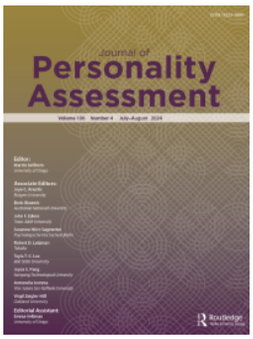

Routledge
Taylor & Francis Group

Journal of Personality Assessment

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/hjpa20

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

Hwayong Shin, Priti Shah & Stephanie D. Preston

To cite this article: Hwayong Shin, Priti Shah & Stephanie D. Preston (25 Jan 2024): The Reasoning through Evidence versus Advice (EvA) Scale: Scale Development and Validation, Journal of Personality Assessment, DOI: 10.1080/00223891.2023.2297266

To link to this article: https://doi.org/10.1080/00223891.2023.2297266

JOURNAL OF PERSONALITY ASSESSMENT
https://doi.org/10.1080/00223891.2023.2297266

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, Ann Arbor, Michigan; [2]Department of Psychology, University of Michigan, Ann Arbor, Michigan

ABSTRACT

Our well-being can improve when people heed evidence rather than simply follow familiar or charismatic advisors who neglect evidence. We developed the Reasoning through Evidence versus Advice (EvA) scale to measure individual differences in reasoning through evidence like science and statistics versus following advisors such as politicians and celebrities. No existing scales directly measure these tendencies; moreover, it was theoretically unknown whether they reflect a single dimension (from evidence- to advice-based) or distinct tendencies to value or distrust each. Our scale validation process included qualitative interviews and four studies that involved 1583 respondents (753 college graduates, 830 non-college graduates) in which we conducted exploratory and confirmatory factor analyses and tests of convergent validity, discriminant validity, and measurement invariance by gender and education. This process yielded a 16-item EvA scale with four dimensions: Pro-evidence, Anti-evidence, Pro-advice, and Anti-advice. In assessing criterion validity, these tendencies identified individual differences in important, real-world attitudes and behaviors, including susceptibility to health misinformation, adherence to CDC guidelines on social distancing, confidence in the COVID vaccine, science curiosity, and religiosity. The EvA scale extends our understanding of individual differences in reasoning tendencies that shape critical attitudes, decisions, and behaviors and can help promote informed decisions.

ARTICLE HISTORY
Received 28 December 2022
Revised 15 November 2023
Accepted 28 November 2023

Before the introduction of the measles vaccine in 1963, an estimated 30 million people worldwide were infected with measles and more than 2 million died from measles each year; an estimated 50,000 hospitalizations occurred annually in the United States alone (Rota et al., 2016). Subsequently, mass immunization through the measles vaccine dramatically reduced the number of cases, preventing an estimated 17.1 million deaths between 2000 and 2014 worldwide (Perry et al., 2015). Despite the well-documented evidence in support of the measles vaccine, some still refuse the vaccine for nonscientific reasons, such as a general distrust in science or because they follow anecdotal claims (e.g., celebrities' claims that vaccines cause autism) or religious doctrines (e.g., porcine components of the measles vaccine are prohibited in some religions) (Browne et al., 2015; Martinez-Berman et al., 2020; Wombwell et al., 2015). The resulting avoidance of vaccination has led to measles outbreaks amongst several intentionally unvaccinated communities, threatening the health of vaccinated individuals and those who cannot receive the vaccine and imposing a burden upon society (Nelson, 2019; Tanne, 2019).

Public health crises, and society overall, benefit when individuals actively seek and base decisions upon "evidence" over "advice"—two major ways of knowing. Evidence refers to aggregated data that describe observed relationships, acquired through the scientific method and statistics (Brown et al., 2010; McNeill & Martin, 2011). Evidence-based reasoning involves drawing conclusions from the best available evidence, obtained through transparent and reproducible procedures, which reduces misleading biases (Gambrill, 1999; Jamieson et al., 2019). In contrast, advice-based reasoning can be characterized by arriving at conclusions through opinions, anecdotes, or declarations of trusted others, such as family, friends, politicians, religious leaders, and celebrities, which may lead to the uncritical acceptance of opinions that counter evidence (Gambrill, 1999; Guzelian & Guzelian, 2004).

On controversial issues, people often reason through sources of information that correspond to evidence or advice: evidential justifications (i.e., empirical data, testable effects in research) or deferential justifications (i.e., source of information that individuals defer to) (Lobato & Zimmerman, 2019). Evidence-based practices and decisions have been recommended in multiple fields, including medicine, management, and public policy to improve outcomes (Akobeng, 2005; Pfeffer & Sutton, 2006; Sanderson, 2002). Evidence is the preferred basis for knowing and deciding, yet some still believe scientifically unsupported claims, even when made aware of established scientific evidence (Kahan, 2015; Rynes et al.,

CONTACT Hwayong Shin hwayong@umich.edu Department of Political Science, University of Michigan, 5700 Haven Hall, 505 S State St, Ann Arbor, MI 48109, USA.
Supplemental data for this article can be accessed online at https://doi.org/10.1080/00223891.2023.2297266.
Hwayong Shin is now located at the Rockefeller Center for Public Policy and the Social Sciences, Dartmouth College, 2 Webster Ave, Hanover, NH 03755.

2018), after hearing from non-expert advisors, including politicians, celebrities, religious leaders, and friends (Druckman et al., 2013; Martinez-Berman et al., 2020; Waldinger, 2004).

Understanding people's reliance upon evidence and advice is important in the current polarized and fragmented information environment, where individuals can selectively accept or avoid either source of information. Partisan polarization in the US, where Democrats and Republicans are increasingly at odds with and hostile to one another (Abramowitz & Webster, 2016; Iyengar et al., 2019), can motivate individuals to heed advice from trusted others over evidence. For instance, people are vulnerable to falsehoods or conspiracy theories from well-known politicians and can be skeptical of scientific findings resisted by their party's leaders (Bolsen & Druckman, 2018; Miller et al., 2016). Furthermore, people are exposed to both credible information from experts and falsehoods from unreliable sources, making it more important than ever to seek and employ objective evidence to inform decisions and prevent false beliefs (Bronstein et al., 2019; Vosoughi et al., 2018).

It is imperative that we understand people's general tendencies to value or devalue evidence and advice to limit harmful misperceptions and ill-informed decisions. As a first step toward the systematic study of these reasoning tendencies, we developed the Reasoning through Evidence versus Advice (EvA) scale, to measure individual differences in the tendency to seek versus suspect evidence and to rely upon or resist advisors' opinions when making decisions.

Measuring reasoning through evidence and advice

There are no existing scales that measure tendencies toward reasoning through evidence and advice. There are related scales such as the Need for Cognition (Cacioppo et al., 1984), which measures the degree that people enjoy deliberation and effortful thinking. There are also scales that measure people's belief and trust in science and scientists (Farias et al., 2013; Nadelson et al., 2014). Other related scales measure people's belief in hierarchical order, such as Right-Wing Authoritarianism (Altemeyer, 1998) and general trust in others, such as Dispositional Trust (Bianchi & Brockner, 2012). While these constructs overlap with our interest, none directly address how people seek, employ, or avoid evidence versus advice as sources of information when making decisions. For example, Need for Cognition captures one's willingness to exert cognitive reasoning efforts, but not how they approach or avoid sources of information (e.g., scientific research, personal advice). Existing scales on belief or trust in science measure support for science at large or scientists per se, without measuring a potentially distinct tendency to seek or discredit scientific evidence. None of the existing scales measure one's tendency to follow the advice of trusted others when forming beliefs or making decisions. Right-Wing Authoritarianism measures support for authorities and a hierarchical social order—including the government, the elderly, laws, and God, without considering other prominent advisors like friends, celebrities, bloggers, and politicians that exert considerable influence. Dispositional Trust addresses trust or suspicion of others generally, without specifically addressing science or personal advice. Neither Right-Wing Authoritarianism nor Dispositional Trust is directly contrasted with trust in evidence. Thus, existing scales address aspects of the problem we seek to solve, but none directly measure people's tendency to rely upon evidence versus advice for decisions.

This research is also theoretically important because it is unknown whether support for evidence and advice relies upon common or distinct constructs. There could be one underlying tendency to base decisions upon evidence on one end of the spectrum to advice on the other. Alternatively, people's reliance upon evidence or advice could depend upon distinct dimensions, with each including seeking versus distrusting those sources. For example, people can vary on how much they seek or suspect evidence. Individuals who value evidence may also accept the advice of trusted individuals or institutions. Because people's reasoning may not involve a simple tradeoff between evidence versus advice, we considered four potentially distinct dimensions: tendencies to seek versus distrust evidence and to rely upon versus suspect advice. To examine the underlying conceptual structure, we included items representing each of these four possible constructs in developing the scale.

We developed a scale that measures individuals' tendencies to reason through Evidence versus Advice (EvA) in a multi-step process: (1) developing an initial set of items refined through qualitative, in-person interviews; (2) performing exploratory factor analysis to identify underlying dimensions; (3) performing confirmatory factor analysis and establishing measurement invariance across demographic subgroups; (4) assessing convergent and discriminant validity by comparing EvA tendencies to related existing scales; (5) assessing criterion validity by examining behaviors or attitudes related to EvA tendencies.

Study 1: Exploratory factor analysis and initial scale characterization

Study 1 was designed to explore the underlying dimensions of reasoning tendencies through evidence and advice from our initial 57 items. In this study and the following ones, we recruited a similar number of participants with and without a college degree, because of the widespread assumption that education level is positively correlated with scientific reasoning but negatively correlated with nonscientific reasoning, such as conspiratorial thinking (Huber & Kuncel, 2016; van Prooijen, 2017).

