

Findings From an Empirical Exploration of Evaluators' Values

American Journal of Evaluation
1-23

© The Author(s) 2022


Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/10982140211046537

journals.sagepub.com/home/aje



John M. LaVelle¹ , Clayton L. Stephenson²,
Scott I. Donaldson³, and Justin D. Hackett⁴

Abstract

Psychological theory suggests that evaluators' individual values and traits play a fundamental role in evaluation practice, though few empirical studies have explored those constructs in evaluators. This paper describes an empirical study on evaluators' individual, work, and political values, as well as their personality traits to predict evaluation practice and methodological orientation. The results suggest evaluators value benevolence, achievement, and universalism; they lean socially liberal but are slightly more conservative on fiscal issues; and they tend to be conscientious, agreeable, and open to new experiences. In the workplace, evaluators value competence and opportunities for growth, as well as status and independence. These constructs did not statistically predict evaluation practice, though some workplace values and individual values predicted quantitative methodological orientation. We conclude by discussing strengths, limitations, and next steps for this line of research.

Keywords

evaluation, values, research on evaluation, theory-practice, values-practice

Findings From an Empirical Exploration of Evaluators' Values

Evaluation practice and its attendant descriptions have commonly involved language around making systematic judgments about the merit, value, significance, credibility, and utility of whatever is being evaluated (e.g., Patton, 2008; Scriven, 1991), often with a focus on programs, policies, and interventions, while sometimes extending into arenas such as the evaluation of personnel, portfolios, and products. Values, value statements, and values-driven decisions are omnipresent in evaluation (House, 2015), and although they are often alluded to by practitioners and scholars, they are not always attended to with care and diligence, or explicitly reflected in practice. Julnes even goes so far as to suggest, "evaluators have often been unreflective, and even sloppy, in their approaches to

¹ University of Minnesota, Minneapolis, MN, USA

² University of Southern California, Los Angeles, CA, USA

³ University of Southern California San Diego, San Diego, CA, USA

⁴ California University of Pennsylvania, California, PA, USA

Corresponding Author:

John M. LaVelle, University of Minnesota, Minneapolis, MN 55455, USA.

Email: jlavelle@umn.edu

valuing” (2012, p. 4), which may be indicative of the fields’ lack of understanding about evaluators’ individual values and the ways those values lead to either valuing behaviors or other aspects of evaluation practice.

The difficulty of conscientiously incorporating values in practice is understandable. It is challenging to identify and articulate one’s own values and valuing processes, as well as discuss them with others that may not share the same set of values. Scholars have long argued for understanding the role individual values should play in evaluation and advocated for their explicit consideration in practice (e.g., Greene, 1997, 2005; House, 2015; Schwandt, 1997; Scriven, 1991; Shadish, 1998; Smith, 1980). Additional scholars have also called for a deeper understanding of group and cultural values and how they influence evaluation practice (e.g., Hood, Hopson, & Frierson, 2015; Kawakami, Aton, Cram, Lai, & Porima, 2008; Kirkhart, 2010). The observations by these and other prominent evaluators often appear alongside nuanced, deliberate discussions and prescriptions about valuing in evaluation, with a specific focus on justifying who, how, and when evaluators or stakeholders ought to render value judgments.

Prescriptions about the processes and importance of valuing, while important to discuss, neglect the importance of understanding the individual values held by evaluators and the ways in which those values might indeed influence evaluation practice. After all, not all evaluators prioritize the valuing aspect of evaluation in their practice, and indeed some prioritize use, learning, social justice, and methodology (Christie & Alkin, 2013). But, all evaluators do carry with them individual worldviews, values, and personality traits. Embedded, but perhaps not explicit in the aforementioned works, is a call for understanding who we are as individual evaluators and what values we collectively represent as a group of professionals. Shadish (1998) suggested evaluation theory is who we are, and that may indeed be largely accurate, but evaluator values may represent the *why* of who they are, and the personality traits of evaluators may explain *how* those values are expressed (Parks-Leduc, Feldman, & Bardi, 2015). Because of this, more empirical attention must be paid to broadly describe evaluator values and individual characteristics and exploring the links between those characteristics and the decision they make in their evaluation practice.

Values Development

Schwartz (1994) described values as a set of goals that transcend situations and contexts and serve as a guiding group of principles in the life of a person or group. Schwartz theorized that values constitute a coherent system that underpins decision-making, attitudes, and ultimately, behavior (Schwartz et al., 2012). Research on values suggests a range of origins and transmission mechanisms including parents (e.g., Barni, Ranieri, Scabini, & Rosnati, 2011), peers (e.g., Wetherill, Neal, & Fromme, 2010), educational institutions (e.g., DePoy & Merrill, 1988), religious institutions (e.g., Kim, 2008), cultural groups (e.g., Brown, 2002), and entry into a workforce or organization (Church, Burke, & Van Eynde, 1994). These conceptualizations leave open the idea that values can be both intrinsic in a person as well as acquired and modified through experience. Because values vary across individuals, groups, and sub-groups, it is possible that within any given population, there will be differences in values based on demographic characteristics such as sex, race, national origin, education level, political affiliation, and first-generation to college (e.g., Schwartz & Rubel-Lifschitz, 2009).

In the field of evaluation, Smith (1980, 1981) has long posited that individual and group values influence many aspects of evaluation practice, including the way in which an evaluation is perceived. That is, the same study could conceivably be considered exceptional, trivial, or even unethical depending on one’s perspective (Smith, 1980). In his argument, Smith detailed three main sources of values influencing evaluation practice including: (1) the evaluation context itself, (2) the political aspects of evaluation, and (3) organizational influences. Smith later expanded this perspective to

include other sources, such as organizational micro-cultures (Smith, 2007), ethics (Smith, 1998), social enterprises (Smith, 1995), and cultures (Smith, Chircop, & Mukherjee, 1993). He also suggested evaluation practice is, by its nature, a values-driven construct, and how even “neutral” evaluation approaches actually reflect value positions (Smith, 2007); thereby concluding that all evaluation approaches must be implemented purposefully, thoughtfully, and with awareness of the value positions they represent. Many of these sentiments are echoed in Schwandt’s (2008) discussions of evaluation as “something more” than the technocratic application of inquiry methodology, suggesting that because evaluation is a human endeavor, it must be influenced by human values, ethics, and worldviews. Schwandt (1997) not only argued for understanding evaluator values in terms of applied ethics, suggesting that they must be viewed in light of stakeholders’ perspectives, but also in the ways both values and ethics influence how evaluations are proposed, developed, and applied. Others (e.g., Gullickson & Hannum, 2019) have suggested ways in which evaluators’ values might be included in evaluator education programs to help nascent evaluators be more explicit in how they render value judgments in practice.

As Mark (2008) pointed out, discussions about evaluator values are limited in that they are primarily prescriptive (i.e., “these are the values that should guide your practice”), they focus on individual reflections about practice (i.e., “this is what I did and what I learned”), or focus on values-based inquiry (i.e., “tell me about your values and how they influence your practice”). Individual values are a difficult construct to study empirically for many reasons, not the least of which is the intensely personal nature of values, which leads to many people having difficulty discussing them. House (2015) summed this challenge up succinctly when he quasi-jokingly wrote, “I’ve been tempted to write about how personal values affect the work of evaluators I know ... [but] I can’t afford to alienate those remaining [friends]” (p. 3). However, the lack of data on evaluators’ values is a limitation for the field; especially if House (2015), Schwandt (1997, 2008), Smith (1980, 1981), and others are correct in their assertions that evaluators’ values guide all facets of evaluation practice and are influenced by multiple sources. Building from Smith’s assertion that many kinds of values influence evaluation and evaluators, we propose an empirical investigation of evaluators’ individual values, workplace values, career commitment values, political values, evaluation practice values, and personality traits.

Individual Values

According to Schwartz (1994), values represent personal concepts and beliefs, promote specific behaviors, transcend situations, and tend to be ordered by relative importance. In other words, values are guiding principles that influence attitudes and behaviors. In addition, Schwartz proposed a model of values based on a set of three universal human needs or motivations: (1) needs of individuals as biological organisms, (2) requirements of coordinated social interactions, and (3) survival needs of groups. Cross-cultural research supports Bardi and Schwartz’s (1996) conceptualization that 10 value types represent basic human motivations and goals: self-direction (pursuing independent thought and action), stimulation (having excitement, novelty, and challenges in one’s life), hedonism (pursuing pleasure and sensuous gratification), achievement (finding personal success through demonstrating competence according to social standards), power (obtaining social status and prestige, control or dominance over people and resources), security (safety, harmony, and stability of society, of relationships, and of self), conformity (restraining action, inclinations, and impulses likely to upset others and violate social expectations or norms), tradition (respecting and accepting customs and ideas that traditional culture or religion provide), benevolence (preservation and enhancement of the welfare of people with whom one is in frequent personal contact), and universalism (understanding, appreciation, tolerance and protection for the welfare of *all* people and nature). Individual values (power, national security, influential, equality, social justice, obedience, or social

power) nested within these value types derive meaning from the motivations they represent to the individual. For instance, freedom and independence are both self-direction values, with the motivation of encouraging the pursuit of independent thought and action.

