Evaluation and Program Planning 76 (2019) 101681

Contents lists available at ScienceDirect

Evaluation and Program Planning

journal homepage: www.elsevier.com/locate/evalprogplan

Evaluation logic in practice
Findings from two empirical investigations of American Evaluation Association members

Satoshi Ozeki[a], Chris L.S. Coryn[b,*], Daniela C. Schröter[b]

[a] Miyazaki International College, Miyazaki, Japan
[b] Western Michigan University, Kalamazoo, MI, United States

ARTICLE INFO

Keywords:
Logic of evaluation
Evaluation logic
American Evaluation Association
Evaluation practice
Research on evaluation

ABSTRACT

Surveys of two independent random samples of American Evaluation Association (AEA) members were conducted to investigate application of the logic of evaluation in their evaluation practice. This logic consists of four parts: (1) establish criteria, (2) set standards, (3) measure performance on criteria and compare to standards, and (4) synthesize into a value judgment. Nearly three-fourths (71.84% ± 5.98%) of AEA members are unfamiliar with this logic, yet a majority also indicate its importance and utility for evaluation practice. Moreover, and despite unfamiliarity with the four steps of the logic of evaluation, many AEA members identify evaluative criteria (82.41% ± 3.34%), set performance standards (60.55% ± 7.39%), compare performance to standards (62.14% ± 5.98%), and synthesize into an evaluative conclusion (75.00% ± 5.80%) in their evaluation practice. Much like the working logic of evaluation, however, application of the general logic varies widely.

1. Background and introduction

Cook (1997) once lamented "I cringe inside when I talk to evaluators who haven't read his writings. Not to know Scriven is to be illiterate in evaluation…" (p. 44). In part, his remarks originate from the premise that Scriven's logic of evaluation is, perhaps, the only approximation of a 'meta-theory' of evaluation (Shadish, Cook, & Leviton, 1991). Scriven (e.g., 1967, 1980, 1986, 1991, 2007, 2012, 2013) has written extensively about the logic of evaluation. Specifically, this logic is "…concerned with (i) how, if at all, professional evaluation is possible; (ii) its nature and its location in the organization of knowledge, and (iii) the logical structure of its inferences" (Scriven et al., 2007, p. 1). This logic of evaluation delineates, defines, and describes the fundamental, underlying logical structure and pattern of reasoning associated with evaluative inquiry. In the absence of such a logic "…circumscribing the inquiry process, there is only a loose set of activities" (Fournier, 1995, p. 16). At its essence, this logic is concerned with value claims and could be considered the foundation of all evaluation practice (Schwandt, 2015). Collectively, four procedures encapsulate the general logic of evaluation: (1) establish criteria, (2) set standards, (3) measure performance on criteria and compare to standards, and (4) synthesize into a value judgment.

In Scriven's logic of evaluation (1967, 1980, 1986, 1991, 2007, 2012, 2013), criteria are essentially the traits, characteristics, or properties that define or exemplify a good, valuable, or important evaluand (that by which merit, worth, or significance are determined). Criteria are often latent and cannot be directly observed (that is, measured) and often are used by Scriven and others to reflect values (see Davidson, 2005; House & Howe, 1999; Schwandt, 2015; Stufflebeam & Coryn, 2014). Synonyms sometimes used by evaluators to describe criteria include indicators, measures, or variables (though these are semantically different from criteria). In the evaluation of university faculty, for instance, the criteria often applied are research, teaching, and service. Indicators of the research criterion might include number of publications, impact factors of journals in which publications appear, citations to publications, external research funding obtained, and so forth. Standards are cut scores (one or more) against which performance on a criterion are compared to, with cut scores representing two or more discrete states, conditions, or degrees of performance. In the case where only a single cut score is set, its application results in the creation of only two possible performance categories such as pass or fail, met or not met. The creation of multiple cut scores results in more than two performance categories such as the method used in some grading systems (e.g., A, B, C, D, or F) or ordered classification systems (e.g., very poor, poor, average, good, very good, excellent). Methods for creating standards are typically either norm-referenced