Pilot study

We first generated a pool of 57 items to measure individuals' tendencies to reason from evidence (e.g., scientific research, statistics) or advice (e.g., parents, friends, politicians, celebrities, religious leaders), with a similar number of items for each dimension (17 Pro-evidence, 12 Anti-evidence, 13 Pro-advice, and 12 Anti-advice; Table S1). Some items were adapted from relevant scales that addressed our concepts of interest: Schommer Epistemological Questionnaire (5 items; Schommer, 1998), Epistemic Beliefs Inventory (1 item; Schraw et al., 2002), Updated Dogmatism Scale (4 items,

Shearman & Levine, 2006), and Attitudes Toward Science Scale (2 items; Francis & Greer, 1999). The remaining items were generated by our study team to create a range of statements assessing people's reasoning through evidence versus advice in a variety of relevant domains. From in-person interviews (5 undergraduate students, 4 adults without college education), four items were clarified (Table S2) but none were removed. Details about the procedures and findings from the pilot study are available in supplementary materials. In this study and the following studies, written informed consent was obtained from participants after they read a consent form that described the study instrument and ensured that their responses would be kept anonymous and the study involved minimal risks. All studies were reviewed and approved by the University of Michigan's Institutional Review Board.

Methods and materials

Following the guidelines on the minimum ratio of participants to items (5:1 or 10:1) for exploratory factor analysis (Gorsuch, 1983; Worthington & Whittaker, 2006), we recruited 579 participants via CloudResearch (Litman et al., 2017). For data quality, we followed the 95% approval rating criteria, to screen out inattentive respondents and ensure data quality (Peer et al., 2014). We excluded 28 who failed at least one of four attention checks, which can identify low-quality responses (Berinsky et al., 2019). Two respondents who did not complete the scale were excluded (Newton et al., 2021). 549 respondents (college: 243; non-college: 306) were retained for analysis (demographics in Table S3), leaving an acceptable participant-item ratio of 9.6:1. Participants responded to the 57 items for the EvA scale (seven-point scale; "strongly disagree" to "strongly agree"; order randomized) before demographic questions (e.g., gender, education). Question wordings for all studies are provided in supplementary materials.

Results

Exploratory Factor Analysis

We conducted exploratory factor analysis (EFA) with maximum likelihood estimation to examine the underlying structure of items (using 'psych' package in R; Baker et al., 2010; Fabrigar et al., 1999). Our Kaiser-Meyer-Olkin (KMO) was .91, which exceeded the recommended values of .60 and higher for an adequate sample size for EFA (Beavers et al. 2013; Tabachnick & Fidell, 2001). Because EvA reasoning tendencies are best characterized as correlated with each other rather than orthogonal, we used oblique (promax) rotation (Brown, 2015; Fabrigar et al., 1999).

The scree plot and parallel analysis suggested six factors (Figure S1; Cattell, 1966; Hayton et al., 2004). The EFA indicated six factors explaining 27%, 24%, 15%, 14%, 12%, and 9% of the variance, respectively. Following recommended item deletion criteria (Baker et al., 2010; Haws et al., 2012; Worthington & Whittaker, 2006), eight items were dropped due to cross-loading (>.30), ten were dropped due to weak factor loading (<.40), and none were dropped due to low communalities (<.40) (full EFA results, factor loadings, kurtosis, skewness, communalities, and item deletion criteria are in supplementary materials, Table S4).

The last step of EFA was to shorten the scale and retain a similar number of items per factor (Baker et al., 2010) to increase scale efficiency (e.g., respondent fatigue) while retaining internal consistency (Worthington & Whittaker, 2006). Empirical and substantive rationales were applied to select a sensible set (Worthington & Whittaker, 2006). To balance factors, we reduced each dimension to four items, deleting those with the following properties: low loadings, high cross-loadings, low contribution to internal consistency, and low conceptual consistency with other items (Brown, 2015; Worthington & Whittaker, 2006) (see Table S5). The retained four items per six dimensions are provided in Table S6.

All six factors could potentially be retained, but conceptual interpretability and theoretical relevance are also important for factor retention decisions (Worthington & Whittaker, 2006). We reasoned that it was sensible to drop Factor 6 (Anti-evidence 2), because it was specific to the aversion to medicine and chemicals. Among the two factors related to Pro-advice, we removed Factor 3 and kept Factor 5, because some Factor 3 items measured Anti-advice tendencies and most originated from existing scales, whereas Factor 5 comprised novel items, contributing more as an original scale. Factor 5 items were also more relevant to our goal of assessing individuals' tendency to rely upon advisors that they like and follow, including politicians, celebrities, and friends. Conversely, Factor 3 items comprised only hierarchical or traditional authorities (e.g., law, government, God) that are less relevant to our broader concern with recent trends in information sources. The final 16-item scale (items in Table 1) was efficient and balanced, with four factors, each containing four items.

Confirmatory Factor Analysis

Following the convention to employ model fit statistics to compare alternative models after EFA (Baker et al. 2010; Cassidy et al., 2005; Svedholm-Häkkinen & Lindeman, 2017), we assessed the soundness of our factor selection decision compared to alternative models through confirmatory factor analysis (CFA). For CFA, the variance–covariance matrices were analyzed using latent variable software programs and maximum-likelihood minimization functions ('lavaan' package in R, Rosseel, 2012). We evaluated model fit following recommended criteria: RMSEA and SRMR ≤ .08, CFI and TLI ≥ .90 (Bentler, 1990; Byrne, 1994; Fabrigar et al. 1999; McDonald & Ho, 2002). The proposed four-dimensional model had a good fit, $\chi^2(df)$ = 228.89 (98), RMSEA = .049, SRMR = .046, CFI = .954, TLI = .943 (Table 1; Table S7). We additionally used simulation-based dynamic fit index (DFI) cutoffs to assess model fits ('dynamic' package in R, McNeish & Wolf, 2023), which indicated the proposed model had acceptable model fits at different levels of misspecification (Table S8, Level 1: RMSEA ≤ .050, SRMR ≤ .057, CFI ≥ .956; Level 2: RMSEA ≤ .065, SRMR ≤ .077, CFI ≥ .933). Item-total correlations

Table 1. Confirmatory factor analysis and item-total correlations.

	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
I carefully examine research on important issues to make sure it is valid and unbiased.	.76	.68
When debating an important issue, I try to fact-check things that people state as statistics	.72	.61
When someone makes a statement that sounds like a fact, I want to know the evidence behind it.	.78	.67
Anti-evidence		
Scientific evidence is overrated; there are often better ways to understand the world.	.77	.70
Even if scientific studies are done carefully and transparently, I still don't really believe them.	.81	.73
People make too much of scientific studies in the news when I know that the research is biased anyway.	.78	.68
When new evidence reverses a previous scientific theory, I just stop paying attention to it and make my own decisions.	.72	.64
Pro-advice		
When I have to vote, I see what my politician says and follow their lead.	.65	.53
I often make changes to my diet based on what my friends tell me is more healthy.	.53	.42
When I think a politician has a confident, assertive personality, I naturally like them and vote for them.	.60	.45
I assume that when my favorite blogger or social media personality gives advice, they know what they are talking about.	.59	.44
Anti-advice		
I am concerned that news reports are based on people's opinions rather than actual evidence.	.72	.59
Government officials often say things that are untrue in their public statements.	.66	.57
Hosts of major television news shows do not know enough to be reliable sources of information.	.69	.55
People who are telling us how to act don't always have an incentive to tell the truth.	.50	.42
CFA fit statistics		
CFI	.954	
TLI	.943	
SRMR	.046	
RMSEA	.049	
$\chi^2(df)$	228.89 (98)	
N	547	

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

Table 2. Fit statistics for alternative models for the EvA scale.

	$\chi^2(df)$	RMSEA	SRMR	CFI	TLI	AIC
Proposed 4-dim model	228.89 (98)	.049	.046	.954	.943	27851.28
Alternative 6-dim model	687.43 (237)	.059	.068	.901	.884	43233.08
Alternative 5-dim model	404.01 (160)	.053	.050	.932	.920	35693.44
Alternative 4-dim model	351.33 (98)	.069	.073	.920	.902	27937.55

Note. Proposed 4-dimensional model: Four-factor solution with Factors 1, 2, 4, 5; *Alternative 6-dim:* Six-factor solution with six factors (Factors 1, 2, 3, 4, 5, 6); *Alternative 5-dim:* Five-factor solution with five factors (Factors 1, 2, 4, 5, 6); *Alternative 4-dim:* Four-factor solution with Factors 1, 2, 3, 4); dim = dimension. Details about Factors 1-6 are available in Table S6.

indicated that all items contributed to scale homogeneity (.40–.74; Streiner et al., 2015; e.g., Duckworth et al., 2007; Lipkus et al., 2001).

We examined the descriptive goodness-of-model fits of our proposed 4-factor solution (Factors 1, 2, 4, 5), compared to the 6-factor solution (all factors retained), a 5-factor solution (Factors 6 also retained), and a 4-factor solution (Factor 3, rather than 5, represents Pro-advice). Compared to the proposed model, alternative models had worse fits, indicated by greater RMSEA and SRMR or smaller CFI and TLI (Table 2). The proposed and alternative models were non-nested (i.e., did not share the same set of parameters), so we additionally compared models through the Akaike Information Criterion (AIC; after Schermelleh-Engel et al., 2003). The proposed model had a lower AIC, suggesting a better model fit compared to alternatives. These results corroborated our item and factor selections.

Factor correlations supported our decision to use oblique rotation, which treats factors as distinct but correlated (–.22–.36, Table S9). Correlations among EvA subscales suggested neither poor discriminant validity nor a single, higher-order factor. None were above .80, thus not suggesting poor discriminant validity, and these correlations varied across factors, suggesting that higher-order factors were unlikely (Brown, 2015). The four EvA constructs had acceptable internal reliability with both Cronbach's alpha (.68–.85, Table S9; Bland & Altman, 1997; DeVellis, 2017; Tavakol & Dennick, 2011; 'ltm' package in R) and McDonald's omega total (.68–.85, Table S9; McNeish, 2018; 'MBESS' package in R).