Additionally, this model specifies relationships between the 10 value types (Schwartz, 1994). The importance placed on particular values results in certain value types being compatible with one another while other value types are more opposed to one another. Plotting the correlations between the values, Schwartz (1994) identified a two-dimensional structure defined by orthogonal axes. One axis contrasts conservation values (e.g., social order, family security, national security) with openness values (e.g., creativity, freedom, independence). The second axis contrasts self-enhancement values (e.g., authority, wealth, power) with self-transcendence values (e.g., helpfulness, honesty, loyalty). Placing greater importance on conservation values is associated with less personal importance on openness values. Likewise, placing greater importance on self-transcendence values is related to less importance on self-enhancement values. Thus, knowing the importance placed on the values within each of these domains should help to not only understand the organization of values but also which values, or which value dimension, may influence attitudes and behaviors. Schwartz et al. (2012) later expanded the taxonomy to include 19 dimensions, though many of the additions were sub-constructs of the original Schwartz taxonomy.

The theoretical framework proposed by Schwartz (1994; see Table 1) allows for the study of values that moves beyond simply considering the role and influence of individual values. Specifically, this approach focuses on the broader motivational domains that values serve. Particular domains in which to test this framework are career and work decisions. Understanding the broader domains, and not focusing exclusively on the influence of a few individual values, may provide greater insight into evaluators as individuals and as a professional group, as well as the activities in which evaluators engage in their practice.

Work Values

Many people spend a lot of their time at work or working, and so it makes sense that they would want to work for an organization whose values and priorities match their own (Leuty & Hansen, 2011). In industrial/organizational psychology, this is broadly described as the person-job fit, and researchers have suggested aligning the person with the job is critical for understanding job satisfaction (Dawis & Lofquist,

Table 1. Operational Definitions of Schwartz's Theory of Basic Individual Values.

Value	Operational Definition
Universalism	Understanding, appreciation, tolerance, and protection for the welfare for all people and nature
Benevolence	Preservation and enhancement of the welfare of people with whom one is in frequent personal contact
Conformity	Restraint of actions, inclinations, and impulses likely to upset or harm others or violate social expectations or norms
Tradition	Respect, commitment, and acceptance of the customs and ideas that traditional culture or religion provides
Security	Safety, harmony, and stability of society, of relationships, and of self
Power	Social status and prestige, control or dominance over people and resources
Achievement	Personal success through demonstrating competence according to social standards
Hedonism	Pleasure and sensuous gratification for oneself
Stimulation	Excitement, novelty, and challenge in life
Self-Direction	Independent thought and action; choosing, creating, exploring

Note. Operational definitions adapted from Schwartz (1994).

1984) and organizational commitment (Meyer, Irving, & Allen, 1998). Although meta-analyses have not found a linear relationship between the constructs (Mathieu & Zajac, 1990), some believe the personal values-career commitment relationship is mediated by a person's workplace values (Meyer et al., 1998). Scholars have noted a remarkable similarity between the tools used to measure person-job fit (Berings et al., 2004; Rounds & Armstrong, 2005). For example, Manhardt (1972) constructed a tool to assess person-job fit by presenting participants with a series of prompts and asking them how important it was that a job have those characteristics. Meyer et al., later refined Manhardt's measure into three major constructs: comfort and security (regular routine, job security, and comfortable working conditions), competence and growth (intellectually stimulating, work makes a social contribution, and encourages continued development of knowledge and skills), and status and independence (permits working independently, is respected by others, and is respected by other people).

Political Values

The political nature of evaluation has been discussed by scholars such as Azzam (2010), Azzam and Levine (2014), House (2015), and Weiss (e.g., 1987, 1998). These discussions have often focused on the political ways in which evaluations are conceptualized, developed, implemented, reported, and used (e.g., Palumbo, 1987). Even more fundamental to the discussion, however, is the fact that politics and political decisions originate in both individuals and groups, necessitating an understanding of individuals' perspectives on both social and fiscal issues. Congruence theory, when applied to political decision-making or attending to weighing arguments (e.g., Taber, Cann, & Kucsova, 2009), suggests that when groups of people make decisions, the process is smoother when their perspectives on social programming and fiscal responsibility are similar. It is reasonable to believe that a similar phenomenon might take place in the space between evaluators and their stakeholders. However, little data exists about evaluators' personal political values. Given evaluation's roots in the United States in social programming and accountability (Christie & Alkin, 2013; Shadish, Cook, & Leviton, 1991), and that most evaluators work in contexts trying to help individuals and groups "do better" or "perform better" or "work together better," we might assume that evaluators tend toward being socially liberal and fiscally conservative, but data does not yet exist to test this assumption.

Personality Traits

Although personality traits and values are conceptually similar, there are some important differences. The most significant difference being that traits are descriptive whereas values are motivational (see Parks-Leduc et al., 2015). Further, personality traits are traditionally defined in terms of "habitual patterns of cognition, affect, and behavior" (Winter, John, Stewart, Klohnen, & Duncan, 1998, p. 232); particularly overt, observable behavior, but research has expanded the descriptions to include the importance of covert cognitive and affective responses as well (Zillig, Hemenover, & Dienstbier, 2002). McCrae and Costa (1987, 1997, 1999) proposed a robust and well-researched taxonomy of personality traits called the Big 5. The Big 5 are sometimes referred to by the acronym OCEAN: Openness to new experiences (inventive/curious), Conscientiousness (efficient/organized), Extraversion¹ (outgoing/energetic), Agreeableness (friendly/compassionate), and Neuroticism (sensitive/nervous). It is important to note that while all people have aspects of each personality trait, they are not intended to render value judgment about the worth of the person or the behavior (e.g., the neuroticism trait has negative connotations). In addition to being psychometrically sound, the Big 5 has the advantage of reducing hundreds of personality characteristics into five major domains, and when researchers discuss them, the disagreements tend to focus on nuanced interpretation of the model instead of the quality of the model itself (Costa & McCrae, 1992; Parks-Leduc et al., 2015; Zillig et al., 2002).

Previous research has suggested positive relationships between specific values and personality traits, such as between benevolence and tradition (values) with agreeableness (trait), self-direction and universalism (values) with openness (trait), and achievement and conformity (values) with conscientiousness (Roccas, Sagiv, Schwartz, & Knafo, 2002). In the case of evaluators, understanding their personality traits may help contextualize the motivations that guide why and how evaluators express their values through their overt behavior. For example, an evaluator that scores highly on the agreeableness scale might be motivated by the different values and mechanisms, such as a benevolent worldview or a value set that prioritizes tradition. Furthermore, understanding the values that motivate the behavior may (1) help an evaluator explain the rationale for their behavior, (2) help others understand that evaluators' behavior better, and (3) help the field understand the ways in which evaluators' values and traits are developed.

Evaluation Practice and Practice Values

Evaluation practice is difficult to assess with empirical data, partially because of the prevalence of prescriptive evaluation theories of practice and the broad absence of descriptive evaluation theory (Shadish et al., 1991). That is, evaluative scholars and theorists prescribe activities and considerations they believe they and others should incorporate into their practice (e.g., Key Evaluation Checklist, Scriven, 2007, p. 17 Steps to Utilization-Focused Evaluation, Patton, 2013), but it is not clear how well those prescriptions translate into on-the-ground practice for working evaluators. Recognizing this challenge, described as the theory-practice gap in evaluation, Christie (2003) developed a tool to broadly describe evaluation practice across two dimensions: stakeholder involvement ("the degree to which the stakeholder process is directed by stakeholders," p. 16), and methodological proclivity ("the degree to which the evaluation is guided by a prescribed technical approach," p. 20). Christie analyzed these constructs using multidimensional scaling to understand how the constructs related to each other using nonlinear analysis. Furthermore, Christie mapped the behavior of a number of prominent evaluation theorists to those constructs and showed the degree to which the theorists' prescriptive behaviors mapped onto practitioner behavior.

Purpose of the Study

A limitation of the previous scholarship on evaluators' values is that it is either prescriptive (i.e., "these are the values you should have") or reflective (i.e., "these are my values and here is how I think about them in my practice"). To the best of our knowledge, none of the previous scholarship on values has actually described evaluators' values or traits, or used the data to predict practice. The purpose of this study was to get a broad understanding of the values and personality traits of evaluators, and explore how those constructs relate to their practice. Specifically, we sought to address four main questions. First, how do evaluators' values map onto Schwartz's taxonomy of Basic Human Values, workplace values, and political values? Second, how do the personality traits of evaluators map onto the Big 5 personality traits? Third, are the values and traits different based on respondent's sex, race, generation status, or highest degree earned? Fourth, to what degree do evaluators' values and personality traits predict evaluation practice?

Method

Protection of Human Subjects

This study was reviewed and approved by the Institutional Review Board at both the lead and partner institutions.

Participants

An invitation to participate via an online survey link was emailed to a simple random sample of 1,000 American Evaluation Association members. Of the 1,000 people in the sample, 330 individuals opened the invitation, equating to a 33% initial response rate. Two hundred forty-four people reached the end of the survey. Two hundred eight participants completed at least 80% of the total items, resulting in a usable response rate of 20.8%.

Table 2 shows the participants' demographics. The respondents' racial and sex demographics were consistent with the data available from the 2008 AEA member scan, which reported that 67% of AEA members were female and 73% were Caucasian. In terms of education level, the study respondents were slightly different from the member scan. Respondents with a master's degree were more prevalent in the study compared with the 2008 member scan (44.6% vs. 42%), and individuals with doctoral degrees were less prevalent in the current study (48.98% vs. 52%).