* Corresponding author.
E-mail address: chris.coryn@wmich.edu (C.L.S. Coryn).

https://doi.org/10.1016/j.evalprogplan.2019.101681
Received 19 March 2019; Received in revised form 8 June 2019; Accepted 16 July 2019
Available online 20 July 2019

(relative performance, in comparison to something else) or criterion-referenced (absolute performance, not compared to something else). Often, standards are used synonymously with targets (criterion-referenced) or external benchmarks (norm-referenced). Synthesis is the process of integrating measures of performance (which can be quantitative, qualitative, or both) against standards (fact-value synthesis) or performances across multiple criteria (value-value synthesis). In recent years, rubrics have been endorsed as a means for synthesis (King, McKegg, Oakden, & Wehipeihana, 2013). And, many times, the synthesis process involves differential weighting of criteria when performing a synthesis across multiple criteria.

In practice, variations on the general logic of evaluation are referred to as working logics (Fournier, 1995). Working logics are expressed as (1) problem, (2) phenomena, (3) question, and (4) claim, which are somewhat analogous to the four steps of the general logic. Moreover, as Fournier (1995) notes, "…because evaluation practice varies widely along these four parameters, evaluation practice can be viewed as a proliferation of individual working logics" (p. 19). No matter the evaluation approach used (e.g., theory-driven, objectives-based, utilization-focused [see Stufflebeam & Coryn, 2014]), object of inquiry (e.g., program, policy, product), purpose (e.g., formative, summative), role of the evaluator (e.g., internal, external), degree of stakeholder involvement (see Cullen, Coryn, & Rugh, 2011), or methodological or ideological orientation, for example, all implicitly apply this logical framework (Shadish et al., 1991).

According to findings reported by Christie (2003) and Shadish and Epstein (1987), very few practicing evaluators use evaluation theory to guide their practice. Similarly, the extent to which practicing evaluators have knowledge of and intentionally apply the general logic of evaluation—or one or more discernable working logics—in their practice is unknown and has not been subjected to empirical scrutiny. Although central to professional, disciplined evaluation, evaluation logic and related concepts have not been endorsed or recognized as a competency for evaluation practitioners in North America (American Evaluation Association (AEA), (2018); Canadian Evaluation Society (CES), (2018)). Relatedly, working knowledge of the logic of evaluation is likely not required as part of most formal evaluation education, training, and preparation (Christie, Quiñones, & Fierro, 2014; Davies & MacKay, 2014; LaVelle, 2011, 2014, 2018; LaVelle & Donaldson, 2010, 2015).

Ostensibly, knowledge and application, or lack thereof, of evaluation logic and reasoning by practicing evaluators is of sufficient importance to warrant empirical investigation, and thus the impetus for this study. More generally, the investigation is positioned in and in response to the increasing appeals for empirical investigations into evaluation theory and practice (e.g., Christie, 2003, 2011; Coryn et al., 2016, 2017, 2019; Henry & Mark, 2003; Mark, 2008; Stufflebeam & Coryn, 2014; Szanyi, Azzam, & Galen, 2013).

2. Study objectives and questions investigated

Two independent studies were conducted to investigate AEA members' knowledge, perceptions, and applications of the logic of evaluation in their evaluation practice. AEA's mission is to improve evaluation practices and methods, increase evaluation use, promote evaluation as a profession, and support the contribution of evaluation to the generation of theory and knowledge about effective human action. As of 2019, AEA had more than 7000 members representing all 50 states in the United States as well as members from more than 80 foreign countries.