Study 1 and Study 2: Confirmatory factor analysis on the EvA Scale

The purpose of Study 2 was to assess the factor structure of the EvA scale by conducting confirmatory factor analysis (CFA) on the 16 items from Study 1 on an independent sample.

Methods

We recruited 201 US adults through CloudResearch. We excluded 12 participants who missed at least one of three attention checks, leaving 189 for analysis (college: 88; non-college: 101; demographics in Table S10).

Results

In Study 2, the proposed 16-item, four-factor EvA scale was supported by CFA, with acceptable model fit and strong factor loadings ($\chi^2(df)$ = 172.349 (98), RMSEA = .063, SRMR = .069, CFI = .928, TLI = .911; Table S7; DFI cutoffs in Table S8).

Studies 1 and 2 explored the structure of the EvA scale by examining four plausible alternative models (after Cable & DeRue, 2002; Medsker et al., 1994). We tested: (A) a more restricted version of our four-factor model, wherein the EvA factors are unrelated to one another (orthogonal), keeping all other specifications the same; (B) a model that assumes only one EvA tendency across all four factors; (C) a model that assumes two dimensions: Evidence-oriented reasoning (Pro-evidence, Anti-advice) and Advice-oriented reasoning (Pro-advice, Anti-evidence) with a two-factor solution with eight items each; (D) a model that assumes two alternative dimensions: Pro/anti-evidence (Pro-evidence, reverse-coded Anti-evidence) and Pro/anti-advice (Pro-advice, reverse-coded Anti-advice); (E) a hierarchical model with two second-order factors as specified in model C; (F) a hierarchical model with two-second order factors as specified in model D. Chi-square difference tests (Brown, 2015; Schermelleh-Engel et al., 2003) indicated a significantly better model fit for our proposed model compared to alternatives (Table 3; $ps < .01$). All five alternatives did not meet the criteria for an acceptable fit with respect to all model fit indices, indicating that it is more likely that there are four factors for accepting and rejecting evidence and advice, which are related and unlikely to reflect a single dimension, two combined factors, or two second-order factors. These results suggest convergent and discriminant validity of four related yet distinct EvA subscales: Anti-evidence, Pro-evidence, Anti-advice, and Pro-advice. This pattern of relative model fits was again confirmed in another independent sample from Study 3 (Table S11).

Discussion

Development, refining, and testing of our initial 57 items produced a smaller, 16-item EvA scale with four factors representing reasoning tendencies with respect to evidence and advice: Pro-evidence, Anti-evidence, Pro-advice, and Anti-advice. Thus, people can have distinct motivations and tendencies to seek or discredit evidence and rely upon or resist advice.

Study 3: Confirmatory factor analysis and convergent and discriminant validity

There were three purposes for Study 3: (1) assess the dimensionality of the EvA scale using CFA on an independent sample, (2) assess preregistered hypotheses regarding convergent and discriminant validity of the EvA scale, and (3) examine factor structure invariance between gender and education groups. Preregistration of Study 3 is available at: https://aspredicted.org/RTS_PWL.

Confirmatory factor analysis

Methods

We recruited 323 US adults through Prolific, an online crowdsourcing platform. Studies found that Prolific provides higher quality data compared to alternative platforms, demonstrated through better performance on attention checks, reproducibility, and lower dishonesty

Table 3. Fit statistics for the proposed and alternative models.

		χ^2 (*df*)	χ^2_{diff} (Δdf)	RMSEA	SRMR	CFI	TLI
Study 1 (*N* = 547)							
Proposed	4-dim	228.9 (98)		.049	.046	.954	.943
Alternative	A: 4-dim, orthogonal	477.4 (104)	248.5 (6)***	.081	.132	.867	.847
	B: 1-dim	1725.6 (104)	1496.8 (6)***	.169	.174	.424	.336
	C: 2-dim (evidence-oriented, advice-oriented)	930.7 (103)	701.8 (5)***	.121	.123	.706	.658
	D: 2-dim (pro/anti-evidence, pro/anti-advice)	1495.8 (103)	1267.0 (5)***	.157	.163	.506	.424
	E: 4-dim (hierarchical, C)	367.1 (103)	138.2 (5)***	.068	.095	.906	.891
	F: 4-dim (hierarchical, D)	461.9 (103)	233.0 (5)***	.080	.135	.873	.852
Study 2 (*N* = 189)							
Proposed	4-dim	172.4 (98)		.063	.069	.928	.911
Alternative	A: 4-dim, orthogonal	293.7 (104)	121.3 (6)***	.098	.149	.815	.787
	B: 1-dim	763.4 (104)	591.0 (6)***	.183	.194	.358	.259
	C: 2-dim (evidence-oriented, advice-oriented)	428.9 (103)	256.5 (5)***	.129	.134	.683	.630
	D: 2-dim (pro/anti-evidence, pro/anti-advice)	585.1 (103)	412.8 (5)***	.157	.162	.530	.453
	E: 4-dim (hierarchical, C)	243.9 (103)	71.5 (5)***	.085	.106	.863	.840
	F: 4-dim (hierarchical, D)	269.4 (103)	97.0 (5)***	.092	.137	.838	.811

Note. Proposed model: Items load on four factors (Pro-evidence, Anti-evidence, Pro-advice, Anti-advice) correlated to each other (oblique rotation); *A:* No relationships among four factors (orthogonal rotation); Alternatives B-F assume factors are correlated to each other; *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; dim = dimension. Results on Study 3 are presented in Table S11. **$p < .1$; **$p < .05$; ***$p < .01$.

(Palan & Schitter, 2018; Peer et al., 2017). Seven participants who did not pass at least one of four attention-checks were excluded, leaving 316 for analysis. Similar numbers of individuals with and without a college degree were recruited (college: 156; non-college: 160; demographics in Table S12).

Results

The EvA scale had an acceptable model fit in CFA with this independent sample, $\chi^2(df)$ = 209.99 (98), RMSEA = .060, SRMR = .071, CFI = .928, TLI = .912 (Table S7; DFI cutoffs in Table S8). No items were highly skewed to suggest removal (> ±2.0; Cassidy et al., 2005). All items loaded onto their corresponding factors with high standardized factor loadings (.56–.83, ps < .01) and all subscales had acceptable internal consistency (α = 68–.84) (Table S9). Model comparison results using Chi-square difference tests indicated that our proposed model had a significantly better fit compared to alternatives (ps < .01; Table S11). These results suggested again that EvA tendencies reflect four distinct factors, consistent with our preregistered hypothesis.

Assessing convergent and discriminant validity of the EvA Scale

Construct validity refers to the extent that an operationalization measures the construct it purports to measure (Westen & Rosenthal, 2003) by confirming that it is associated with existing indicators in ways that conform to theoretical expectations. 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).

Because Pro-evidence reflects the tendency to seek objective evidence to inform decisions, it was expected to relate more to the motivation toward deliberative thinking and not simply reflect an aptitude for numerical calculations. Anti-evidence, the tendency to suspect the validity of scientific or statistical evidence, was expected to relate more to the tendency to distrust science than the general tendency to be pessimistic about future outlooks. Because Pro-advice reflects the tendency to defer to others in decision-making, it was expected to be more similar to the tendency to abide by authorities in a social hierarchy over the general tendency to think people are other-regarding. Anti-advice, the tendency to suspect the credibility of others' claims and advice, was expected to be more similar to the tendency to resist hierarchical order or discipline than the tendency to think people are self-interested. Building upon theoretical expectations, we preregistered our hypotheses in relation to existing measures as follows: Pro-evidence reflects effortful thinking more than numerical ability; Anti-evidence reflects distrust in science more than general pessimism; Pro-advice reflects respect for conventions more than dispositional trust; Anti-advice reflects defiance to authority more than dispositional distrust. We additionally expected EvA tendencies to be distinct from the desire to be seen as socially acceptable.

Methods

Participants completed existing scales (the order of scales and items was randomized) after the EvA, prior to demographic items. In all studies, all variables were constructed as the average of the constituent items, scaled to range from 0 to 1.

Need for cognition. The extent to which individuals enjoy or engage in effortful thinking was measured with the Need for Cognition Scale (Coelho et al., 2018). Participants assessed six statements (e.g., "I would prefer complex to simple problems") on a seven-point scale ("strongly disagree" to "strongly agree"). The composite score averaged the six items, with two reverse-coded items (e.g., "Thinking is not my idea of fun"), α = .86.

Distrust in science. To measure the degree that individuals distrust science, we adopted the Trust in Science and Scientist Inventory (Nadelson et al., 2014). The original inventory included 21 items but we adopted the 12 items that are specific to 'distrusting' science (e.g., "Scientific theories are weak explanations"), measured on a five-point scale ("strongly disagree" to "strongly agree"), α = .93.

Respect for convention and defiance to authority. To measure the tendency to follow social conventions and resist authority in social hierarchies, we used the Right-Wing Authoritarianism scale (Bizumic & Duckitt, 2018). Participants indicated the degree to which they agree or disagree with two Conventionalism items (e.g., "God's laws about abortion, pornography, and marriage must be strictly followed before it is too late."), $|r|$ = .77, p < .01, and two Authoritarianism Submission items (e.g., "It's great that many young people today are prepared to defy authority"), $|r|$ = .30, p < .01, on a seven-point scale ("strongly disagree" to "strongly agree"). In our study, Respect for Convention refers to the Conventionalism scale, and Defiance to Authority refers to the reverse-coded Authoritarian Submission scale.