Survey Procedure

The survey was programmed using the Qualtrics survey platform. Participants were first instructed to reflect on their individual values as well as their typical evaluation practice. Participants then responded to a series of validated scales composed of Likert-type items and open-ended questions

Table 2. Participant Demographics.

Demographic variable	<i>n</i>	% or <i>M</i> (<i>SD</i>)
Gender		
Female	143	68.8
Male	59	28.4
Other	3	1.4
Declined to state	3	1.4
Ethnicity		
Asian American/Asian	15	7.2
Black/African American	7	3.4
Decline to state	7	3.4
First Nations Peoples (or, Native American/Alaskan Native)	2	1.0
Hispanic (or, of Spanish Origin)	9	4.3
Hispanic or Latino	9	4.3
Other	12	5.8
White/Caucasian	156	75.0
Age	190	42.46 (12.15)
Educational attainment		
Bachelor	12	5.7
Master	95	45.5
Advanced graduate work or PhD	102	48.8
Graduate student		
Yes	39	18.7
No	170	81.3
First generation		
Yes	60	28.8
No	148	71.2

Note. Not all participants responded to each question, so the section totals may not sum to 208.

(e.g., advice for colleagues, advice for educators). The instruments were ordered randomly for each participant to minimize the chance of order effects and fatigue effects. Participants were not directly compensated for their participation, but the researchers made a \$5 donation to the American Evaluation Association Student Travel Fund for each usable survey. Participants received two invitations to complete the survey before the study was closed to new responses.

Instruments

Schwartz values inventory (SVI). Schwartz and colleagues' (2012) long-form instrument was utilized for the current study. The SVI measure instructed respondents to indicate the degree to which 58 items were personally important in their life using a 7-point Likert scale anchored with the terms "not at all important" and "very important." All items began with a root concept in capital letters, followed by a short explanation in parentheses. Example items include "EQUALITY (equal opportunity for all)," "FREEDOM (freedom of action and thought)," "MEANING IN LIFE (a purpose in life)," and "WEALTH (material possessions, money)." The questions map onto the ten constructs listed in Table 1. The version of the SVI used in the current study had construct alpha values ranging from .74 to .89 (Stern, Dietz, & Guagnano, 1998).

Workplace values inventory (WVI). Meyer et al.'s (1998) measure of workplace values asks participants to respond to twenty-one Likert questions on how important specific aspects of the workplace are important to them ("not at all important" to "very important"). The questions group into three factors: Comfort and Security, Competence and Growth, and Status and Independence. Sample items began with the preface, "Please use the scale to rate how important each of the following values is as a guiding principle in your work." Example statements included "permits a regular routine in time and place of work," "has clear-cut rules and procedures to follow," and "permits you to develop your own methods of doing tasks." This version of the WVI has alphas ranging from .68 to .8 for each of the constructs (Fields, 2013). Our Cronbach alpha values were .73 ($n = 213$) for Comfort and Security, .80 ($n = 211$) for Competence and Growth, and .71 ($n = 213$) for Status and Independence.

Political values (PV). We used a very general measure of respondents' political values (Malsch, 2005). The scale consists of three questions asking respondents to respond to a 7-point Likert scale (very liberal to very conservative) to indicate their perceived political values (1) in general, (2) on social issues, and (3) on fiscal issues. Our Cronbach's alpha for the combined three items was .85 ($n = 213$).

Big 5 inventory (BFI). The Big 5 Inventory (John & Srivastava, 1999) is made of 44 Likert items that ask respondents to indicate their level of agreement questions that tap into the OCEAN framework. Sample questions include "I remain calm in tense situations," "I have an assertive personality," and "I value artistic, aesthetic, experiences." The alpha value for each construct ranges from .81 to .87 (John & Srivastava, 1999). Our alpha levels were close or within this range with an alpha of .78 ($n = 211$) for Openness, .82 ($n = 212$) for Conscientiousness, .87 ($n = 210$) for Extraversion, .76 ($n = 212$) for Agreeableness, and .82 ($n = 212$) for Neuroticism.

Christie evaluation practice scale. We used Christie's (2003) measure of evaluation practice to examine evaluators' practice. The original scale suggested that evaluation practice could be mapped across two major dimensions using multidimensional scaling. The first dimension was the degree to which the evaluator includes stakeholders in the evaluation process, and the second dimension was the evaluators' methodological proclivity and adherence to positivist, post-positivist, pragmatist, and constructivist epistemological perspectives. The participants responded to 17 items using a

5-point Likert scale. Example questions include “I believe that evaluation conclusions are a mixture of facts and values,” “When I conduct an evaluation, the primary users help conceptualize and decide upon the evaluation questions,” and “My evaluation approach intends to create changes in the culture of the organization (agency) where the evaluation is being conducted.” The scales’ psychometric values are described in the Results section.

Analytical Procedures

The current study collected both quantitative and qualitative (i.e., open-ended questions) data. There was an emphasis on the quantitative data with qualitative data acting as a supplement (QUAN + qual). The data were not mixed to answer the research questions, and the qualitative data, which focused on advice for evaluator educators, will be reported in future projects.

Quantitative analysis. SPSS v.25 was used to analyze the quantitative data. First, the scales were examined using basic descriptive statistics including means, standard deviations, and confidence intervals. Second, because the majority of measures had already been validated on other populations, we used the previously proposed model structures to scale our study variables with the exception of Christie’s (2003) Evaluation Practice Scale. Christie’s scale had not been validated through ongoing research; thus, we checked Cronbach’s alpha for Christie’s two factors and conducted an exploratory factor analysis (EFA) of our own using Principal Axis Factoring (PAF). Third, we tested the models for between-group differences. Fourth, we used the model structures to predict evaluation practice as the dependent measure. We used confidence intervals to inform our interpretation of the data because they are better at estimation about the population (see Cumming, 2014; Cumming & Finch, 2005; Lindsay, 2015) and lead to a better understanding about meaningful (i.e., practical) inferences in the data. More specifically, confidence intervals are built so that interpretation can be directly connected to the scale and how meaningful a 1-point difference might practically mean. Confidence intervals also inform how precise a measurement was from a specific sample.

Power analysis. An a priori power analysis was conducted using G*Power (Faul, Erdfelder, Buchner, & Lang, 2009) to determine the sample size needed to conduct a regression analysis. All 19 constructs from the SVI, WVI, PV, and Big 5 measures were entered as the number of tested and total predictors in the analysis with an alpha-level of .05, a moderate effect size ($f^2 = .15$), and a power of .80, which resulted in a total required sample size of 153. Our obtained power for each model is provided in the results section.

Results

The purpose of this study was to provide empirical data on evaluators’ values, test to see if the values differ based on respondent characteristics, and explore the relationship between values and practice. We first report the results of each scale individually, followed by demographic analysis, and conclude with exploratory multiple regression predicting evaluation practice. For all analyses, cases were excluded using a pairwise method so that small amounts of missing data in one scale would not exclude participant data from being used in other scales and analyses (George & Mallery, 2008). See Table 3 for a summary of all constructs in the study, including a number of respondents, means, standard deviations, and the upper and lower limits for 95% confidence intervals.

Schwartz Values Inventory

The SVI maps the respondents onto 10 major constructs: Achievement, Benevolence, Conformity, Hedonism, Power, Security, Self-direction, Stimulation, Tradition, and Universalism. The scale ranges from 1 to 7, with low scores indicating less importance and higher scores indicating greater importance. Descriptive statistics illustrated the data were distributed normally without significant skew or kurtosis. Especially important are the confidence interval ranges: they are precise (i.e., narrow) and show that people in the field of evaluation are more similar than different in their values. Our data suggest that the values evaluators espouse most are Self-direction, Universalism, Achievement, and Benevolence. Power was the least important construct with Tradition also receiving a lower score (see Table 3). Figure 1 shows a visual of evaluators' values on the Schwartz scale by illustrating how the average score on each construct maps onto the overall taxonomy. For example, evaluators scored highly on Universalism, and as such the data point is plotted close to the edge of the radar map. Conversely, Power was not as important of a value; thus, it was closer to the center of the figure.

Table 3. Means, Standard Deviations, and Confidence Intervals for All Measures.

	<i>n</i>	<i>M</i>	<i>SD</i>	95% CI	
				LL	UL
Schwartz values inventory					
Achievement	210	5.84	0.71	5.74	5.94
Benevolence	208	5.84	0.71	5.74	5.93
Conformity	210	4.90	0.71	4.74	5.04
Hedonism	210	4.84	1.05	4.70	4.98
Power	211	3.74	1.06	3.60	3.89
Security	206	5.14	0.78	5.03	5.25
Self-direction	211	5.94	0.76	5.84	6.04
Stimulation	212	4.81	1.12	4.66	4.96
Tradition	212	4.09	0.99	3.95	4.22
Universalism	208	5.93	0.72	5.83	6.03
Workplace values inventory					
Comfort and security	212	3.47	0.76	3.36	3.57
Competence and growth	211	4.24	0.52	4.17	4.31
Status and independence	213	3.53	0.62	3.45	3.62
Political values					
General view	210	2.24	1.21	2.07	2.40
Social issues	210	1.93	1.19	1.77	2.09
Fiscal issues	210	2.91	1.58	2.69	3.12
Overall average	209	2.36	1.17	2.20	2.52
Big 5 inventory					
Agreeableness	212	3.99	0.52	3.91	4.06
Conscientiousness	212	4.05	0.58	3.97	4.13
Extraversion	210	3.20	0.86	3.09	3.32
Neuroticism	212	2.64	0.68	2.55	2.73
Openness	211	3.89	0.55	3.82	3.97
Christie evaluation practice scale					
Evaluation processes	208	4.13	0.66	4.04	4.22
Quantitative methodological orientation	207	2.99	0.82	2.87	3.10

Note. Sample sizes differ across constructs due to using Pairwise Method. CI = confidence interval; LL = lower limit; UL = upper limit.