2.1. Study 1

The focal questions investigated as part of Study 1 were:

1 To what extent do evaluators apply the logic of evaluation in practice?
   a How frequently do evaluators perform each step of the logic of evaluation?
   b How difficult it is to perform each step of the logic of evaluation?
2 What are evaluators' perceptions of the logic of evaluation?
   a How familiar are evaluators with the logic of evaluation?
   b How useful do evaluators consider the logic of evaluation?
   c How important do evaluators consider the logic of evaluation?

2.2. Study 2

The focal questions investigated as part of Study 2 were:

1 How do evaluators establish performance standards?
   a Do they use standards to formulate evaluative conclusions?
   b Who is involved in establishing performance standards?
   c When are performance standards established?
2 How do evaluators formulate evaluative conclusions?
   a How frequently do evaluators formulate evaluative conclusions?
   b Who is responsible for formulating evaluative conclusions?
   c What are the challenges in formulating evaluative conclusions?

3. Method

3.1. Design

A cross-sectional design was used to investigate and address the focal research questions associated with both Study 1 and Study 2. More specifically, the design consisted of two independent, simple random sample surveys of AEA members. Two independent samples were taken in order to reduce the burden of response on individual AEA members given the length of each survey questionnaire.

3.2. Samples

The member list of AEA was obtained following an application procedure with and approval of the AEA Research Request Task Force. The list consisted of a total of $N = 7231$ individuals who were members of AEA as of March 2016. Two independent simple random samples were drawn from the list (i.e., sampling frame). With a bound on the error of estimation of ± 5% and assuming a population proportion of $p = 00.50$, an estimated simple random sample of size $n = 365$ was necessary for Study 1. To accommodate for potential nonresponse and maintain precision, a 20% oversample ($n = 73$) was taken, resulting in a total sample size of $n = 438$ for Study 1. So as not to potentially duplicate sample elements, the $n = 438$ elements drawn for Study 1 were excluded from the sampling frame for Study 2, leaving $N = 6793$. For Study 2, with a bound on the error of estimation of ± 5% and assuming a population proportion of $p = 00.50$, a simple random sample of size $n = 364$ was estimated. To accommodate for potential nonresponse and maintain precision, a 20% oversample ($n = 73$) was taken, resulting in a total sample size of $n = 437$ for Study 2. Undeliverable e-mail addresses and requests to opt-out resulted in final usable samples of $n = 434$ and $n = 430$ for Study 1 and Study 2, respectively.

For Study 1, a sample of $n = 108$ was obtained resulting in a response rate of 24.88%. For Study 2, a sample of $n = 106$ was obtained for a response rate of 24.65%. The obtained sample sizes are consistent with those typically obtained from AEA members (i.e., approximately 25% on average, Coryn et al., 2019). As shown in Table 1, the samples obtained for both Study 1 and Study 2 are generally congruent with the characteristics of all AEA members. Most respondents (52.04% for Study 1 and 44.68% for Study 2) indicated that they are predominately external evaluators with an average of 10.18 ($SD = 7.36$) and 12.45 ($SD = 9.54$) years of experience conducting evaluations, for Study 1 and Study 2, respectively.

Table 1
AEA member characteristics and study 1 and 2 sample characteristics.

Characteristic[a]	AEA member population (N = 7026)[b]	Study 1 sample (n = 108)	Study 2 sample (n = 106)
Gender			
Male	26.27%	26.53%	27.66%
Female	64.56%	70.41%	69.15%
Prefer not to answer		3.06%	3.19%
Highest level of education			
Doctorate	41.52%	51.02%	46.24%
Masters	41.94%	43.88%	48.39%
Bachelors	05.61%	4.08%	4.30%
Other	00.85%	1.02%	1.08%
Country			
United States	80.03%	84.69%	76.34%
Other	14.86%	15.31%	23.66%
Primary work setting			
College/University	30.84%	22.45%	31.91%
Non-profit organization	21.02%	21.43%	21.28%
Private business	20.35%	26.53%	25.53%
Federal agency	05.31%	8.16%	6.38%
State agency	03.19%	7.14%	2.13%
Local agency	02.08%	9.18%	5.32%
School system	02.32%	3.06%	4.26%
Other	06.12%	2.04%	2.13%
Role as evaluator			
External		52.04%	44.68%
Internal		20.41%	28.72%
Mix of both		27.55%	26.66%
Number of years conducting evaluation (M, SD)		10.18 (7.36)	12.45 (9.54)

[a] Within subgroups, characteristic percentages do not always total 100% due to item nonresponse/missingness and/or rounding error.
[b] Estimates from Coryn et al. (2016).