Numeracy. The ability to understand and use numerical information was measured by the Numeracy scale (Weller et al., 2013). Numeracy was measured as the proportion of correct answers out of eight numerical tasks (e.g., "If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 1000 people?"), α = .67.

Pessimism. The general tendency to have a negative view of life and future prospects was measured by the

Pessimism items (Scheier et al., 1994). Respondents indicated agreement or disagreement with three statements (e.g., "I hardly ever expect things to go my way") on a five-point scale ("disagree a lot" to "agree a lot"), α = .80.

Dispositional trust/distrust. To measure the general tendencies to believe others as being other-regarding or self-interested, we used the Dispositional Trust/Distrust items (Bianchi & Brockner, 2012). Respondents indicated their views on three statements (e.g., "Do you think most people would try to take advantage of you if they got a chance or would they try to be fair?") on a three-point scale (e.g., "take advantage" to "fair"), α = .70. The average of reverse-coded items was considered as indicating the dispositional 'distrust.'

Social desirability. To measure the desire to be seen as socially acceptable, we used the Social Desirability scale (Fischer & Fick, 1993) that asked respondents to indicate whether they think ten statements (e.g., "I never resent being asked to return a favor") describe them ("true" or "false). However, this scale unexpectedly failed to load together as a single construct (Table S13). Thus, social desirability was examined again in Study 4 using an alternative measure (Hart et al., 2015).

Results

To demonstrate construct validity, we adopted the correlational approach (Haws et al., 2012; Stöber, 2001; Watson et al., 1995), indicating convergent validity with moderate to strong correlation coefficients (e.g., .3–.6) and discriminant validity with weaker coefficients (e.g., ≤ .2) (Anastasi & Urbina, 1997; Ward et al., 2009). We additionally conducted tests of differences in dependent correlations (Haws et al., 2012; Steiger, 1980) to compare the relative strength of the relationships between the EvA scale and convergent versus discriminant scales.

Convergent and discriminant validity was indicated, with most existing scales correlating with their respective EvA tendency as expected, with significant differences between strongly versus weakly related constructs (Table 4). Pro-evidence was more strongly correlated with Need for Cognition (r = .39) than Numeracy (r = .07), and the difference between the two correlations was significant, t = 4.73, p < .01. Anti-evidence strongly correlated with Distrust in Science (r = .78) but weakly with Pessimism (r = −.02), and the difference was significant, t = 14.7, p < .01. Pro-advice was more strongly correlated with Respect for Convention (r = .45) than Dispositional Trust (r = .04), and the difference was significant, t = 5.61, p < .01. Anti-advice was strongly correlated with Defiance to Authority (r = .31) and was slightly more strongly related with Dispositional Trust than expected for discriminant validity (r = .23), and their difference was not significant, t = 1.14, p = .26. Despite this relatively weak finding, we considered these values close enough to recommended criteria to indicate convergent and discriminant validity (Anastasi & Urbina, 1997; Ward et al., 2009).

Table 4. Convergent and discriminant validity of the EvA scale with other relevant scales.

EvA tendency	Correlation with convergent construct	Correlation with discriminant construct	Test of relative correlation strength (t-value)
Pro-evidence	.39*** (Need for Cognition)	.07 (Numeracy)	4.73***
Anti-evidence	.78*** (Distrust in Science)	−.02 (Pessimism)	14.7***
Pro-advice	.45*** (Respect for Convention)	.0.04 (Dispositional Trust)	5.61***
Anti-advice	.31*** (Defiance to Authority)	.23*** (Dispositional Distrust)	1.14

Note. Entries are bivariate correlations between the EvA subscale and convergent/discriminant scales. *p < .1; **p < .05; ***p < .01.

Assessing measurement invariance of the EvA Scale

To assess generalizability of the EvA, we tested for measurement invariance across education and gender groups, to ensure equivalent relationships across subpopulations (Brown, 2015). We conducted multiple-group CFAs in college (n = 156) and non-college (n = 160) samples (Table S15), and examined measurement invariance in three steps: configural, metric, and scalar (Table S16; Steinmetz et al., 2009; Putnick & Bornstein, 2016). We additionally evaluated measurement invariance using the goodness-of-fit index approach, where the change of CFI between the constrained and unconstrained model (|−ΔCFI|) smaller than or equal to 0.01 indicates invariance across subgroups (Cheung & Rensvold, 2002).

Configural invariance testing indicated that the simultaneous equal form solution had an acceptable model fit, $\chi^2(196)$ = 322.56, RMSEA= .064, SRMR= .071, CFI = .913, TLI = .893. This result supports the same number of factors and pattern of fixed and free parameters (Steinmetz et al., 2009) between college and non-college groups. Metric (weak) measurement invariance was supported as the equality constraints on factor loadings across education groups did not significantly degrade model fit in chi-square differences, $\chi^2_{diff}(12)$ = 12.92, p = .37, or the goodness-of-fit difference, ΔCFI = −0.001. Scalar (strong) measurement invariance was supported as equality constraints on factor loadings and item intercepts did not significantly degrade model fit, $\chi^2_{diff}(12)$ = 16.58, p = .17; ΔCFI = −0.004. These data suggest that the EvA generalizes as a measure of reasoning tendencies between lower and higher education groups.

Using the same steps for gender groups (male: 145; female: 162; Table S17), the initial model that tests configural invariance had an acceptable fit, $\chi^2(196)$ = 321.26, RMSEA= .065, SRMR= .077, CFI = .92, TLI = .90, supporting equal factor structures between women and men (Table S18). Metric invariance was also supported, as equality constraints on factor loadings across genders did not significantly degrade fit, $\chi^2_{diff}(12)$ = 15.24, p = .23; ΔCFI = −0.002. Scalar invariance was supported, as equality constraints on factor loadings and item intercepts did not significantly degrade model fit, $\chi^2_{diff}(12)$ = 11.02, p = .53; ΔCFI = −0.001.

These results suggest that the EvA generalizes as a measure of reasoning tendencies between men and women.

It is worth noting that residual (strict) measurement invariance, the final step of measurement invariance, was supported, albeit weakly, between gender, but not education groups. The equality constraints on the sum of specific variances (variance of the items not shared with the factor) and error variance (measurement error) degraded model fit only weakly (i.e., $|-\Delta CFI| < 0.01$) for gender groups, $\chi^2_{diff}(16) = 24.14$, $p = .09$; $\Delta CFI = -0.006$, but significantly for education groups, $\chi^2_{diff}(16) = 37.68$, $p < .01$; $\Delta CFI = -0.019$. Thus, correlations among EvA tendencies and other manifest variables should not be compared across education levels. However, because residuals are not part of the latent factor and residual invariance can be too strict and unrealistic for group comparisons, residual invariance is often not considered a prerequisite for latent mean comparisons (Steinmetz et al., 2009; Putnick & Bornstein, 2016).

Discussion

In Studies 2 and 3, CFA verified that the EvA scale consists of 16 items that are best explained by four underlying constructs: Pro-evidence, Anti-evidence, Pro-advice, and Anti-advice, confirming our preliminary findings from the EFA in Study 1. Our findings also supported convergent and discriminant validity, since all four EvA subscales were more related to theoretically similar constructs (Need for Cognition, Distrust in Science, Respect for Convention, Defiance to Authority) than the less specific constructs (Numeracy, Pessimism, Dispositional Distrust and Trust). Even though our Anti-evidence scale was strongly correlated with Distrust in Science, this result does not undermine the uniqueness of our scale because Distrust in Science was measured with a subset (12 reverse-coded items) of the 21-item Trust in Science and Scientists scale (Nadelson et al., 2014). A strong correlation between Distrust in Science and Anti-evidence indicates that the EvA scale offers a new and more efficient measure of the specific tendency to resist scientific evidence, compared to Nadelson and colleagues' much longer scale that encompasses both trusting and distrusting science and scientists. Data also supported the idea that the EvA factor structure generalizes between lower and higher education groups and men and women.

Study 4: Criterion and discriminant validity of the EvA Scale

Criterion validity of a scale is demonstrated when the scale is highly correlated with the observable outcomes of external criteria, such as conceptually-related attitudes or behaviors (DeVellis, 2017; Motta et al., 2021). For instance, a measure of hockey players' aggressiveness should be highly correlated with minutes spent in the penalty box for aggression (Bushman & Wells, 1998). We focused on concurrent over predictive validity (Drost, 2011) because our criterion variables were measured at the same time as the EvA tendencies.

We assessed criterion validity by examining whether EvA tendencies are closely related to decision making contexts where individuals arrive at decisions through evidence or advice. We aimed to measure multiple types of decisions that have important impacts on the life of individuals and society.

Susceptibility to misinformation on social media was chosen to capture a decision making context where people are routinely exposed to expert evidence and non-expert advice; misinformation also has important consequences for society, such as reinforcing the partisan cleavages during presidential elections (Allcott & Gentzkow, 2017). Thus, we examined whether individuals who are more resistant to evidence or more reliant on advice would be more susceptible to misinformation.

Adherence to CDC guidelines and trust in the COVID-19 vaccine were added as another context with important real-world consequences for health and the transmission of disease (Vlasceanu & Coman, 2022); of interest to our key tendencies, the CDC can be construed as a source of expert evidence and as a trusted advisor. We examined whether participants who rely more on evidence would adhere to CDC guidelines and trust vaccines more, which can also be shaped by the tendency to follow or resist advisors.

Science curiosity was chosen to assess how the tendency to seek or devalue evidence was related to joy in consuming scientific information. For example, scientists aim to educate the public with public-facing blog posts or television documentaries, but they will not achieve the aim of educating the populace if people are not interested or if such interest is not related to using such evidence in decision making.