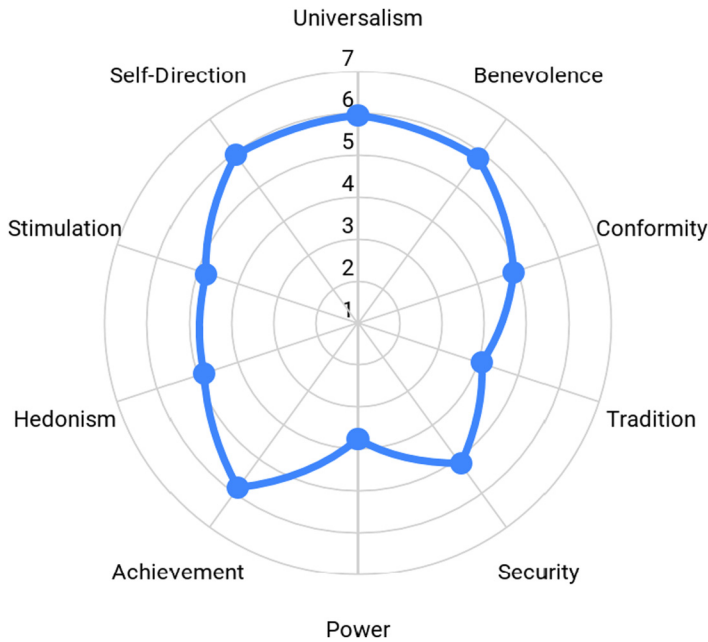


Figure 1. Evaluators average scores mapped into Schwartz Taxonomy of Basic Individual Values.

Workplace Values Inventory

The WVI (Meyer et al., 1998) breaks down into three factors (Comfort and Security; Competence and Growth; Status and Independence) that scale from 1 (not very important) to 5 (very important). Evaluators responded that Competence and Opportunities for growth are most important in their workplace. Status and Independence and Security and Comfort were equally ranked second behind Competence and Opportunities. Similar to the constructs in the Schwartz Values Inventory, the confidence intervals for the WVI constructs are quite narrow suggesting precision in representing the hypothetical “true” score across the population of evaluators.

Political Values

As previously stated, we used a general measure of respondents’ political values. Participants responded to three items on a 7-point Likert scale (1 = very liberal to 7 = very conservative) about their political values in general, social issues, and fiscal issues. The data suggest that evaluators’ values skew toward liberal. Respondents’ average score on the general political view indicated that 73.3% were either “very liberal” or “liberal.” Evaluators’ view on social issues was even more liberal with 82.4% of the respondents being “very liberal” or “liberal.” In contrast, only 49.3% of evaluators were “very liberal” or “liberal” when considering fiscal issues. Confidence intervals for each item were still quite precise, but showed a little more range compared to the other scales. We were surprised by the large percent difference between social issues and fiscal issues and checked the confidence intervals for the mean difference, $M_{Diff}=0.98$, $SD=1.38$, 95% CI [0.79, 1.16] compared to the mean difference between social issues and general political views, $M_{Diff}=0.31$, $SD=0.70$, 95% CI [0.21, 0.41] and between general political views and fiscal issues, $M_{Diff}=0.67$, $SD=1.15$, 95% CI [0.51, 0.82].

Big 5 personality traits

The Big 5 personality traits map onto five major constructs: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. The scales range from 1 to 5, with a low score suggesting the trait does not represent the evaluator well, and a high score indicating the trait is illustrative of evaluators. Confidence intervals for all traits were narrow, again, suggesting with reasonable precision that evaluators, overall, have low Neuroticism, moderate Extraversion, and high Conscientiousness, Agreeableness, and Openness (see Table 3 for descriptives and Figure 2 for a radar map).

Christie Evaluator Practice Scale

Christie's (2003) research used multidimensional scaling (MDS) to develop item groupings that broadly map onto two major factors: Stakeholder Involvement and Methodological Proclivity. Because Christie's scales had not yet been replicated and validated by ongoing research, we conducted an exploratory analysis to ascertain the degree to which the items in the scales group together. In our sample, the initial Cronbach alpha values were somewhat acceptable for Stakeholder Involvement ($\alpha = .67$) and unacceptable for Methodological Proclivity ($\alpha = .30$). Due to these low-reliability scores, we elected to refine the items into scales that were more amenable to regression analysis.

We conducted an EFA using the following methods based on Costello and Osborne's (2005) best practices guidelines. PAF with oblique rotation was used to identify latent factors for the initial 17-item scale with the goal of refining Christie's (2003) scale into discrete constructs with high internal consistency. The factor extraction was guided by analysis of a scree plot, factors with eigenvalues greater than one, and expertise with theoretical and applied evaluation practice. Items with high cross-loadings (i.e., $> .40$) on other hypothesized factors, or low factor loadings on its specific factor (i.e., $< .32$; Devellis, 2012; Yong & Pearce, 2013) were removed. Only items loading onto one factor and possessing high loadings ($> .40$) were retained, resulting in a two-factor solution. Factor one, which we are describing as *Evaluation Processes*, is composed of nine items with a

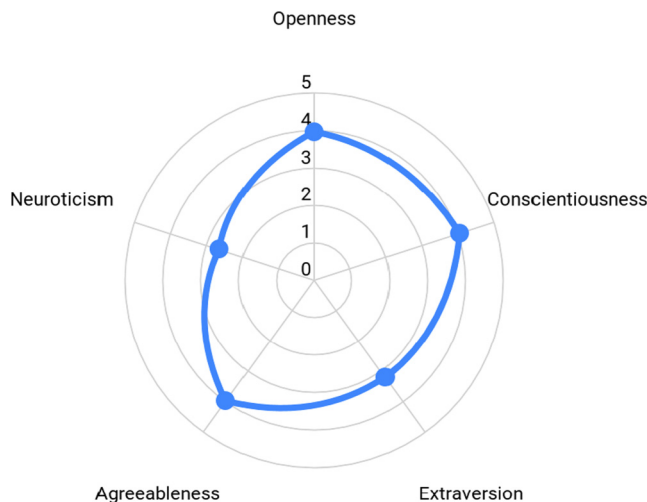


Figure 2. Evaluators average scores mapped into Big 5 personality traits.

Cronbach's alpha of .83, and an eigenvalue of 4.00, explained 28.56% of the variance. Factor two, which we describe as *Quantitative Methodological Orientation*, is comprised of five items with a Cronbach's alpha of .69 with an eigenvalue of 2.40, explained 17.11% of the variance; together, the two-factor solution explains 45.66% of the variance in the data (see Table 4 for items and loadings). Both scales range from 1 to 5 with a low score illustrating a low value of that particular characteristic, and a high score illustrating a high value or commitment. The data from the resulting scales suggest evaluators lean towards consistent evaluation processes in their work ($M = 4.13$, $SD = 0.66$), but are neutral regarding their inquiry technique such that a low score indicates a preference for qualitative techniques and a higher score indicates a preference for quantitative techniques ($M = 2.99$, $SD = 0.82$, see Table 4 for all means and standard deviations).

Table 4. Revised Factor Loadings for Christie's (2003) Scale.

Item	Evaluation processes	Quantitative methodological orientation
I believe that during the final stage of the evaluation, an important role of the evaluator is to work with the primary users to help determine the next steps.	.78	-.08
When I conduct an evaluation, if while the evaluation is underway the plan is not working, corrections can be made and new strategies can be identified and implemented.	.74	-.05
When I conduct an evaluation, the primary users help conceptualize and decide upon the evaluation questions.	.70	-.08
When I conduct an evaluation, the primary users help interpret the meaning of the evaluation's data.	.68	-.10
Using my evaluation approach, stakeholders' assumptions about a program are integrated into the evaluation process in order to ensure its relevance and usefulness.	.53	.001
In my evaluation approach, the application of research method(s) in an evaluation is most effective when it is guided by the program's conceptual framework, model, or theory.	.51	.18
I believe the ultimate purpose of program evaluation is to enhance knowledge and strategies for designing and implementing programs.	.50	.30
I believe that evaluation conclusions are a mixture of facts and values.	.48	-.17
My evaluation approach intends to create changes in the culture of the organization (agency) where the evaluation is being conducted.	.45	-.07
My approach as an evaluator is to assess objectively the evaluation's quality against validated technical standards.	.04	.74
When conducting an evaluation, I believe it is critical to understand the main question and then use all scientifically-tested instruments available to answer it.	-.03	.60
I believe that the evaluator's interpretation of the findings and conclusions can be unbiased.	-.21	.57
My primary method choice for answering evaluation questions is quantitative.	-.17	.44
I believe the ultimate purpose of program evaluation is to assess the effectiveness of programs.	.16	.43

Note. $N = 213$. The extraction method was principal axis factoring with an oblique rotation. Factor Loadings above .32 are in bold.

