionally keep their work secret.
- Scientists don't value the ideas of others.
- Scientists don't care if laypersons understand their work.
- When scientists form a hypothesis they are just guessing.
- We cannot trust scientists because they are biased in their perspectives.
- Scientist will protect each other even when they are wrong.
- We cannot trust scientists to consider ideas that contradict their own.
- Today's scientists will sacrifice the well being of others to advance their research.
- We cannot trust science because it moves too slowly.

- **Right-Wing Authoritarianism** (RWA; Bizumic & Duckitt, 2018)

Response options:

Strongly disagree (1) – Disagree (2) – Somewhat disagree (3) – Neither disagree nor agree (4) – Somewhat agree (5) – Agree (6) – Strongly Agree (7)

- **Defiance to Authority (Reverse-code of the "Authoritarianism Submission")**
  - What our country needs most is discipline, with everyone following our leaders in unity.
  - It's great that many young people today are prepared to defy authority (R).

- **Respect for Convention ("Conventionalism")**

- God's laws about abortion, pornography, and marriage must be strictly followed before it is too late.
- There is nothing wrong with premarital sexual intercourse (R).

- **Numeracy** (Weller et al., 2013)

Instruction:

Please answer the following questions by entering your answer into the box using numbers only.

Questions:

- Imagine that we roll a fair, six-sided die 1000 times. Out of 1000 rolls, how many times do you think the die would come up as an even number? ____
- In the BIG BUCKS LOTTERY, the chances of winning a $10.00 prize are 1%. What is your best guess about how many people would win a $10.00 prize if 1000 people each buy a single ticket from BIG BUCKS? ____
- In the ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1000. What percent of tickets of ACME PUBLISHING SWEEPSTAKES win a car? ____ %
- If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 1000 people? ____
- If the chance of getting a disease is 20 out of 100, this would be the same as having a ____% chance of getting the disease.
- A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost? ____ cents
- If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? ____ minutes
- In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? ____ days

- **Pessimism** (Scheier et al., 2012)

Response options:

Disagree a lot (1) – Disagree a little (2) – Neither agree nor disagree (3) – Agree a little (4) – Agree a lot (5)

Statements:

- If something can go wrong for me, it will.
- I hardly ever expect things to go my way.
- I rarely count on good things happening to me.

- **Dispositional Trust/Distrust** (Bianchi & Brockner, 2012)

Do you think most people would try to take advantage of you if they got a chance or would they try to be fair?

Take advantage (1) - Depends (2) - Fair (3)

Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?

Cannot trust (1) - Depends (2) - Can trust (3)

Would you say that most of the time people try to be helpful, or that they are mostly just looking out for themselves?

Look out for themselves (1) - Depends (2) - Helpful (3)

- **Social Desirability** (Fischer & Fick, 1993)

Note: We discuss concerns about this scale in the section "Measurement Issue with Social Desirability" (Study 3) in this document. Due to the concerns, we use an alternative social desirability measure in Study 4.

Instruction:

Please answer the following questions by choosing whether or not each statement describes you, using true or false.

Response options:

True (1) – False (2)

Statements:

- I'm always willing to admit it when I make a mistake.
- I always try to practice what I preach.
- I never resent being asked to return a favor.
- I have never been irked when people expressed ideas very different from my own.
- I have never deliberately said something that hurt someone's feelings.
- I like to gossip at times. (R)
- There have been occasions when I took advantage of someone. (R)
- I sometimes try to get even rather than forgive and forget. (R)
- At times I have really insisted on having things my own way. (R)
- There have been occasions when I felt like smashing things. (R)

Study 4

- **Susceptibility to Health Misinformation** (Scherer et al., 2021)

Instruction:

[Screen 1]
Now, you will view pictures and information that have been shared publicly on social media.
Your job is to evaluate the accuracy of each one. We are interested in your personal opinion.

Some social media posts may contain multiple claims. In these cases, tell us what you think of it overall.
Later, we will ask you some questions about yourself, and then you will be done.
Click the arrow button for a few more instructions.

[Screen 2]
You will be judging social media posts about a health-related topics: cancer treatments

[Screen 3]
Now you are ready to begin. The first thing you see will be a social media post. Please read it and then answer the two questions below it.
You will rate a total of 8 social media posts.
Click the arrow button to start rating the social media posts.