3.3. Instrumentation

For Study 1, the survey questionnaire consisted of 39 closed-response, free-response, and demographic questions. For Study 2, the survey questionnaire consisted of 34 closed-response, free-response, and demographic questions.[1]

In the survey questionnaire for Study 1, respondents were asked about their familiarity with the logic of evaluation, their views on the logic of evaluation, the frequency with which they apply each step of the logic of evaluation, and the difficulty in completing each step. For Study 2, the questions emphasized formulating evaluative conclusions and participants were asked whether they utilized performance standards in their evaluation practice, who is typically involved in standard setting, what type of performance standards are used, and what challenges they encounter in setting performance standards. For both survey questionnaires, skip patterns were used to elicit additional information as well as reduce response burden.

Excluding free-response items and demographic questions, internal consistency (i.e., the lower-bound estimate of reliability) over all ordinal-level items was estimated using ordinal α and Ω, assuming a congeneric (i.e., unidimensional) measurement model. For Study 1, ordinal $\alpha = 0.86$ and ordinal $\Omega = 0.89$ and for Study 2 ordinal $\alpha = 0.81$ and ordinal $\Omega = 0.84$.

3.4. Procedure

Both survey questionnaires were administered using the Qualtrics web-based survey system. An initial e-mail message inviting the two randomly selected samples of AEA members and informing them of the study and its purposes was sent one week prior to the initial administration of the surveys. Three reminder e-mails were delivered weekly over the administration period thereafter to those who were selected for the samples but had not yet responded. Throughout the planning and administration of both surveys, the principles of Dillman, Smyth, and Christian (2009) 'tailored design method' for conducting surveys were carefully applied in an effort to increase the quality and quantity of responses.

3.4.1. Institutional review

The study was reviewed and approved by the Western Michigan University (WMU) Human Subjects Institutional Review Board (HSIRB). Both samples of AEA members read an electronic informed consent prior to participating in the study.[2]

3.5. Data processing and analysis

Closed-response data were downloaded from the Qualtrics survey system as comma separated values (CSV) files and then imported into R for processing and analysis. Where relevant, bounds on errors of estimation, B (notated by ± [i.e., sampling error], for statistical estimates of population parameters were calculated. Free-response data were downloaded as text files and inductively analyzed following construction of an emergent coding scheme derived from an initial screening of responses using MAXqda. Throughout the presentation of findings below, results of free-response analyses are interwoven with closed-response analyses in order to facilitate interpretation and provide greater understanding than either would provide in isolation.

4. Findings

4.1. Study 1

Nearly three-fourths (71.84% ± 5.98%) of AEA members are either 'not at all familiar' or only 'a little familiar' with the logic of evaluation, with the remaining 28.16% (± 14.04%) being either 'familiar' or 'very familiar.' Simultaneously, a majority report that the logic of evaluation is 'useful' or 'very useful' and 'important' or 'very important' for planning, implementation, and reporting evaluations (see Fig. 1). Even those who are unfamiliar with the logic indicate often using the logic, knowingly or unknowingly (e.g., "I have been using this all along, not knowing that there was a theoretical basis."). The logic's usefulness, however, varies widely among practitioners and is exemplified in statements such as "It might work in academia, but not in the real world" and, conversely, "The logic of evaluation underlies all evaluation inquiry. If the logic is not used as a guide, then even the most sensical attempts to do evaluation just end up being social science research."