Differences Between Groups

We used a series of tests, strictly for exploratory purposes, to test the average difference in each of the measures based on participant's sex, whether they were a first-generation college student or not, whether they were a graduate student or postgraduate, and whether they were in the United States or International. We were not able to run any comparative analysis based on race or type of degree because each demographic had three or more levels with a large difference in sample sizes in each level (see Table 2). This led to a total of 88 comparison tests (i.e., four demographic variables and 22 dependent variables). Here, we report only significant differences, and it is important to note that any significant differences reported may simply be due to chance when running a large number of comparisons. We emphasize that the following results are strictly for descriptive purposes.

To start, there were no statistically significant differences found for the status of first-generation college graduates versus not first-generation on any of the variables (i.e., 22 non-significant comparisons). For participant's sex, independent-samples *t*-tests revealed differences in three variables (i.e., 13.64% of the variables). First, a small average difference in conscientiousness between men ($M=3.9$, $SD=0.62$) and women ($M=4.12$, $SD=0.54$), $t(199)=2.54$, $p=.01$, $d=0.39$, $M_{Diff}=0.22$, 95% CI [0.05, 0.40]. We also found that females had slightly higher scores ($M=4.28$, $SD=0.50$) on Competence and Growth compared to men ($M=4.11$, $SD=0.56$), $t(198)=2.18$, $p=.03$, $d=0.33$, $M_{Diff}=0.18$, 95% CI [0.02, 0.34]. Lastly, females had slightly lower scores ($M=2.27$, $SD=1.12$) on the political scale compared to men ($M=2.65$, $SD=1.28$), $t(198)=2.11$, $p=.04$, $d=.32$, $M_{Diff}=.38$, 95% CI [0.02, 0.74].

When comparing the United States and international evaluators, only two differences appeared (i.e., 9.09% of the variables). First, international evaluators had higher scores on Conformity ($M=5.28$, $SD=0.97$) than U.S. evaluators ($M=4.80$, $SD=1.10$), $t(205)=2.50$, $p=.01$, $d=0.47$, $M_{Diff}=0.48$, 95% CI [0.10, 0.86]. Second, international evaluators had higher scores on Power ($M=4.38$, $SD=1.08$) than U.S. evaluators ($M=3.60$, $SD=0.99$), $t(205)=4.33$, $p<.001$, $d=0.75$, $M_{Diff}=.78$, 95% CI [0.43, 1.14].

There was a difference between graduate students and postgraduates on five variables (i.e., 22.73% of the variables). First, graduate students had slightly higher scores on Status and Independence ($M=3.76$, $SD=0.72$) than postgraduates ($M=3.48$, $SD=0.58$), $t(207)=2.62$, $p=.009$, $d=0.47$, $M_{Diff}=0.28$, 95% CI [0.07, 0.50]. Second, postgraduates had higher scores on Career Commitment ($M=2.67$, $SD=1.05$) than graduate students ($M=2.20$, $SD=0.87$), $t(204)=2.61$, $p=.01$, $d=0.49$, $M_{Diff}=0.18$, 95% CI [0.12, 0.83]. Third, graduate students had higher scores on Achievement ($M=6.15$, $SD=0.65$) than postgraduates ($M=5.77$, $SD=0.70$), $t(205)=3.11$, $p=.002$, $d=0.57$, $M_{Diff}=0.39$, 95% CI [0.14, 0.63]. Fourth, graduate students had higher scores on Conformity ($M=5.21$, $SD=0.92$) than graduate students ($M=4.82$, $SD=1.11$), $t(205)=2.03$, $p=.04$, $d=0.39$, $M_{Diff}=0.39$, 95% CI [0.01, 0.78]. Lastly, graduate students had higher scores on Power ($M=4.11$, $SD=1.19$) than postgraduates ($M=3.66$, $SD=1.00$), $t(206)=2.42$, $p=.02$, $d=0.41$, $M_{Diff}=0.45$, 95% CI [0.08, 0.81].

As previously stated, the results of these between-group analyses should be interpreted cautiously for a few different reasons. First, only 10 out of the 88 (i.e., 11.36%) of comparisons were significant and likely due to chance. Second, even though the analyses achieved statistical significance, the effect sizes were small, and the confidence intervals of the mean differences approached zero; thus, the practical meaningfulness is minimal. Third, the purpose of comparing the groups was strictly for exploratory and descriptive reasons (i.e., we had no a priori hypotheses for any of the comparisons).

Predicting Evaluation Practice

We analyzed each factor we obtained from Christie's (2003) scale, individually, as the dependent variables in separate analyses. If one of the factors in a scale appeared to be an adequate predictor,

we included it in an overall regression model. None of the variables examined in this study statistically significantly predicted Evaluation Processes. An identical analysis revealed, in the final model, that Competence and Growth (WVI), Status and Independence (WVI), and Benevolence (SVI), Conformity (SVI), and Security (SVI), significantly predicted Quantitative Methodological Orientation, $F(3, 183) = 9.20, p < .001, f^2 = .43, 95\% \text{ CI } [.25, .68]$ (see Table 5). Competence and Growth (WVI) and Benevolence (SVI) negatively predicted Quantitative Methodological Orientation while Status and Independence (WVI), Conformity (SVI), and Security (SVI) had a positive predictive relationship (see Table 5).

Discussion

Values, much like evaluation itself, are a loaded construct. Embedded within those six letters are implicit concepts of rightness and wrongness, decisions and anticipated consequences, worldviews, and more. Values are difficult to measure and interpret, and so we must ask, "What have we learned from this process?" First, the study established that evaluators' values and traits data are normally distributed without significant skew or kurtosis, except for political values which skewed liberal. The data are within the normal range for all constructs, map well onto the theoretical constructs proposed by researchers, and are contained within tight confidence intervals, suggesting a level of homogeneity in the respondents and perhaps in the field itself. Future comparative research will need to explore if these narrow confidence intervals are unique to evaluators or if other professions share this characteristic within their membership.

Second, we learned that evaluation practice or, even practice values, is subtle, complex and that predicting its totality or even its elements is difficult. We were surprised by the lack of predictive power in the regression analysis for Evaluation Practice given the comprehensive nature of the wide range of values typologies and measures included in the study. A key to understanding this may be found in the items themselves in the refined Evaluation Practice measure (see Table 4), which retained questions that address topics such as helping primary users determine next steps, making adjustments to a plan that is not working well, and including primary users in making decisions about the key evaluation questions, as well as including them in interpreting meaning from the data. These are similar to the tenets espoused in many contemporary evaluation approaches, including Utilization-Focused Evaluation (Patton, 2008), Transformative Evaluation (Mertens, 2009), and Empowerment Evaluation (Fetterman & Wandersman, 2005), Participatory Evaluation (Cousins & Earl, 1992), and Culturally Responsive evaluation (Hood et al., 2015). We wonder if this means that there wasn't enough variance in the Evaluation Practice measure for any of the other measured variables to predict it. Further, given the flexibility in contemporary evaluation practice, we could imagine respondents agreeing to the items on the scale while actually implementing them differently in practice. For example, evaluators in the midst of a project undergoing major redirection might find themselves needing to wholly revisit the key evaluation questions and methods, whereas a project undergoing minor change might only see subtle changes related to data collection or analysis. These data should be interpreted cautiously, but could there be much more similarities than differences in how AEA evaluators practice even though their reasons for engaging in these constant practices may indeed differ? Future research will need to delve more deeply into this question.

When it came to predicting Quantitative Methodological Orientation the model that fit the data best included variables with negative predictive weights (Competence and Growth [WVI] and Benevolence [SVI]) as well as variables with positive weights (Status and Independence [WVI], Conformity [SVI], and Security [SVI]), and the β weights suggest that each variable contributes a relatively equal amount to predicting an evaluator's methodological orientation. These data suggest that evaluators who value Competence and Growth (WVI) and Benevolence (SVI) tend

Table 5. Hierarchical Regression Analyses for Revised Christie's Scale (Methodological Technique).