Questionnaires for each social media post:

- To the best of your knowledge, how accurate is the information in this social media post?
    - Completely false (1) - Mostly false (2) - Mostly true (3) - Completely true (4)

- If you were diagnosed with cancer, would this information influence your decision about your treatment?
    - No, definitely not (1) - Probably not (2) - Probably yes (3) - Yes, definitely (4)

Social media posts:[2]

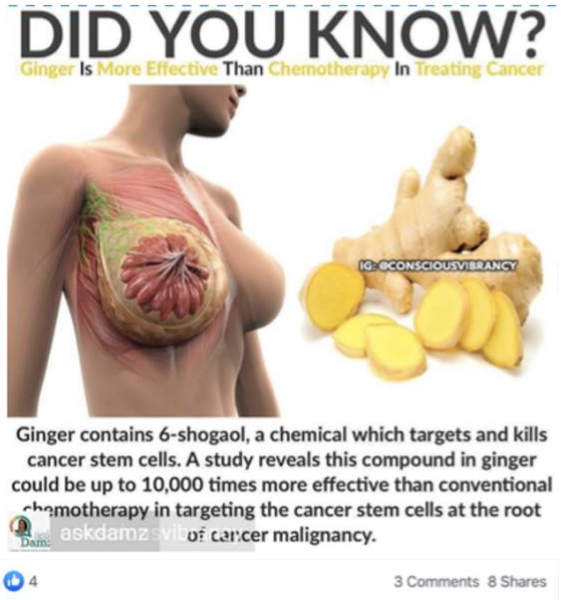
- Social media posts with FALSE information:

1

2

[2] These experimental stimuli were adopted from supplemental materials (cancer treatments) for Scherer et al. (2021).

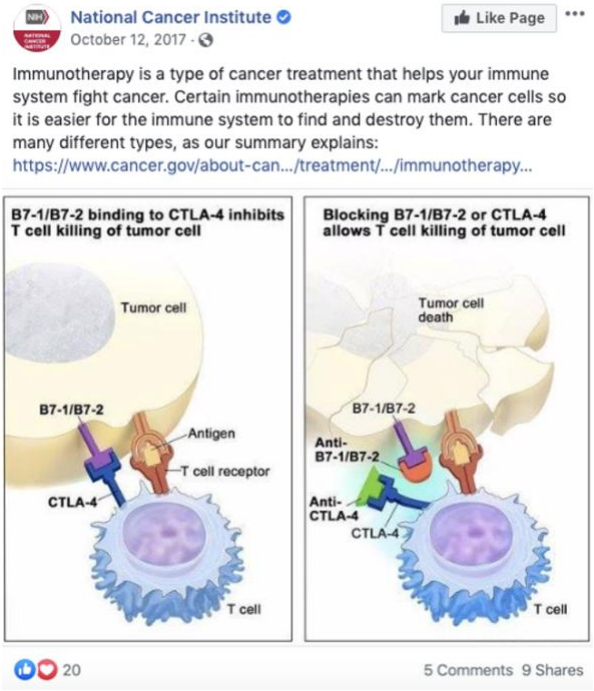
3



4

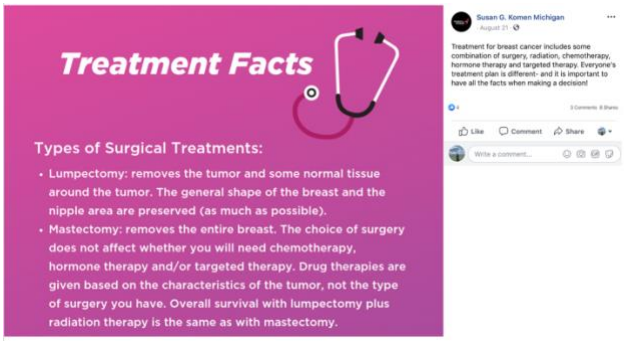
- Social media posts with TRUE information:

1

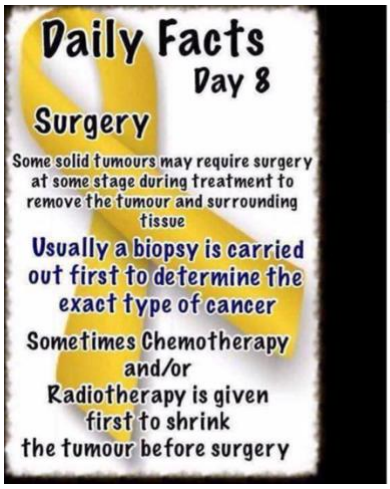


2

3



4



- **Adherence to CDC Guidelines on COVID-19** (Graupensperger et al., 2021)

Instruction:

Please rate the degree to which you engaged in each activity during the past month.

Response options:

Never (1) – Rarely (2) – Sometimes (3) – Often (4) – All the time (5)

Statements:

- Hand washing with soap and water for 20 seconds when available
- Using hand sanitizer in between activities
- **Wearing a face mask when in indoor public spaces (e.g., shopping)**
- **Staying 6 feet away from other people you don't live with**
- **Doing things at home rather than in public, when possible (e.g., work)**
- Clean and disinfect frequently touched surfaces (e.g., tables, doorknobs)
- **Avoiding dining in restaurants by cooking meals at home and using takeout/delivery options**
- **Avoiding crowded indoor hang-out spots (e.g., bars, pubs, lounges)**
- **Avoiding large indoor gatherings such as weddings, shows, or parties**
- **Avoiding indoor social gatherings (e.g., friends' houses)**
- Staying home and getting tested when feeling sick
- **Avoiding contact with at-risk individuals (e.g., older people)**
- **Avoiding physical contact with others you do not live with (e.g., handshakes, hugs)**
- **Wearing a face mask while using public transportation (e.g., buses, trains, planes)**

Note: Bolded items were used to construct a measure of distancing behaviors (explained in Table S21).