Religiosity was selected to test the hypothesis that people who rely more on advice from non-expert sources would also be more religious, given the strong role that religious advisors and authorities play in the lives of many people. This result is important particularly when common advice from experts and religious sources conflict, such as the conflicting advice from medical experts and religious leaders on vaccines (O'Neill, 2021; Wombwell et al., 2015).

Study 4 also assessed a preregistered hypothesis about the discriminant validity of the EvA scale compared to social desirability. Preregistration of Study 4 is available at: https://aspredicted.org/9HC_B7G.

Methods

We recruited 540 US adults through the survey platform Prolific. To address a Prolific issue at the time with gender imbalances (Charalambides, 2021), we balanced recruitment between men and women. 11 participants who missed at least one of four attention checks were removed, leaving 529 for analysis (college: 266, non-college: 263; male: 255, female: 269, gender self-identifying: 5; demographics in Table S19).

Criterion validity was assessed with ordinary least squares (OLS) with robust standard errors following our preregistered model specification (Table 5). We additionally confirmed the robustness of our findings with control variables (gender, age, education, and income), which did not alter

Table 5. Relationships between EvA reasoning tendencies and criterion behaviors or attitudes.

EvA Reasoning Tendencies	Susceptibility to health misinformation			Adherence to the CDC guide on social distancing			Confidence in COVID vaccine		
	b	95% CI	*t*	*b*	95% CI	*t*	*b*	95% CI	*t*
Pro-evidence	0.003	[−0.09; 0.10]	0.1	0.24	[0.09; 0.40]	3.0***	0.19	[0.05; 0.33]	2.6***
Anti-evidence	0.39	[0.31; 0.47]	9.6***	−0.23	[−0.35; −0.12]	−3.9***	−0.79	[−0.90; −0.68]	−14.1***
Pro-advice	0.17	[0.08; 0.26]	3.8***	−0.03	[−0.15; 0.10]	−0.4	0.37	[0.24; 0.51]	5.4***
Anti-advice	−0.05	[−0.14; 0.05]	−0.9	−0.13	[−0.27; −0.00]	−2.0**	−0.21	[−0.36; −0.07]	−2.8***
Constant	0.09	[−0.01; 0.19]	1.8*	0.71	[0.56; 0.87]	9.0***	0.81	[0.65; 0.96]	10.2***
N		529			529			529	
Adjusted R^2		.27			.08			.33	

EvA Reasoning Tendencies	Science Curiosity			Religiosity		
	b	95% CI	*t*	*b*	95% CI	*t*
Pro-evidence	0.64	[0.54; 0.74]	12.2***	0.06	[−0.09; 0.22]	0.8
Anti-evidence	−0.14	[−0.24; −0.05]	−2.9***	0.41	[0.28; 0.55]	6.1***
Pro-advice	0.06	[−0.05; 0.17]	1.2	0.21	[0.05; 0.36]	2.7***
Anti-advice	−0.04	[−0.16; 0.07]	−0.7	−0.18	[−0.34; −0.02]	−2.2**
Constant	0.09	[-0.02; 0.20]	1.6	0.20	[0.04; 0.37]	2.4**
N		529			529	
Adjusted R^2		.22			.13	

Note. *b* = unstandardized regression coefficient from OLS with robust standard errors. CI = confidence interval. *t* = t-value for regression coefficient. All variables were scaled to range from 0 to 1. *p < .1; **p < .05; ***p < .01.

the direction or statistical significance of the relationships (Table S20). These analyses examined the incremental associations between each EvA tendency and each criterion variable, controlling for the other three EvA tendencies. This approach was used because we conceptualize the four EvA tendencies as intercorrelated (supported by Table S9). We also reported bivariate correlations (Table S21), which were largely consistent with the regression coefficients in Table 5.

Measures

Susceptibility to health misinformation. Participants viewed eight social media posts about cancer treatments (from Scherer et al., 2021) in random order; four contained true information and four contained false information. For each post, participants indicated the degree that they perceived the information as accurate on a four-point scale ("completely false" to "completely true"). Susceptibility to misinformation was measured as the average perceived accuracy of false social media posts, α = .74. We confirmed that 97.5% of participants regularly used at least one social media platform (Facebook: 68.1%, Twitter: 54.4%, Instagram: 62.8%), rendering the task externally valid.

Adherence to CDC guidelines on COVID-19. To measure the degree to which individuals abided by the Centers for Disease Control and Prevention (CDC) guidelines on COVID-19, we used fourteen items from the CDC recommendations, such as six-foot social distancing, hand washing, and wearing a face mask (after Graupensperger et al., 2021) on a five-point scale ("never" to "all the time"). We adjusted items to reflect updated CDC guidelines at the time of data collection (CDC, 2022), such as changing avoiding all social gatherings to those that took place indoors. In these items, the CDC was not mentioned to avoid response bias. EFA 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); one item, "getting tested when feeling sick" did not meaningfully load onto either (Table S22). Because sanitizing is confounded by distancing (i.e., if you distance well, you need not sanitize), we used the composite of the ten distancing behaviors to indicate adherence to CDC guidelines, α = .93.

We additionally measured confidence in the safety and efficacy of COVID-19 vaccines (Vaccine Confidence Survey Question Bank; CDC, 2021). Participants answered four items (e.g., "How likely are you to recommend getting the COVID-19 vaccine to others?") on a five-point scale (e.g., "not at all likely" to "extremely likely"), α = .95.

Science curiosity. To measure the degree to which individuals enjoy consuming scientific information (Kahan et al., 2017), for efficiency, we adopted the reduced-form science curiosity scale (Motta et al., 2021). As planned in our preregistration, we replaced an item about attending public lectures, which was precluded by the pandemic, with an item about conversations about science from the original long-form scale (Kahan et al. 2017). Questions on a variety of topics (e.g., politics, religion, scientific research, celebrities) were presented, so that respondents would not infer the purpose of the questions, all on a four-point scale. Four items were relevant to measuring science curiosity (e.g., frequency of discussing scientific research with friends, family, or coworkers, "never" to "often"), α = .74.

Religiosity. To measure religiosity, we selected five items from the Religiosity scale (Rohrbaugh & Jessor, 1975) that did not contain archaic language and were applicable across religions. Participants answered five items (e.g.,

"How often have you attended religious services during the past year?") on a five-point scale (e.g., "never" to "every week"), α = .93.

Results

Assessing criterion validity of the EvA Scale

The relationships between the four EvA tendencies and potentially relevant attitudes and behaviors were largely consistent with our expectations about criterion validity, while illustrating the types of behaviors or attitudes that each tendency uniquely predicts.

Susceptibility to health misinformation. People who were more Anti-evidence and more Pro-advice were more likely to perceive health misinformation as accurate; Pro-evidence and Anti-advice were unrelated. These associations mimic dynamics on social media wherein people who resist scientific evidence or follow the advice of their own preferred sources of information can be misinformed by weakly-supported claims by figures such as politicians, influencers, or celebrities (Brennen et al., 2020; Bruns et al., 2021).

COVID-19 behaviors and attitudes. Social distancing was more strictly followed by people who were more Pro-evidence and was less followed among those who were more Anti-evidence and Anti-advice. Thus, the CDC does appear to operate as both a source of credible scientific evidence and as an advisor, supporting the utility of our four-dimensional EvA tendencies for understanding people's varied responses. Similarly, confidence in the COVID-19 vaccine was higher for individuals who were more Pro-evidence and Pro-advice and lower for people who were more Anti-evidence or Anti-advice.

Science curiosity. Individuals who were more Pro-evidence and those who were less Anti-evidence reported greater science curiosity, which was unrelated to Pro- or Anti-advice tendencies.

Religiosity. Individuals who were more Pro-advice or Anti-evidence were more religious whereas those who were more Anti-advice were less religious, consistent with the phenomenon whereby some highly religious people follow religious leaders whose teachings diverge from scientific evidence (Harding, 2014; O'Neill, 2021). The pattern was distinct from that of COVID-19 recommendations or vaccine trust—whereby people scored lower if they were either Anti-advice or Anti-evidence—but it is sensible given the time of the data collection, in which the CDC was particularly represented by an advisor in the director, Dr. Anthony Fauci (Vlasceanu & Coman, 2022). This difference between COVID-related attitudes and religiosity highlights how EvA tendencies can characterize people's unique response to advisors in different domains. For instance, federal health agencies promote science but can also be associated with an unevenly-trusted advisor (Lee et al., 2016) and trusted religious authorities may give advice not grounded in scientific evidence (Harding, 2014; O'Neill, 2021).

Assessing discriminant validity of the EvA Scale

We tested our preregistered hypothesis that EvA tendencies are distinct from social desirability using the correlational approach (Haws et al., 2012; Stöber 2001). Social desirability was measured by the Impression Management items on the Balanced Inventory of Desirable Responding (Hart et al., 2015), which is related to personality traits such as agreeableness, honesty, and self-enhancement tendencies (Müller & Moshagen, 2019). As hypothesized, social desirability was only weakly correlated with all four EvA tendencies, Pro-evidence (r = .17), Anti-evidence (r = .05), Pro-advice (r = −.06), Anti-advice (r = −.10), indicating discriminant validity (with correlations ≤ .2; Anastasi & Urbina, 1997; Ward et al., 2009).

Discussion

Study 4 demonstrates the criterion validity of EvA scale, highlighting its capacity to explain or predict important, real-world attitudes or behaviors that are germane to people's reliance on evidence or advice. The criterion measures included susceptibility to health misinformation (higher Anti-evidence and Pro-advice) and science curiosity (stronger Pro-evidence, weaker Anti-evidence). The EvA scale contributes to our understanding of individual differences in confidence in advice that is sometimes but not always guided by science (e.g., the CDC versus social media or religion). Interestingly, Pro-advice tendencies promoted confidence in the COVID vaccine but increased people's vulnerability to health misinformation on social media—because both can be considered sources of advice to be heeded. Individuals who are more Anti-evidence adhered less to CDC distancing recommendations on COVID-19 and were more religious. These EvA tendencies were also distinct from social desirability.