	B	95% CI for B		SE B	β	p	R ²	ΔR^2	df	p (model)
		LL	UL							
Step 1							.05	.04	2,190	.01
Constant	3.11	2.14	4.07	0.49						
Conscientiousness	0.12	-0.08	0.32	0.10	0.09	.22				
Neuroticism	-0.22	-0.39	-0.06	0.08	-0.19	.009				
Step 2							.16	.14	3,187	<.001
Constant	3.58	2.20	4.96	0.70						
Conscientiousness	0.02	-0.18	0.21	0.10	0.01	.87				
Neuroticism	-0.23	-0.39	-0.08	0.08	-0.20	.004				
Comfort and security (WVI)	0.15	0.001	0.31	0.08	0.14	.05				
Competence and growth (WVI)	-0.41	-0.66	-0.16	0.13	-0.24	.001				
Status and independence (WVI)	0.34	0.13	0.54	0.11	0.25	.002				
Step 3							.19	.17	1,186	.005
Constant	2.91	2.53	3.28	0.19						
Conscientiousness	0.02	-0.16	0.21	0.10	0.02	.80				
Neuroticism	-0.21	-0.37	-0.05	0.08	-0.18	.01				
Comfort and security (WVI)	0.16	0.01	0.31	0.08	0.15	.04				
Competence and growth (WVI)	-0.38	-0.62	-0.13	0.12	-0.22	.003				
Status and independence (WVI)	0.29	0.08	0.50	0.11	0.22	.007				
Political values scale	0.13	0.04	0.22	0.05	0.19	.005				
Step 4							.30	.27	3,183	<.001
Constant	2.29	1.74	2.83	0.28						
Conscientiousness	-0.04	-0.21	0.15	0.09	-0.03	.70				
Neuroticism	-0.13	-0.28	0.02	0.08	-0.11	.10				
Comfort and security (WVI)	0.07	-0.08	0.22	0.08	0.07	.35				
Competence and growth (WVI)	-0.33	-0.56	-0.09	0.12	-0.19	.008				
Status and independence (WVI)	0.23	0.03	0.43	0.10	0.17	.02				
Benevolence (Schwartz)	-0.28	-0.48	-0.08	0.10	-0.22	.006				
Conformity (Schwartz)	0.19	0.05	0.32	0.07	0.24	.006				
Security (Schwartz)	0.27	0.09	0.45	0.09	0.25	.004				

Note. Enter method used in regression analysis. CI = confidence interval; LL = lower limit; UL = upper limit.

to have a qualitative orientation, whereas high scores on Status and Independence (WVI), Conformity (SVI), and Security (SVI) are more indicative of a quantitative orientation. Although these data are difficult to interpret in isolation, we are reminded that both Shadish et al. (1991) and Tashakkori and Teddlie (1998) argued that methodological orientations do not exist in isolation, but result from ontological and epistemological intersections. Indeed, these make sense in light of both Schwartz's (1994) taxonomy of values and Meyer et al.'s (1998) perspective on work values given that conformity, security, and status/independence principles broadly align with positivist/post-positivist/normative perspectives, whereas benevolence and growth broadly

align with a more constructivist/expansionist/humanistic orientations (Nilsson & Strupp-Levitsky, 2016). That is, the Schwartz and Meyer et al., values that align with the perception of the world as static and with a mechanical/transactional perspective aligned with higher scores on the quantitative methodology orientation scale and its attendant designs and tools (i.e., there is an objective truth, and I can measure it, albeit imperfectly). Conversely, values that are synonymous with trust, development, and multiple interpretations of reality aligned and its attendant tools and techniques aligned lower scores on the quantitative methodological orientation scale (i.e., there are multiple interpretations of even a single event, and to understand another person's truthful interpretation, one must use multiple interpretative approaches). However, we also note that the average score on quantitative methodological orientation was 2.99 with a standard deviation of 0.82, which we take to mean that most evaluators would identify as pragmatists or mixed-methodologists (Tashakkori & Teddlie, 1998), which is in keeping with a practice that is adaptive and responsive to stakeholder questions.

These findings about evaluation practice and methodological orientation need to be unpacked and clarified through future research. An important consideration is that even though these findings are interesting and intuitive, they are only an initial step in trying to empirically link evaluators' values to practice. It is possible there are unmeasured or untested mediators (e.g., efficacy, outcome expectations, stakeholder trust) and moderators (e.g., evaluation context, resources, evaluator demographics) that could help clarify the proposed values-practice relationship, and we hope this is a relationship that future researchers will explore.

As measured by the Big 5, evaluators are relatively conscientious, agreeable, and open to new experiences. These data are consistent with the major characteristics of evaluation practice. For example, Russ-Eft and Preskill (2009) described evaluation as a profession that demands great attention to detail, and many evaluators would likely agree that being agreeable is a part of the "personal factor" that is so critical to evaluation practice (Patton, 2008). Low average scores on the neuroticism construct support this vision of evaluation because evaluators need to be able to tolerate anxiety, fear, and other negative feelings, even though having these feelings is common in both evaluators and in the general population (LaVelle, Jones, & Donaldson, 2018). We wonder if this is because of the fluid and often-changing nature of both internal and external evaluation practice, and individuals that struggle with managing neurotic tendencies might self-elect towards other professional tracks. We are intrigued by the high agreeableness score, which is made of traits such as friendliness and compassion, and may indeed support previous discussions about the importance of the "personal factor" in evaluation practice. Further research might explore this finding and the ways in which it is expressed in practice.

In terms of political values, the data paint a picture consistent with what the authors imagine an evaluator might look like. That is, evaluators, work with programs, policies, and interventions that seek to change individuals, groups, and organizations (LaVelle & Dighe, 2020), and supporting programs that encourage change is a stereotypically liberal perspective. The respondents felt more conservative about their economic-political outlook, although they were still quite liberal in their fiscal outlook, which supports the position that evaluation is rooted in both social programming and accountability (Christie & Alkin, 2013). Put together, we make sense of evaluators' political worldview by suggesting that they are supportive of social programs and policies, but that they want to know that they are effective. It is not clear what would happen if program stakeholders hold similar views about social or fiscal issues, or the degree to which such congruence is necessary for effective evaluator-stakeholder relationships, but hope future research can explore this hypothetical intersection.

The data gathered from evaluators suggest that the most important aspect of their workplace is competence and opportunities for growth. This is consistent with the idea that evaluators, similar to other fields dependent on systematic inquiry, value challenges and refining their skills. We

interpret this to support Skolits, Morrow, and Burr's (2009) assertion that an attractive facet of evaluation to evaluators is that the work is often varied and that evaluators must act in many different roles throughout the engagement. Much lower in importance is Status and Independence, as well as Security and Comfort. In tandem, these paint a picture of the evaluator as a person who mainly values the challenge of evaluation practice instead of the independence and job security it offers. We are intrigued about the factors and experiences that influence evaluators and their work, such as the varied influences of individual workplaces, community engagement, and association-wide stances on sociopolitical movements (e.g., see Donaldson, 2015, for discussion on the AEA Statement and the Not-AEA Statement on credible evidence).

Finally, we come to the Schwartz Values Inventory, which provides illustrative data on the overarching views of evaluators. In some ways, the results from the SVI are not surprising because they are consistent with the data from the other measures in this study. Evaluators largely value individual achievement and self-direction, which agrees with the highly technical and procedural aspects of evaluation practice. They have a strong benevolent worldview that likely translates into treating people and the environment with kindness and respect. As a group, evaluators place much less value on power over others, tradition, and stimulation, which agree with the ever-evolving nature of evaluation practice and the call to "give evaluation away." Indeed, we believe that a review of the Presidential Keynotes from associations such as the American Evaluation Association, Canadian Evaluation Society, European Evaluation Society, African Evaluation Association, Australasian Evaluation Society, and others would find echoes of these values in the speakers' views for evaluation into the future.

Although in need of further exploration, the data also may be suggestive of the origins of some frustrations that evaluators have experienced in practice (Hutchinson, 2019) perhaps because their worldviews may have been difficult to reconcile with those of their stakeholders or colleagues. For example, an evaluator who is guided by benevolence/self-direction/universalism, working diligently to build the capacity of a stakeholder team might indeed be frustrated with individual and organizational conformity and tradition. This suggests the importance of early discussions with stakeholders about their plans, goals, processes, and of course, their readiness for learning and evaluation (e.g., Russ-Eft & Preskill, 2009), evaluative thinking (Buckley, Archibald, Hargraves, & Trochim, 2015), anticipated form(s) of use (Patton, 2008) and conceptual program development (Donaldson, 2007), and other desirable outcomes from the evaluation process so that the evaluators' values and stakeholders' values align as well as their expectations.

The results of this study highlight the need for early discussions about the differences between monitoring and evaluation, which draw from a similar skill set, but are expressed very differently in practice, and are likely guided by different value positions. Further research might explore evaluators' value alignment with their preferred evaluative frameworks such as theory-driven evaluation science, culturally responsive evaluation, participatory evaluation, realist evaluation, building evaluation capacity, and the like. Shadish (1998, p. 1) is credited with the phrase "evaluation theory is who we are," and though we largely agree, we respectfully suggest that evaluative values and theory could be aspirational too: *evaluation theory is who we aspire to be*.

Strengths and Limitations

This study represents the first of its kind, and while we are pleased with the process and the results, there were some inherent limitations in the study. First, the study relied upon an online-only data collection procedure coupled with the ever-present possibility that respondents' social desirability inflated their scores on some items and factors (e.g., Benevolence) and minimized their scores on others (e.g., Power). It is also possible that some of the survey items challenged respondents' self-perception of their personality traits (e.g., Neuroticism), or political ideology. Although the 20.8%

usable response rate is within an acceptable range given the number of study invitations distributed (Dillman, Smyth, & Christian, 2014), we would have liked a stronger response rate. We note as well that the respondents were mainly White, and because the number of respondents from other demographics groups was small, we believe further studies should look specifically at the values of all demographic and identity groups so that their voices are not whitewashed in the research process. Further, given the nature of the incentive (\$5 donation to the AEA Student Travel Fund per completed survey), it is possible that respondents with a particular value set participated at a higher rate than people that did not see it as compelling compensation. Last, although the study used empirically tested scales and their attendant response categories, it is possible that alternate measures of these constructs would yield slightly different results. Furthermore, it is possible that the tight confidence intervals on both the predictor and dependent variables may have constrained the size of the observed relationships in the regression analyses. These limitations are offset, however, by the quality of the data itself.