As regards each of the four steps of the logic of evaluation, nearly all AEA members indicate that identifying criteria (93.52% ± 1.23%), setting performance standards (81.48% ± 3.51%), comparing performance to standards (82.35% ± 4.22%), and synthesizing into an evaluative conclusion (86.14% ± 3.69%) are either 'important' or 'very important,' as shown in Fig. 2. Even so, many AEA members have no clear understanding of the meaning of either criteria or standards and frequently use the terms synonymously or incorrectly when probed. In addition to semantic irregularities and inconsistencies are statements such as "I typically collect data, write-up results, and provide interpretations and recommendations, and *then let the client decide what the criteria should be* [italics added for emphasis]"

In their practice, as shown in Fig. 3, the majority of AEA members most 'frequently' or 'always' identify criteria (82.41% ± 3.34%) and synthesize into an evaluative conclusion (75.00% ± 5.80%). AEA members less 'frequently' or 'always' compare performance to standards (62.14% ± 5.98%) and set performance standards (60.55% ± 7.39%). Some AEA members indicate that evaluative "criteria and performance standards are generally predetermined prior to an evaluation" and, therefore, this is "not a task in which I engage." Challenges associated with identifying criteria and setting standards

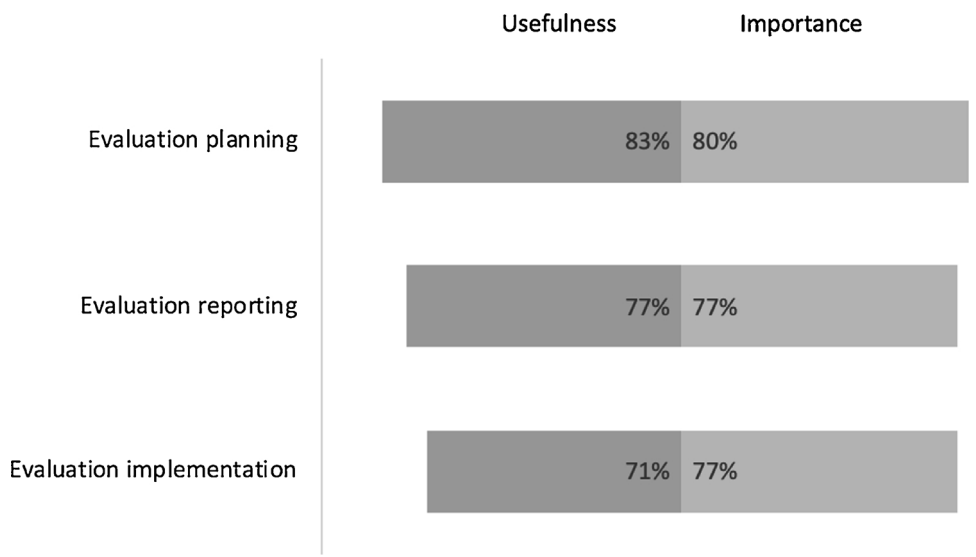

Fig. 1. Usefulness and importance of the logic of evaluation for planning, reporting, and implementation.

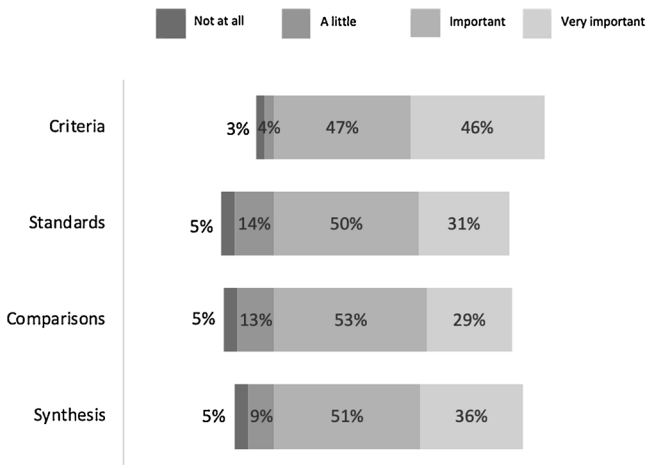

Fig. 2. Importance of each of the four steps of the logic of evaluation.