General discussion

There are many instances in human history where evidence improved the well-being of individuals and society, such as in our opening example of the measles vaccine (Rota et al., 2016). Despite this power to improve lives, some people are inclined to discredit evidence and value the opinion of preferred advisors (Martinez-Berman et al., 2020). In our current, high-choice media environment, where anti-science sentiments exist alongside misinformation, it is important to understand when people rely upon evidence or advice so that we can promote informed decisions and mitigate the

impact of misinformation. There were no existing scales that captured these reasoning tendencies or their underlying conceptual structure, so we developed and validated the Reasoning through Evidence versus Advice (EvA) scale.

Our results suggest that people's reasoning involves tendencies for valuing or devaluing scientific evidence, as well as valuing or discrediting advice. There were multiple possible underlying dimensional structures, which had never been tested, such as a single dimension from relying more on evidence to advice or one dimension for trusting versus untrusting for both evidence and advice. Our rigorous testing revealed and replicated the four distinct underlying dimensions of the EvA tendencies, through factor analyses on four large, independent samples. The four-factor model generalized across education and gender groups. In future research, four EvA subscales (Pro-evidence, Anti-evidence, Pro-advice, Anti-advice) can also be used separately, to efficiently meet the goals of one's research, such as to predict beliefs in false claims (Pennycook & Rand, 2019) or partisan-motivated misinformation or conspiracy theories (Miller et al., 2016). This scale builds upon related, existing scales that measure effortful thinking or attitudes toward hierarchical advice, but distinctly measures tendencies to approach or avoid evidence and advice for daily decisions. As expected, EvA tendencies are strongly correlated with conceptually similar existing scales but weakly related to more generic tendencies.

Our study has important implications for society, in which figures like politicians, celebrities, social media influencers, religious leaders, or even acquaintances can promote behaviors that defy evidence (Brennen et al., 2020; Bruns et al., 2021; Harding, 2014). We linked EvA tendencies to attitudes and behaviors in key social issues, including health misinformation, COVID-19 avoidance, science, and religion. Individuals who are more suspicious of evidence and follow advice are more susceptible to health misinformation. Those who trust evidence and advice also trust the COVID vaccine more and people who discredit evidence and advice followed CDC social distancing guidelines less. People who seek evidence or are less suspicious about scientific evidence were also more curious about science. Individuals who are more prone to following or less resistant to advice were also more religious. The fact that the pattern of the four EvA tendencies changes by context speaks to their utility in specifying a range of natural behavior and in helping us identify and improve reasoning across essential, everyday contexts like medicine, politics, and public policy that impact people's lives.

Preregistration for Studies 3 and 4 labeled the EvA constructs as evidence versus "authority." We used the word "authority" to mean non-expert figures from which people take advice in daily life. We later substituted the word with "advice" to more sharply distinguish EvA from scales on hierarchical authorities—the latter being a more narrow and distinct construct from our common, everyday one. Because the EvA construct itself was consistent throughout, we did not deviate from preregistration. Researchers who study reasoning by reliance on authority figures in everyday life may use the EvA scale to investigate trusted advisors specifically.

Our study is limited by the use of online, crowdsourced samples. Online samples are more demographically diverse than college student samples, and yield similar results to nationally representative samples (Buhrmester et al., 2011; Coppock, 2019; Paolacci & Chandler, 2014). We improved generalizability and validity by recruiting similar numbers of individuals with higher and lower levels of education and by excluding participants who indicated lapses in attention. However, our samples still overrepresented younger and more liberal people compared to the US population. We also did not yet test this scale in a non-US population; some of these relationships may be specific to or more or less pronounced in the US, or similar countries that are currently enmeshed in a polarized, partisan environment (Silver, 2022). Cross-cultural research is needed, for example in a nation characterized by a high level of religiosity where even one's political leaders may also be religious leaders, or in a less partisan and more educated nation.

Future work could perform additional validity tests on the EvA scale. For example, we only used a subset of scale items for some validation tests to minimize respondent fatigue (e.g., Trust in Science and Scientist Inventory, Nadelson et al., 2014), use internally consistent items (e.g., Right-Wing Authoritarianism, Bizumic & Duckitt, 2018), and eliminate archaic language or concepts (e.g., Religiosity scale, Rohrbaugh & Jessor, 1975). Validity could be re-tested using those complete scales. We also used dispositional trust to measure both trust and distrust (Bianchi & Brockner, 2012) to reduce survey length and respondent fatigue; discriminant validity could be re-tested using distinct constructs, for instance dispositional trust for Pro-advice and suspiciousness for Anti-advice. There may be additional traits associated with EvA tendencies that could be considered in the future, such as belief superiority, epistemic trust, inquisitiveness, or curiosity.

In response to widespread concerns about the politicization of science and the spread of fabricated news (Bolsen & Druckman, 2015; Vosoughi et al., 2018), our study examined reasoning through evidence versus advice, laying the groundwork for future research on information processing, belief formation, and decision making across contexts—including the high-choice media environments characterized by partisan polarization. The ability to assess these reasoning tendencies can facilitate instructional programs, communication strategies, and interventions that improve evidence-based reasoning, tailored to individuals or populations, ultimately contributing to the health of society and our democracy.

Author Note

This study was approved by the Institutional Review Board of the University of Michigan (HUM00167758) and involves no conflicts of interests.

This research was supported by a generous grant from the Martin and Ruth Jaffe Fund to Stephanie D. Preston. We thank helpful comments from Christopher Fariss, Yusaku Horiuchi, and Dean Lacy, and research assistance from Tyler Hill and Nicholas Hollman for the initial scale development

and pilot study, and Hannah Groenke and Amanda Szczesniak for conducting a survey and organizing results for Study 2. We also appreciate the constructive and helpful feedback from the editor and anonymous reviewers.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Open scholarship

This article has earned the Center for Open Science badges for Open Data, Open Materials and Preregistered. The data are openly accessible at https://osf.io/qeav5/?view_only=cd009d6d768344a58f90b0b4c10ba2f0 (DOI: 10.17605/OSF.IO/QEAV5). The preregistration for Study 3 is available at https://aspredicted.org/mw35r.pdf. The preregistration for Study 4 is available at: https://aspredicted.org/7ny46.pdf.

ORCID

Hwayong Shin http://orcid.org/0000-0002-1639-5444
Priti Shah https://orcid.org/0000-0001-5908-0453
Stephanie D. Preston https://orcid.org/0000-0002-2733-6108

References

Abramowitz, A. I., & Webster, S. (2016). The rise of negative partisanship and the nationalization of U.S. elections in the 21st century. *Electoral Studies*, *41*, 12–22. https://doi.org/10.1016/j.electstud.2015.11.001

Adcock, R., & Collier, D. (2001). Measurement validity: A shared standard for qualitative and quantitative research. *American Political Science Review*, 95(3), 529–546. https://doi.org/10.1017/S0003055401003100

Akobeng, A. (2005). Principles of evidence based medicine. *Archives of Disease in Childhood*, *90*(8), 837–840. https://doi.org/10.1136/adc.2005.071761

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, *31*(2), 211–236. https://doi.org/10.1257/jep.31.2.211

Altemeyer, B. (1988). *Enemies of freedom: Understanding right-wing authoritarianism*. Jossey-Bass.

Ames, S. L., Wolffsohn, J. S., & McBrien, N. A. (2005). The development of a symptom questionnaire for assessing virtual reality viewing using a head-mounted display. *Optometry and Vision Science* , *82*(3), 168–176. https://doi.org/10.1097/01.opx.0000156307.95086.6

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Prentice Hall.

Baker, R., Thomas, S., Thomas, P. W., Gower, P., Santonastaso, M., & Whittlesea, A. (2010). The emotional processing scale: Scale refinement and abridgement (EPS-25). *Journal of Psychosomatic Research*, *68*(1), 83–88. https://doi.org/10.1016/j.jpsychores.2009.07.007

Beavers, A. S., Lounsbury, J. W., Richards, J. K., Huck, S. W., Skolits, G. J., & Esquivel, S. L. (2013). Practical considerations for using exploratory factor analysis in educational research. *Practical Assessment, Research & Evaluation*, *18*, 6.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238–246. https://doi.org/10.1037/0033-2909.107.2.238

Berinsky, A. J., Margolis, M. F., Sances, M. W., & Warshaw, C. (2019). Using screeners to measure respondent attention on self-administered surveys: Which items and how many? *Political Science Research and Methods*, *9*(2), 430–437. https://doi.org/10.1017/psrm.2019.53

Bianchi, E. C., & Brockner, J. (2012). In the eyes of the beholder? The role of dispositional trust in judgments of procedural and interactional fairness. *Organizational Behavior and Human Decision Processes*, *118*(1), 46–59. https://doi.org/10.1016/j.obhdp.2011.12.005

Bizumic, B., & Duckitt, J. (2018). Investigating right wing authoritarianism with a very short authoritarianism scale. *Journal of Social and Political Psychology*, *6*(1), 129–150. https://doi.org/10.5964/jspp.v6i1.835

Bland, J. M., & Altman, D. G. (1997). Statistics notes: Cronbach's alpha. *BMJ (Clinical Research ed.)*, *314*(7080), 572–572. https://doi.org/10.1136/bmj.314.7080.572

Bolsen, T., & Druckman, J. N. (2015). Counteracting the politicization of science. *Journal of Communication*, *65*(5), 745–769. https://doi.org/10.1111/jcom.12171

Bolsen, T., & Druckman, J. N. (2018). Do partisanship and politicization undermine the impact of a scientific consensus message about climate change? *Group Processes & Intergroup Relations*, *21*(3), 389–402. https://doi.org/10.1177/1368430217737855

Brennen, S., Simon, F. M., Howard, P. N., Nielsen, R. K. (2020). Types, sources, and claims of COVID-19 misinformation. *Reuters Institute for the Study of Journalism*. https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation.