The data for each of the constructs closely align to the conceptual models proposed by the original researchers, and with the exception of political values, the data are normally distributed with few outliers. Moreover, the data seem trustworthy because respondents' demographics generally align with the membership profile of AEA and because each of the measures was presented randomly to reduce the possibility of ordering effects. However, because the study looked exclusively at AEA evaluators, we do not know how these data compare non-AEA evaluators, or with members of other professions. Indeed, many practicing evaluators in the United States and across the globe are not members of AEA, and it is possible that non-AEA evaluators' values and practice will be different from the people that self-select into being AEA members. Further research will be needed to compare evaluators' scores against other populations' data, as well as that of the general population. These future studies will be especially telling in interpreting the data regarding educational attainment and values, given that the respondents with baccalaureate degrees had still self-selected to be members of AEA.

One construct still worries us, however, and that is the measure of evaluation practice itself. While we appreciate the insights offered by Christie's (2003) original and refined measures, they do not seem like a comprehensive measure of evaluation practice. Understanding evaluation processes and methodological orientations are essential for understanding evaluation, of course, but deeper questions are why and how those perspectives shape practice, as well as the degrees of actual variance in evaluation practice. Using Christie's studies and measure alongside the current study, we envision other scholars and practitioners of evaluation discussing the depth and nuances of their practice, and finding new and innovative ways of measuring evaluation practice. A particularly fruitful research stream might even go so far as to develop tools to measure the fidelity of evaluation practice compared with prescribed theory (evaluation activities) and link them with some of the desired outcomes of evaluation practice (Mark, 2008), such as utilization, empowerment, and social justice. We hope this study and other lines of inquiry will help evaluation move more fully from a prescriptive discipline onto one that is descriptive of its values and practice, predictive of the outcomes of that practice on the individuals and communities we serve, and be both aspirational and inspirational to current and future evaluators.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported by a grant from the Office of the Vice President for Research, University of Minnesota.

ORCID iD

John M. LaVelle  <https://orcid.org/0000-0002-8513-5232>

Note

1. Extraversion is sometimes spelled “extroversion,” though the terms are identical in meaning.

References

- Azzam, T. (2010). Evaluator responsiveness to stakeholders. *American Journal of Evaluation*, 31(1), 45–65. <https://doi.org/10.1177/1098214009354917>
- Azzam, T., & Levine, B. (2014). Negotiating truth, beauty, and justice: A politically responsive approach. In J. C. Griffith & B. Montrosse-Moorhead (Eds.), *Revisiting truth, beauty, and justice: Evaluating with validity in the 21st century* (Vol. 142, pp. 57–70). New Directions for Evaluation. <https://doi.org/10.1002/ev.20085>
- Bardi, A., & Schwartz, S. (1996). Relations among sociopolitical values in Eastern Europe: Effects of the communist experience? *Political Psychology*, 17(3), 525–549. <https://doi.org/10.2307/3791967>
- Barni, D., Ranieri, S., Scabini, E., & Rosnati, E. (2011). Value transmission in the family: Do adolescents accept the values their parents want to transmit? *Journal of Moral Education*, 40(1), 105–121. <https://doi.org/10.1080/03057240.2011.553797>
- Berings, D., De Fruyt, F., & Bouwen, R. (2004). Work values and personality traits as predictors of enterprising and social vocational interests. *Personality and Individual Differences*, 36, 349–364.
- Brown, D. (2002). The role of work and cultural values in occupational choice, satisfaction, and success: A theoretical statement. *Journal of Counseling & Development*, 80(1), 48–56. <https://doi.org/10.1002/j.1556-6678.2002.tb00165.x>
- Buckley, J., Archibald, T., Hargraves, M., & Trochim, W. M. (2015). Defining and teaching evaluative thinking: Insights from research on critical thinking. *American Journal of Evaluation*, 36(3), 375–388. <https://doi.org/10.1177/1098214015581706>
- Christie, C. A. (2003). What guides evaluation? A study of how practice maps onto evaluation theory. *New Directions for Evaluation*, 2003(97), 7–35. <https://doi.org/10.1002/ev.72>
- Christie, C. A., & Alkin, M. C. (2013). An evaluation theory tree. In M. C. Alkin (Ed.), *Evaluation roots: A wider perspective of theorists views and influences* (pp. 12–65). Thousand Oaks, CA: Sage.
- Church, A. H., Burke, W. W., & Van Eynde, D. F. (1994). Values, motives, and interventions of organization development practitioners. *Group & Organization Management*, 19(1), 5–50. <https://doi.org/10.1177/1059601194191002>
- Costa, P. T.Jr., & McCrae, R. R. (1992). Four ways five factors are basic. *Personality and Individual Differences*, 13(6), 653–665. [https://doi.org/10.1016/0191-8869\(92\)90236-I](https://doi.org/10.1016/0191-8869(92)90236-I)
- Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, 10(7), 1–9. <https://doi.org/10.7275/jyj1-4868>
- Cousins, J. B., & Earl, L. M. (1992). The case for participatory evaluation. *Educational Evaluation and Policy Analysis*, 14(4), 397–418. <https://doi.org/10.3102/01623737014004397>
- Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, 25(1), 7–29. <https://doi.org/10.1177/0956797613504966>
- Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist*, 60(2), 170–180. <https://doi.org/10.1037/0003-066X.60.2.170>
- Dawis, R. V., & Lofquist, L. H. (1984). *A psychological theory of work adjustment: An individual differences model and its applications*. Minneapolis: University of Minnesota Press.
- DePoy, E., & Merrill, S. C. (1988). Value acquisition in an occupational therapy curriculum. *The Occupational Therapy Journal of Research*, 8(5), 259–274. <https://doi.org/10.1177/153944928800800501>
- Devellis, R. F. (2012). *Scale Development* (3rd ed.). Sage.

- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed mode surveys: The tailored design method* (4th ed.). Hoboken, NJ: John Wiley & Sons.
- Donaldson, S. I. (2007). *Program theory-driven evaluation science: Strategies and applications*. Mahwah, NJ: Lawrence Erlbaum.
- Donaldson, S. I. (2015). Examining the backbone of contemporary evaluation practice. In S. I. Donaldson, C. A. Christie, & M. M. Mark (Eds.), *Credible and actionable evidence: The foundation for rigorous and influential evaluations* (2nd ed., pp. 3–26). Thousand Oaks, CA: Sage.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analysis using GPower 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. <https://doi.org/10.3758/BRM.4.4.1149>
- Fetterman, D. M., & Wandersman, A. (2005). *Empowerment evaluation principles in practice*. New York, NY: The Guilford Press.
- Fields, D. L. (2013). *Taking the measure of work: A guide to validated scales for organizational research and diagnosis*. Charlotte, NC: Information Age Publishing.
- George, D., & Mallery, P. (2008). *SPSS For windows: Step by step* (8th ed.). Pearson.
- Greene, J. (1997). Evaluation as advocacy. *The American Journal of Evaluation, 18*(1), 25–35. [https://doi.org/10.1016/S0886-1633\(97\)9005-2](https://doi.org/10.1016/S0886-1633(97)9005-2)
- Greene, J. C. (2005). A value-engaged approach for evaluating the Bunche–Da Vinci learning academy. *New Directions for Evaluation, 2005*(106), 27–45. <https://doi-org.ezp3.lib.umn.edu/10.1002/ev.150>
- Gullickson, A. M., & Hannum, K. M. (2019). Making values explicit in evaluation practice. *Evaluation Journal of Australasia, 19*(4), 162–178. <https://doi.org/10.1177/1035719X19893892>
- Hood, S., Hopson, R., & Frierson, H. (Eds.) (2015). *Continuing the journey to reposition culture and cultural context in evaluation theory and practice*. Information Age Publishing.
- House, E. R. (2015). *Evaluating: Values, biases, and practical wisdom*. Information Age Publishing.
- Hutchinson, K. (Ed.) (2019). *Evaluation failures: 22 tales of mistakes made and lessons learned*. Sage.
- John, O. P., & Srivastava, S. (1999). The Big 5 trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102–138). Guilford.
- Julnes, G. (2012). Managing valuation. *New Directions for Evaluation, 2012*(133), 3–15. <https://doi.org/10.1002/ev.20002>
- Kawakami, A. J., Aton, K., Cram, F., Lai, M., & Porima, L. (2008). Improving the practice of evaluation through indigenous values and methods: Returning the gaze from Hawai'i and Aotearoa. In N. L. Smith & P. Brandon (Eds.), *Fundamental issues in evaluation* (pp. 219–239). The Guilford Press.
- Kim, M. (2008). Spiritual values, religious practices, and democratic attitudes. *Politics and Religion, 1*(2), 216–236. <https://doi.org/10.1017/S1755048308000187>
- Kirkhart, K. E. (2010). Eyes on the prize: Multicultural validity and evaluation theory. *American Journal of Evaluation, 31*(3), 400–413. <https://doi.org/10.1177/1098214010373645>
- LaVelle, J. M., & Dighe, S. (2020). A transdisciplinary model of program outcomes for enhanced evaluation practice. *Canadian Journal of Program Evaluation, 35*(1), 20–34. <https://doi.org/10.3138/cjpe.61660>
- LaVelle, J. M., Jones, N., & Donaldson, S. (2018). Imposter phenomenon in evaluators. Paper presented at the Annual Conference of the American Evaluation Association. Cleveland, OH.
- Leuty, M. E., & Hansen, J. C. (2011). Evidence of construct validity for work values. *Journal of Vocational Behavior, 79*, 379–390.
- Lindsay, D. S. (2015). Replication in psychological science. *Psychological Science, 26*(12), 1827–1832. <https://doi.org/10.1177/0956797615616374>
- Malsch, A. M. (2005). Prosocial behavior beyond borders: Psychological sense of global community [Unpublished doctoral dissertation]. Claremont Graduate University. <https://doi.org/10.1037/e628612012-025>