- **COVID-19 Related Attitudes** (CDC, 2021)

Confidence in COVID-19 vaccine:

- How likely are you to recommend getting the COVID-19 vaccine to others?
    - Not at all likely (1) - A little likely (2) - Somewhat likely (3) - Very likely (4) - Extremely likely (5)

- How safe do you think a COVID-19 vaccine is for you?
    - Not at all safe (1) - A little safe (2) - Moderately safe (3) - Very safe (4) - Extremely safe (5)

- How much do you distrust or trust the public health agencies that recommend you get a COVID-19 vaccine?
    - Strongly distrust (1) - Distrust (2) - Neither distrust nor trust (3) - Trust (4) - Strongly trust (5)

- How much confidence do you have that the research and development process have produced COVID-19 vaccines in the U.S. that are safe and effective?
    - None at all (1) - A little (2) - A moderate amount (3) - A lot (4) - A great deal (5)

Trust in sources of COVID-19 information:

Instruction:

How much do you distrust or trust the following as sources of information about COVID-19 vaccines?

Note: These response options and items were presented as a matrix.

Response options:

Strongly distrust (1) - Distrust (2) - Neither distrust nor trust (3) - Trust (4) - Strongly trust (5)

Items:

- Centers for Disease Control and Prevention (CDC)
- Food and Drug Administration (FDA)
- Religious leader(s)
- Social media (e.g., Facebook, Twitter, Instagram, WhatsApp, LinkedIn, or TikTok)

Intake of COVID-19 vaccine:

Q1 Have you received a COVID-19 vaccine?

Yes (1) - No (2)

[Display if Q1==1] Q2 Did you receive a vaccine product that requires only one dose or two doses?

One-dose product (e.g., Johnson & Johnson) (1)
Two-dose product (e.g., Pfizer, Moderna) (2)
Mix and match (e.g., J&J and Pfizer, Moderna and Pfizer, etc.) (3)

[Display if Q1==1] Q3 How many doses of COVID vaccines have you gotten so far?
One dose (1) - Two doses (2) - Three doses (3)

- **Science Curiosity** (Kahan et al. 2017; Motta et al., 2021)

Note: Following Motta et al. (2021) and our preregistration, science curiosity was measured as a composite score of bolded items.

Q1 There are a lot of issues in the news and it is hard to keep up with every area. We will list some topics that are covered in the media.

Please indicate **how closely you follow news relating to each topic** either in the newspaper, on television, on the radio, or on the Internet.

	Not at all (1)	A little but not closely (2)	Closely but not very closely (3)	Very closely (4)
Government or politics (1)				
Religion (2)				
Scientific research or discoveries (3)				
New technologies (4)				
Entertainment or celebrities (5)				

Q2 We'd also like to know whether you read books in your spare time. We will list some book topics. Please indicate **whether you have read a book on that topic <u>in the previous year</u>**.

	Did not read any books on the topic in the previous year (1)	Have read **1 book** on the topic in the previous year (2)	Have read **2 books** on the topic in the previous year (3)	Have read **more than 3 books** on the topic in the previous year (4)
Science fiction (1)				
Mystery novel (2)				
Government or politics (3)				
Religion (other than holy script text) (4)				
Scientific research or discoveries (5)				

Q3 We are also interested in knowing about the sorts of topics you discuss with family members, friends, or co-workers. We will list some conversation topics. Please indicate **how often you discuss these topics with either friends, family members, or co-workers.**

	Never (1)	Rarely (2)	More than rarely but not often (3)	Often (4)
Government or politics (1)				
Religion (2)				
Scientific research or discoveries (3)				
New technologies (4)				
Entertainment or celebrities (5)				

Q4 We will now list some topics that some people are interested in, and some people are not interested in. For each topic, please indicate **how interested you are in that topic.**

	Not at all interested (1)	Slightly interested (2)	Moderately interested (3)	Very interested (4)
Government or politics (1)				
Religion (2)				
Scientific research or discoveries (3)				
New technologies (4)				
Entertainment or celebrities (5)				

- **Religiosity** (Rohrboaugh & Jessor, 1975)

  - How often have you attended religious services during the past year?
    - Never (1)
    - A few times a year (2)
    - Once or twice a month (3)
    - Almost every week (4)
    - Every week (5)