Bronstein, M. V., Pennycook, G., Bear, A., Rand, D. G., & Cannon, T. D. (2019). Belief in fake news is associated with delusionality, dogmatism, religious fundamentalism, and reduced analytic thinking. *Journal of Applied Research in Memory and Cognition*, *8*(1), 108–117. https://doi.org/10.1037/h0101832

Brown, N. J. S., Furtak, E. M., Timms, M., Nagashima, S. O., & Wilson, M. (2010). The evidence-based reasoning framework: Assessing scientific reasoning. *Educational Assessment*, *15*(3–4), 123–141. https://doi.org/10.1080/10627197.2010.530551

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd Ed.). Guilford Publications.

Browne, M., Thomson, P., Rockloff, M. J., & Pennycook, G. (2015). Going against the herd: Psychological and cultural factors underlying the 'vaccination confidence gap. *PLoS One*, *10*(9), e0132562. https://doi.org/10.1371/journal.pone.0132562

Bruns, A., Harrington, S., & Hurcombe, E. (2021). Coronavirus conspiracy theories: Tracing misinformation trajectories from the fringes to the mainstream. In M. Lewis, E. Govender, & K. Holland (Eds.), *Communicating COVID-19: Interdisciplinary perspectives* (pp. 229–249). Springer International Publishing.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science* , *6*(1), 3–5. https://doi.org/10.1177/1745691610393980

Bushman, B. J., & Wells, G. L. (1998). Trait aggressiveness and hockey penalties: Predicting hot tempers on the ice. *The Journal of Applied Psychology*, *83*(6), 969–974. https://doi.org/10.1037/0021-9010.83.6.969

Byrne, B. M. (1994). Testing for the factorial balidity, replication, and invariance of a measuring instrument: A paradigmatic application based on the Maslach Burnout Inventory. *Multivariate Behavioral Research*, *29*(3), 289–311. https://doi.org/10.1207/s15327906mbr2903_5

Cable, D. M., & DeRue, D. S. (2002). The convergent and discriminant validity of subjective fit perceptions. *The Journal of Applied Psychology*, *87*(5), 875–884. https://doi.org/10.1037/0021-9010.87.5.875

Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, *48*(3), 306–307. https://doi.org/10.1207/s15327752jpa4803_13

Cassidy, D. J., Hestenes, L. L., Hegde, A., Hestenes, S., & Mims, S. (2005). Measurement of quality in preschool child care classrooms: An exploratory and confirmatory factor analysis of the early childhood environment rating scale-revised. *Early Childhood Research Quarterly*, *20*(3), 345–360. https://doi.org/10.1016/j.ecresq.2005.07.005

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*(2), 245–276. https://doi.org/10.1207/s15327906mbr0102_10

CDC (2022). *How to Protect Yourself & Others. Centers for Disease Control and Prevention*. https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html.

CDC. (2021). Vaccine confidence survey question bank. *Centers for Disease Control and Prevention, 24*. https://www.cdc.gov/vaccines/covid-19/vaccinate-with-confidence/rca-guide/downloads/CDC_RCA_Guide_2021_Tools_AppendixD_Surveys-508.pdf.

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/.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Coelho, G. L. d H., H. P. Hanel, P., & J. Wolf, L. (2018). The very efficient assessment of need for cognition: Developing a six-item version. *Assessment*, *27*(8), 1870–1885. https://doi.org/10.1177/1073191118793208

Coppock, A. (2019). Generalizing from survey experiments conducted on Mechanical Turk: A replication approach. *Political Science Research and Methods*, *7*(3), 613–628. https://doi.org/10.1017/psrm.2018.10

DeVellis, R. F. (2017). *Scale development: Theory and applications* (Vol. 26). Sage publications.

Drost, E. A. (2011). Validity and reliability in social science research. *Education Research and Perspectives*, *38*(1), 105–123.

Druckman, J. N., Peterson, E., & Slothuus, R. (2013). How elite partisan polarization affects public opinion formation. *American Political Science Review*, *107*(1), 57–79. https://doi.org/10.1017/S0003055412000500

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, *92*(6), 1087–1101. https://doi.org/10.1037/0022-3514.92.6.1087

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272–299. https://doi.org/10.1037/1082-989X.4.3.272

Farias, M., Newheiser, A.-K., Kahane, G., & de Toledo, Z. (2013). Scientific faith: Belief in science increases in the face of stress and existential anxiety. *Journal of Experimental Social Psychology*, *49*(6), 1210–1213. https://doi.org/10.1016/j.jesp.2013.05.008

Ferguson, E., & Cox, T. (1993). Exploratory factor analysis: A users' guide. *International Journal of Selection and Assessment*, *1*(2), 84–94. https://doi.org/10.1111/j.1468-2389.1993.tb00092.x

Fischer, D. G., & Fick, C. (1993). Measuring social desirability: Short forms of the Marlowe-Crowne social desirability scale. *Educational and Psychological Measurement*, *53*(2), 417–424. https://doi.org/10.1177/0013164493053002011

Francis, L. J., & Greer, J. E. (1999). Attitude toward science among secondary school pupils in northern Ireland: Relationship with sex, age and religion. *Research in Science & Technological Education*, *17*(1), 67–74. https://doi.org/10.1080/0263514990170105

Gambrill, E. (1999). Evidence-based practice: An alternative to authority-based practice. *Families in Society: The Journal of Contemporary Social Services*, *80*(4), 341–350. https://doi.org/10.1606/1044-3894.1214

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Lawrence Erlbaum Associates.

Graupensperger, S., Lee, C. M., & Larimer, M. E. (2021). Young adults underestimate how well peers adhere to COVID-19 preventive behavioral guidelines. *The Journal of Primary Prevention*, *42*(3), 309–318. https://doi.org/10.1007/s10935-021-00633-4

Guzelian, P. S., & Guzelian, C. P. (2004). Authority-based explanation. *Science*, *303*(5663), 1468–1469. https://doi.org/10.1126/science.303.5663.1468

Harding, D. (2014). *South African preacher tells parishioners to eat grass to get closer to God*. New York Daily News.

Hart, C. M., Ritchie, T. D., Hepper, E. G., & Gebauer, J. E. (2015). The balanced inventory of desirable responding short form (BIDR-16). *SAGE Open*, *5*(4), 215824401562111. https://doi.org/10.1177/2158244015621113

Haws, K. L., Naylor, R. W., Coulter, R. A., & Bearden, W. O. (2012). Keeping it all without being buried alive: Understanding product retention tendency. *Journal of Consumer Psychology*, *22*(2), 224–236. https://doi.org/10.1016/j.jcps.2011.05.003

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, *7*(2), 191–205. https://doi.org/10.1177/1094428104263675

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Huber, C. R., & Kuncel, N. R. (2016). Does college teach critical thinking? A meta-analysis. *Review of Educational Research*, *86*(2), 431–468. https://doi.org/10.3102/0034654315605917

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, *22*(1), 129–146. https://doi.org/10.1146/annurev-polisci-051117-073034

Jamieson, K. H., McNutt, M., Kiermer, V., & Sever, R. (2019). Signaling the trustworthiness of science. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(39), 19231–19236. https://doi.org/10.1073/pnas.1913039116

Johnson, D. R., & Borden, L. A. (2012). Participants at your fingertips: Using Amazon's Mechanical Turk to increase student–faculty collaborative research. *Teaching of Psychology*, *39*(4), 245–251. https://doi.org/10.1177/0098628312456615

Kahan, D. M. (2015). Climate-science communication and the measurement problem. *Political Psychology*, *36*(S1), 1–43. https://doi.org/10.1111/pops.12244

Kahan, D. M., Landrum, A., Carpenter, K., Helft, L., & Hall Jamieson, K. (2017). Science curiosity and political information processing. *Political Psychology*, *38*(S1), 179–199. https://doi.org/10.1111/pops.12396

Lee, C., Whetten, K., Omer, S., Pan, W., & Salmon, D. (2016). Hurdles to herd immunity: Distrust of government and vaccine refusal in the US, 2002–2003. *Vaccine*, *34*(34), 3972–3978. https://doi.org/10.1016/j.vaccine.2016.06.048

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making*, *21*(1), 37–44. https://doi.org/10.1177/0272989X0102100105

Lobato, E. J. C., & Zimmerman, C. (2019). Examining how people reason about controversial scientific topics. *Thinking & Reasoning*, *25*(2), 231–255. https://doi.org/10.1080/13546783.2018.1521870

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, *49*(2), 433–442. https://doi.org/10.3758/s13428-016-0727-z

Martinez-Berman, L., McCutcheon, L., & Huynh, H. P. (2020). Is the worship of celebrities associated with resistance to vaccinations? Relationships between celebrity admiration, anti-vaccination attitudes, and beliefs in conspiracy. *Psychology, Health & Medicine*, *2020*, 1–10.

McDonald, R. P., & Ho, M.-H R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, *7*(1), 64–82. https://doi.org/10.1037/1082-989x.7.1.64

McNeill, K. L., & Martin, D. M. (2011). Claims, evidence, and reasoning. *Science and Children*, *48*(8), 52–56.