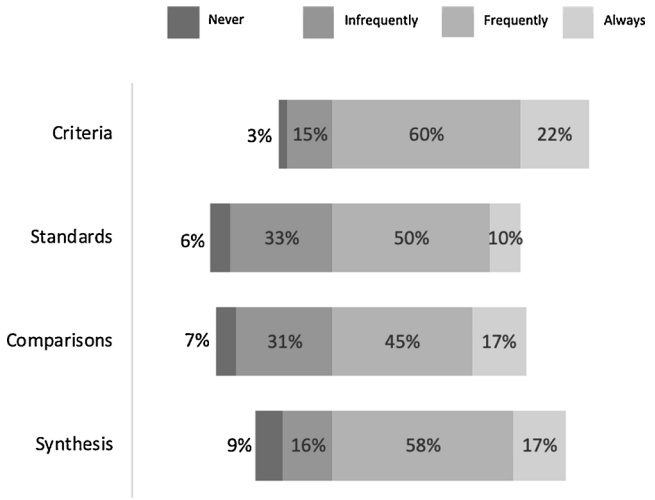

Fig. 3. Frequency of applying each of the four steps of the logic of evaluation.

include, but are not limited to, "getting different stakeholder groups to agree on criteria" and "preexisting standards often do not exist." In comparing performance to standards, the difficulties of agreeing upon or lack of preexisting standards are considered the major challenges, followed by "lack of stakeholder interest." As for synthesizing into evaluative conclusions, many simply resist "making judgements" and that such procedures are "too reductionist and cannot account for complexity."

As illustrated in Fig. 4, AEA members consider identifying criteria (50.47% ± 9.70%) as the most difficult step ('difficult' and 'very difficult' combined) in the logic of evaluation, followed by setting performance standards (48.51% ± 10.41%) and synthesizing into an evaluative conclusion (41.58% ± 12.57%). Fewer consider comparing performance to standards (27.66% ± 14.60%) as either 'difficult' or 'very difficult.' Difficulties arise at each step and occur for a variety of reasons (e.g., "even at the beginning of an evaluation, different stakeholders usually have different criteria for success," "I am often discouraged from using criteria and standards, even after explaining why they should be used," "we measure so many indicators that synthesizing them into a single result would be like administering a Rorschach test").

4.2. Study 2

Nearly all (89.36% ± 3.97%) AEA members report 'frequently' or 'always' formulating evaluative conclusions in their practice (see Fig. 5). In doing so, a majority apply standards of performance as a means for determining evaluative conclusions (58.51% ± 7.97%). Very few indicate 'frequently' or 'always' constructing a singular

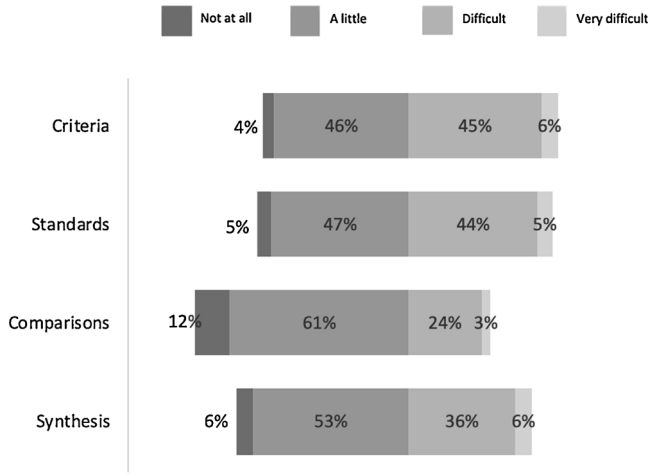

Fig. 4. Difficulty of applying each of the four steps of the logic of evaluation.