  - How often do you pray or practice religious meditation?
    - Never (1)
    - Rarely (2)
    - Occasionally (3)
    - Fairly often (4)
    - Very often (5)

  - When you have a serious personal problem, how often do you take religious advice or teaching into consideration?
    - Never (1)
    - Rarely (2)
    - Occasionally (3)
    - Fairly often (4)

- Very often (5)

- How much influence would you say that religion has on the way that you choose to act and the way that you choose to spend your time each day?
  - No influence (1)
  - A small influence (2)
  - Some influence (3)
  - A fair amount of influence (4)
  - A large influence (5)

- Which of the following statements comes closest to your belief about God? (pick one)
  - I don't believe in a personal God or in a higher power (1)
  - I don't know if there is a personal God or a higher power of some kind, and I don't know if I ever will. (2)
  - I don't know if there is a personal God, but I do believe in a higher power of some kind. (3)
  - Although I sometimes question God's existence, I do believe in God and believe God knows of me as a person. (4)
  - I am sure that God really exists and is active in my life. (5)

- **Social Desirability** (Hart et al., 2015; Impression Management)

Response options:

Strongly disagree (1) – Disagree (2) – Slightly disagree (3) – Neither disagree nor agree (4) – Slightly agree (5) – Agree (6) – Strongly Agree (7)

Statements:

- I sometimes tell lies if I have to. (R)
- I never cover up my mistakes.
- There have been occasions when I have taken advantage of someone. (R)
- I sometimes try to get even rather than forgive and forget. (R)
- I have said something bad about a friend behind his or her back. (R)
- When I hear people talking privately, I avoid listening.
- I never take things that don't belong to me.
- I don't gossip about other people's business.

References

Adcock, R., & Collier, D. (2001). Measurement validity: A shared standard for qualitative and quantitative research. *The American Political Science Review*, *95*(3), 529–546.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238–246.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105.

Charalambides, N. (2021). We recently went viral on TikTok—Here's what we learned. *Prolific*. https://blog.prolific.co/we-recently-went-viral-on-tiktok-heres-what-we-learned/

Clark, L. A., & Watson, D. (2016). *Constructing validity: Basic issues in objective scale development* (p. 203). American Psychological Association.

DeVellis, R. F. (2006). Classical Test Theory. *Medical Care*, *44*(11), S50–S59.

Flora, D. B. (2020). Your coefficient alpha is probably wrong, but which coefficient omega is right? A tutorial on using R to obtain better reliability estimates. *Advances in Methods and Practices in Psychological Science*, *3*(4), 484–501.

Gastañaduy, P. A., Budd, J., Fisher, N., Redd, S. B., Fletcher, J., Miller, J., McFadden, D. J., Rota, J., Rota, P. A., Hickman, C., Fowler, B., Tatham, L., Wallace, G. S., de Fijter, S., Parker Fiebelkorn, A., & DiOrio, M. (2016). A measles outbreak in an underimmunized Amish community in Ohio. *New England Journal of Medicine*, *375*(14), 1343–1354.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Gliem, J. A., & Gliem, R. R. (2003). *Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for likert-type scales*. Presented at the Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education. https://hdl.handle.net/1805/344

Helmes, E., & Holden, R. R. (2003). The construct of social desirability: One or two dimensions? *Personality and Individual Differences*, *34*(6), 1015–1023.

McDonald, R. P., & Ho, M.-H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, *7*(1), 64–82.

McNeish, D., & Wolf, M. G. (2023). Dynamic fit index cutoffs for confirmatory factor analysis models. *Psychological Methods*, *28*(1), 61–88.

Peterson, R. A. (1994). A meta-analysis of Cronbach's coefficient alpha. *Journal of Consumer Research*, *21*(2), 381–391.

Price, J. L., & Mueller, C. W. (1986). *Handbook of organizational measurement*. Pitman: Longman.

Ramanaiah, N. V., & Martin, H. J. (1980). On the two-dimensional bature of the Marlowe-Crowne social desirability scale. *Journal of Personality Assessment*, *44*(5), 507–514.

Raubenheimer, J. (2004). An item selection procedure to maximize scale reliability and validity. *SA Journal of Industrial Psychology*, *30*(4), 59–64.

Raykov, T. (2008). Alpha if item deleted: A note on loss of criterion validity in scale development if maximizing coefficient alpha. *British Journal of Mathematical and Statistical Psychology*, *61*(2), 275–285.

Reynolds, W. M. (1982). Development of reliable and valid short forms of the Marlowe-Crowne Social Desirability Scale. *Journal of Clinical Psychology*, *38*(1), 119–125.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, *8*(4), 350–353.

Taber, K. S. (2018). The use of Cronbach's Alpha when developing and reporting research instruments in science education. *Research in Science Education*, *48*(6), 1273–1296.