- Manhardt, P. J. (1972). Job orientation of male and female college graduates in business. *Personnel Psychology*, 25, 361–368.
- Mark, M. (2008). Building a better evidence base for evaluation theory: Beyond general calls to a framework of types of research on evaluation. In N. L. Smith & P. Brandon (Eds.), *Fundamental issues in evaluation* (pp. 111–134). The Guilford Press.
- Mathieu, J. E., & Zajac, D. M. (1990). A review and meta-analysis of the antecedents, correlates, and consequences of organizational commitment. *Psychological Bulletin*, 108(2), 171–194. <https://doi.org/10.1037/0033-2909.108.2.171>
- McCrae, R. R., & Costa, P. T. Jr. (1987). Validation of a five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, 52(1), 81–90. <https://doi.org/10.1037/0022-3514.52.1.81>
- McCrae, R. R., & Costa, P. T. Jr. (1997). Personality trait structure as a human universal. *American Psychologist*, 52(5), 509–516. <https://doi.org/10.1037/0003-006X.52.5.509>
- McCrae, R. R., & Costa, P. T. Jr. (1999). A five-factor theory of personality. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 139–153). The Guilford Press.
- Mertens, D. M. (2009). *Transformative research and evaluation*. New York: Guilford.
- Meyer, J. P., Irving, P. G., & Allen, N. J. (1998). Examination of the combined effects of work values and early work experiences. *Journal of Organizational Behavior*, 19(1), 29–52. [https://doi.org/10.1002/\(SICI\)1099-1379\(199801\)19:1<29::AID-JOB818>3.0.CO;2-U](https://doi.org/10.1002/(SICI)1099-1379(199801)19:1<29::AID-JOB818>3.0.CO;2-U)
- Nilsson, A., & Strupp-Levitsky, M. (2016). Humanistic and normativistic metaphysics, epistemology, and conative orientation: Two fundamental systems of meaning. *Personality and Individual Differences*, 100, 85–94. <https://doi-org.ezp3.lib.umn.edu/10.1016/j.paid.2016.01.050>
- Palumbo, D. J. (Ed). (1987). *The politics of program evaluation*. Sage.
- Parks-Leduc, L., Feldman, G., & Bardi, A. (2015). Personality traits and personal values: A meta-analysis. *Personality and Social Psychology Review*, 19(1), 3–29. <https://doi.org/10.1177/1088868314538548>
- Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Sage.
- Patton, M. Q. (2013). Utilization-focused evaluation checklist. https://wmich.edu/sites/default/files/attachments/u350/2014/UFE_checklist_2013.pdf
- Roccas, S., Sagiv, L., Schwartz, S. H., & Knafo, A. (2002). The Big five personality factors and personal values. *Personality and Social Psychology Bulletin*, 28(6), 789–801. <https://doi.org/10.1177/0146167202289008>
- Rounds, J. B., & Armstrong, P. I. (2005). Assessment of needs and values. In S. D. Brown & R. W. Lent (Eds.), *Career development and counseling: Putting theory and research to work* (pp. 305–329). John Wiley and Sons, Inc.
- Russ-Eft, D., & Preskill, H. (2009). *Evaluation in organizations: A systematic approach to enhancing learning, performance, and change* (2nd ed.). New York: Basic Books.
- Schwandt, T. A. (1997). The landscape of values in evaluation: Charted terrain and unexplored territory. *New Directions for Evaluation*, 1997(76), 25–39. <https://doi.org/10.1002/ev.1085>
- Schwandt, T. A. (2008). Educating for intelligent belief in evaluation. *American Journal of Evaluation*, 29(2), 139–150. <https://doi.org/10.1177/1098214008316889>
- Schwartz, S. H. (1994). Are there universal aspects in the content and structure of values? *Journal of Social Issues*, 50(4), 19–45. <https://doi.org/10.1111/j.1540-4560.1994.tb01196.x>
- Schwartz, S. H., Cieciuch, J., Vecchione, M., Davidov, E., Fischer, R., Beierlein, C., Ramos, A., Verkasalo, M., Lönnqvist, J. E., Demirutku, K., Dirilen-Gumus, O., & Konty, M. (2012). Refining the theory of basic individual values. *Journal of Personality and Social Psychology*, 103(4), 663–688. <https://doi.org/10.1037/a0029393>
- Schwartz, S. H., & Rubel-Lifschitz, T. (2009). Cross-national variation in the size of sex differences in values: Effects of gender equality. *Journal of Personality and Social Psychology*, 97(1), 171–185. <https://doi.org/10.1037/a0015546>
- Scriven, M. (1991). *Evaluation Thesaurus* (4th Ed.). Sage.

- Scriven, M. (2007). Key evaluation checklist. http://michaelscriven.info/images/KEC_7.25.2013.pdf
- Shadish, W. R. (1998). Evaluation theory is who we are. *American Journal of Evaluation*, 19(1) 1–19. [https://doi.org/10.1016/S1098-2140\(99\)80177-5](https://doi.org/10.1016/S1098-2140(99)80177-5)
- Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). *Foundations of evaluation: Theories of practice*. Sage (Atlanta, GA).
- Skolits, G. J., Morrow, J. A., & Burr, E. M. (2009). Reconceptualizing evaluator roles. *American Journal of Evaluation*, 30(3), 275–295. <https://doi.org/10.1177/1098214009338872>
- Smith, N. L. (1980). Sources of values influencing educational evaluation. *Studies in Educational Evaluation*, 6(2), 101–118. [https://doi.org/10.1016/0191-49X\(80\)90013-9](https://doi.org/10.1016/0191-49X(80)90013-9)
- Smith, N. L. (1981). Evaluation design as preserving valued qualities in evaluation studies. *Studies in Educational Evaluation*, 7(3), 229–237. [https://doi.org/10.1016/0191-491X\(81\)90001-8](https://doi.org/10.1016/0191-491X(81)90001-8)
- Smith, N. L. (1995). The influence of societal games on the methodology of evaluative inquiry. In D. M. Fournier (Ed.), *Reasoning in evaluation: Inferential links and leaps* (pp. 5–14). Jossey-Bass Publications.
- Smith, N. L. (1998). Professional reasons for declining an evaluation contract. *American Journal of Evaluation*, 19(2), 177–190. [https://doi.org/10.1016/S10982140\(99\)80193-3](https://doi.org/10.1016/S10982140(99)80193-3)
- Smith, N. L. (2007). Empowerment evaluation as evaluation ideology. *American Journal of Evaluation*, 28(2), 169–178. <https://doi.org/10.1177/1098214006294722>
- Smith, N. L., Chircop, S., & Mukherjee, P. (1993). Considerations on the development of culturally relevant evaluation standards. *Studies in Educational Evaluation*, 19(1), 3–13. [https://doi.org/10.1016/S0191-49X\(05\)80051-3](https://doi.org/10.1016/S0191-49X(05)80051-3)
- Stern, P. C., Dietz, T., & Guagnano, G. A. (1998). A brief inventory of values. *Educational and Psychological Measurement*, 58(6), 984–1001. <https://doi.org/10.1177/0013164498058006008>
- Taber, C. S., Cann, D., & Kucsova, S. (2009). The motivated processing of political arguments. *Political Behavior*, 31(2), 137–155. <https://doi.org/10.1007/s11109-008-9075-8>
- Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches*. Thousand Oaks, CA: Sage.
- Weiss, C. H. (1987). Where politics and evaluation meet. In D. J. Palumbo (Ed.), *The politics of program evaluation* (pp. 47–71). Sage.
- Weiss, C. H. (1998). *Evaluation* (2nd ed.). Pearson.
- Wetherill, R. R., Neal, D. J., & Fromme, K. (2010). Parents, peers, and sexual values influence sexual behavior during the transition to college. *Archives of Sexual Behavior*, 39(3), 682–694. <https://doi.org/10.1007/s10508-009-9476-8>
- Winter, D. G., John, O. P., Stewart, A. J., Klohnen, E. C., & Duncan, L. E. (1998). Traits and motives: Toward and integration of two traditions in personality research. *Psychological Review*, 105(2), 230–250. <https://doi.org/10.1037/0033-295X.105.2.230>
- Yong, A., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology*, 9, 79–94.
- Zillig, L. M. P., Hemenover, S. H., & Dienstbier, R. A. (2002). What do we assess when we assess a Big 5 trait? A content analysis of the affective, behavioral, and cognitive processes represented in Big 5 personality inventories. *Personality and Social Psychology Bulletin*, 28(6), 847–858. <https://doi.org/10.1177/0146167202289013>