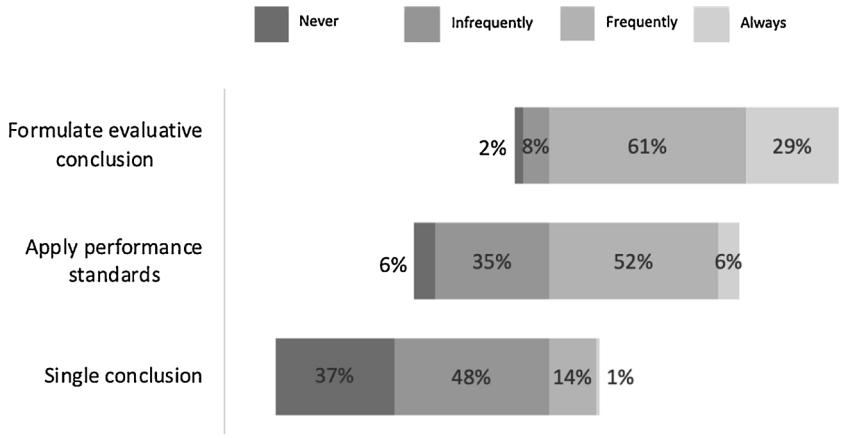

Fig. 5. Frequency and methods of evaluative conclusions.

evaluative conclusion about an evaluand (15.55% ± 12.15%). Of those who do, the vast majority consider such conclusions 'difficult' or 'very difficult' (76.79% ± 6.15%; only 22.32% ± 20.93% consider such conclusions 'not at all' or merely 'a little' difficult). More than half of evaluative conclusions are jointly determined by both evaluators and evaluation stakeholders (56.04% ± 9.12%), whereas 43.96% ± 11.66% of evaluative conclusions are evaluator-only. Difficulties associated with formulating evaluative conclusions (in particular, singular conclusions) vary widely from "insufficient data to support and justify" to "stakeholder disinterest in conclusions that are evaluative rather than descriptive."

More than two-thirds of AEA members construct or determine performance standards prior to conducting an evaluation (69.39% ± 9.11%) and slightly less than one-third do so during the evaluation process (26.53% ± 14.83%). Most often, performance standards are determined at the design phase (72.00% ± 11.29%), and less frequently during data collection (12.00%), data analysis (8.00%), or data interpretation (8.00%) phases. Determining performance standards is most frequently a collaborative effort between evaluators and evaluation stakeholders (79.38% ± 4.13%). Stakeholder-only and evaluator-only-determined standards occur far less frequently (10.31% and 9.28%, respectively). Existing standards are applied very rarely (1.03%). Performance standards are most often some combination of absolute (i.e., criterion-referenced) and relative (i.e., norm-referenced) standards (60.82% ± 7.86%). The use of only relative (22.68% ± 15.57%) or only absolute (13.40%) standards occurs far less frequently. Performance standards, if used, are infrequently (58.76% ± 8.28%) changed during the evaluation process, changed frequently 26.80% (± 14.86%), never changed 13.40%, or always changed only 1.03% of the time. Standards are sometimes changed or altered for a variety of reasons, some of which are exemplified by statements such as "stakeholders worry about not obtaining desired results" and "preliminary data sometimes indicates that original goals are unrealistic and need to be modified."