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, *23*(3), 412–433. https://doi.org/10.1037/met0000144

McNeish, D., & Wolf, M. G. (2023). Dynamic fit index cutoffs for confirmatory factor analysis models. *Psychological Methods*, *28*(1), 61–88. https://doi.org/10.1037/met0000425

Medsker, G. J., Williams, L. J., & Holahan, P. J. (1994). A review of current practices for evaluating causal models in organizational behavior and human resources management research. *Journal of Management*, *20*(2), 439–464. https://doi.org/10.1177/014920639402000207

Miller, J. M., Saunders, K. L., & Farhart, C. E. (2016). Conspiracy endorsement as motivated reasoning: The moderating roles of political knowledge and trust. *American Journal of Political Science*, *60*(4), 824–844. https://doi.org/10.1111/ajps.12234

Morof, D., Steinauer, J., Haider, S., Liu, S., Darney, P., & Barrett, G. (2012). Evaluation of the London measure of unplanned pregnancy

in a United States population of women. *PLoS One*, *7*(4), e35381. https://doi.org/10.1371/journal.pone.0035381

Motta, M., Chapman, D., Haglin, K., & Kahan, D. (2021). Reducing the administrative demands of the science curiosity scale: A Validation Study. *International Journal of Public Opinion Research*, *33*(2), 215–233. https://doi.org/10.1093/ijpor/edz049

Müller, S., & Moshagen, M. (2019). Controlling for response bias in self-ratings of personality: A comparison of impression management scales and the overclaiming technique. *Journal of Personality Assessment*, *101*(3), 229–236. https://doi.org/10.1080/00223891.2018.1451870

Nadelson, L., Jorcyk, C., Yang, D., Smith, M. J., Matson, S., Cornell, K., & Husting, V. (2014). I just don't trust them: The development and validation of an assessment instrument to measure trust in science and scientists. *School Science and Mathematics*, *114*(2), 76–86. https://doi.org/10.1111/ssm.12051

Nelson, R. (2019). US measles outbreak concentrated among unvaccinated children. *The Lancet. Infectious Diseases*, *19*(3), 248. https://doi.org/10.1016/S1473-3099(19)30074-X

Newton, C., Feeney, J., & Pennycook, G. (2021). *On the disposition to think analytically: Four distinct intuitive-analytic thinking styles*. *PsyArXiv* (Working paper).

O'Neill, J. (2021). February 8). Iranian cleric says COVID vax turns people into homosexuals. *New York Post*. https://nypost.com/2021/02/08/iranian-cleric-says-covid-vax-turns-people-into-homosexuals/.

Palan, S., & Schitter, C. (2018). Prolific.ac – A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27. https://doi.org/10.1016/j.jbef.2017.12.004

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science*, *23*(3), 184–188. https://doi.org/10.1177/0963721414531598

Pennycook, G., & Rand, D. G. (2019). Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of Personality*, *88*(2), 185–200. https://doi.org/10.1111/jopy.12476

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, *70*, 153–163. https://doi.org/10.1016/j.jesp.2017.01.006

Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods*, *46*(4), 1023–1031. https://doi.org/10.3758/s13428-013-0434-y

Perry, R. T., Murray, J. S., Gacic-Dobo, M., Dabbagh, A., Mulders, M. N., Strebel, P. M., Okwo-Bele, J.-M., Rota, P. A., & Goodson, J. L. (2015). Progress toward regional measles elimination—worldwide, 2000–2014. *MMWR. Morbidity and Mortality Weekly Report*, *64*(44), 1246–1251. https://doi.org/10.15585/mmwr.mm6444a4

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review* , *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

Pfeffer, J., & Sutton, R. I. (2006). Evidence-based management. *Harvard Business Review*, *84*(1), 62–74, 133.

Rohrbaugh, J., & Jessor, R. (1975). Religiosity in youth: A personal control against deviant behavior. *Journal of Personality*, *43*(1), 136–155. https://doi.org/10.1111/j.1467-6494.1975.tb00577.x

Rota, P. A., Moss, W. J., Takeda, M., de Swart, R. L., Thompson, K. M., & Goodson, J. L. (2016). Measles. *Nature Reviews. Disease Primers*, *2*(1), 16049. https://doi.org/10.1038/nrdp.2016.49

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Rynes, S. L., Colbert, A. E., & O'Boyle, E. H. (2018). When the "best available evidence" doesn't win: How doubts about science and scientists threaten the future of evidence-based management. *Journal of Management*, *44*(8), 2995–3010. https://doi.org/10.1177/0149206318796934

Sanderson, I. (2002). Evaluation, policy learning and evidence-based policy making. *Public Administration*, *80*(1), 1–22. https://doi.org/10.1111/1467-9299.00292

Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the Life Orientation Test. *Journal of Personality and Social Psychology*, *67*(6), 1063–1078. https://doi.org/10.1037//0022-3514.67.6.1063

Scherer, L. D., McPhetres, J., Pennycook, G., Kempe, A., Allen, L. A., Knoepke, C. E., Tate, C. E., & Matlock, D. D. (2021). Who is susceptible to online health misinformation? A test of four psychosocial hypotheses. *Health Psychology*, *40*(4), 274–284. https://doi.org/10.1037/hea0000978

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research*, *8*(2), 23–74.

Schommer, M. (1998). The influence of age and education on epistemological beliefs. *British Journal of Educational Psychology*, *68*(4), 551–562. https://doi.org/10.1111/j.2044-8279.1998.tb01311.x

Schraw, G., Bendixen, L. D., & Dunkle, M. E. (2002). Development and validation of the Epistemic Belief Inventory (EBI). In *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 261–275). Lawrence Erlbaum Associates Publishers.

Shearman, S. M., & Levine, T. R. (2006). Dogmatism updated: A scale revision and validation. *Communication Quarterly*, *54*(3), 275–291. https://doi.org/10.1080/01463370600877950

Silver, L. (2022, November 16). Most across 19 countries see strong partisan conflicts in their society, especially in South Korea and the *U.S. Pew Research Center*. https://www.pewresearch.org/short-reads/2022/11/16/most-across-19-countries-see-strong-partisan-conflicts-in-their-society-especially-in-south-korea-and-the-u-s/.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*(2), 245–251. https://doi.org/10.1037/0033-2909.87.2.245

Steinmetz, H., Schmidt, P., Tina-Booh, A., Wieczorek, S., & Schwartz, S. H. (2009). Testing measurement invariance using multigroup CFA: Differences between educational groups in human values measurement. *Quality & Quantity*, *43*(4), 599–616. https://doi.org/10.1007/s11135-007-9143-x

Stöber, J. (2001). The social desirability scale-17 (sds-17): Convergent validity, discriminant validity, and relationship with age. *European Journal of Psychological Assessment*, *17*(3), 222–232. https://doi.org/10.1027//1015-5759.17.3.222

Streiner, D. L., Norman, G. R., Cairney, J., Streiner, D. L., Norman, G. R., & Cairney, J. (2015). *Health measurement scales: A practical guide to their development and use* (5th ed.). Oxford University Press.

Strahan, R., & Gerbasi, K. C. (1972). Short, homogeneous versions of the Marlowe-Crowne Social Desirability Scale. *Journal of Clinical Psychology*, *28*(2), 191–193. https://doi.org/10.1002/1097-4679(197204)28:2<191::AID-JCLP2270280220>3.0.CO;2-G

Svedholm-Häkkinen, A. M., & Lindeman, M. (2017). Actively open-minded thinking: Development of a shortened scale and disentangling attitudes towards knowledge and people. *Thinking & Reasoning*, *24*(1), 21–40. https://doi.org/10.1080/13546783.2017.1378723

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Harper & Row.

Tanne, J. H. (2019). US county bars unvaccinated children from public spaces amid measles emergency. *BMJ*, *364*, l1481. https://doi.org/10.1136/bmj.l1481

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

van Prooijen, J.-W. (2017). Why education predicts decreased belief in conspiracy theories. *Applied Cognitive Psychology*, *31*(1), 50–58. https://doi.org/10.1002/acp.3301

Vlasceanu, M., & Coman, A. (2022). The impact of information sources on COVID-19 knowledge accumulation and vaccination intention. *International Journal of Data Science and Analytics*, *13*(4), 287–298. https://doi.org/10.1007/s41060-021-00307-8

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Waldinger, M. D. (2004). Lifelong premature ejaculation: From authority-based to evidence-based medicine. *BJU International*, *93*(2), 201–207. https://doi.org/10.1111/j.1464-410x.2004.04495.x

Ward, C., Fischer, R., Zaid Lam, F. S., & Hall, L. (2009). The convergent, discriminant, and incremental validity of scores on a self-report measure of cultural intelligence. *Educational and Psychological Measurement*, *69*(1), 85–105. https://doi.org/10.1177/0013164408322001

Watson, D., Weber, K., Assenheimer, J. S., Clark, L. A., Strauss, M. E., & McCormick, R. A. (1995). Testing a tripartite model: Evaluating the convergent and discriminant validity of anxiety and depression symptom scales. *Journal of Abnormal Psychology*, *104*(1), 3–14. https://doi.org/10.1037//0021-843x.104.1.3

Weller, J. A., Dieckmann, N. F., Tusler, M., Mertz, C. K., Burns, W. J., & Peters, E. (2013). Development and testing of an abbreviated numeracy scale: A Rasch analysis approach. *Journal of Behavioral Decision Making*, *26*(2), 198–212. https://doi.org/10.1002/bdm.1751

Westen, D., & Rosenthal, R. (2003). Quantifying construct validity: Two simple measures. *Journal of Personality and Social Psychology*, *84*(3), 608–618. https://doi.org/10.1037/0022-3514.84.3.608

Wombwell, E., Fangman, M. T., Yoder, A. K., & Spero, D. L. (2015). Religious barriers to measles vaccination. *Journal of Community Health*, *40*(3), 597–604. https://doi.org/10.1007/s10900-014-9956-1

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for nest practices. *The Counseling Psychologist*, *34*(6), 806–838. https://doi.org/10.1177/0011000006288127