5. Discussion

Perhaps the most striking finding arising from the two investigations is that nearly three-fourths of AEA members are unfamiliar with the logic of evaluation, though most indicate applying a working logic in practice. This finding lends support to prior investigations in which it has been reported that few evaluators use evaluation theory to guide their practice (Christie, 2003; Shadish & Epstein, 1987). As a result of their findings, Shadish and Epstein worried that "…there is a danger of scholarly illiteracy in evaluation about its own writings and concepts" (p. 586). In part, these findings suggest practicing evaluators may not have extensive training in or familiarity with evaluation theory. Less than one fourth of master's degrees, Ph.D.'s, and certificates in evaluation require students to take a course specifically about evaluation theory (LaVelle, 2014). This lack of attention to evaluation theory in evaluation training, in combination with inadequate formal evaluation training for the average practitioner (Datta, 2003; Grob, 2018), buttress Shadish and Epstein (1987) conclusions.

Fundamental evaluation concepts such as criteria, standards, and synthesis are frequently misunderstood by a majority of self-identified practitioners. Synthesis, in particular, is a complex process and includes not only synthesis of facts and values (i.e., criteria) but also synthesis across multiple values (Davidson, 2005; Schwandt, 2015; Scriven, 2013), which was not mentioned by any respondent as a challenge associated with formulating evaluative conclusions. Instead of the methodological challenges usually associated with synthesis procedures, respondents generally referred to ideological (e.g., objections to making 'judgements') or structural difficulties (e.g., stakeholder 'disinterest'). Collectively, these findings beg the question of whether evaluation practitioners—specifically, those who are members of AEA—are, in fact, conducting 'evaluation' or are simply applying social science research methods to programs, policies, and other types of evaluands to address questions of interest to their clients and sponsors (Davidson, 2007).

Numerous scholars recognize that many evaluation theories are normative in origin; that is, theories derived from practice or observation (e.g., Fitzpatrick, Sanders, & Worthen, 2011; Mertens & Wilson, 2018; Miller, 2010; Patton, 2008; Schwandt, 2015; Stufflebeam & Coryn, 2014). Simultaneously, other theories are prescriptive and specify how evaluation ought to be conducted (e.g., Fitzpatrick et al., 2011; Mertens & Wilson, 2018; Miller, 2010; Patton, 2008; Schwandt, 2015; Stufflebeam & Coryn, 2014). As Shadish remarked in 1998, "… evaluation theory is "who we are" in the sense that it is the knowledge base that defines the profession. It is what we know that other professions don't know" (p. 5) as well as a defining, identifying feature of the profession, practice, field, and discipline. Even so, contemporary evaluation practice (and, therefore, evaluation theory itself, if normative) is largely controlled by and under the auspices of external forces and demands (e.g., evaluation sponsors and clients) rather than carefully constructed, well-articulated theories.

5.1. Implications

There is, ostensibly, a foundational base of evaluation knowledge. And, at the core of that foundation is the logic of evaluation. Given the ongoing discussions in regard to competencies, certification, credentialing, and professionalization (see entire issue of *New Directions for Evaluation* edited by Altschuld & Engle, 2015), the findings derived from the two investigations, coupled with those of other studies (e.g.,

LaVelle, 2014; LaVelle & Donaldson, 2010), suggest educational and other implications (e.g., all of the aforementioned would require some level of core knowledge requirement). Even so, debates both in support of and objecting to any form of credentialing have been occurring for decades (e.g., Jones & Worthen, 1999; Love, 1994) and are not likely to be resolved in the immediate future. Knowledge of evaluation theory, moreover, is only given a cursory mention in *The 2018 AEA Evaluator Competencies* (i.e., "The competent evaluator…selects evaluation approaches and theories appropriately." [American Evaluation Association (AEA), 2018, p. 2]). Coupled with other investigations (e.g., Christie, 2003; Shadish & Epstein, 1987), the findings arising from the two investigations suggest that evaluation is whatever practicing evaluators do rather than what practicing evaluators ought to do.

5.2. Limitations

The rates of response for both investigations were low (24.88% and 29.95%, respectively), though similar to most AEA member surveys (Coryn et al., 2019). Even so, both were probability samples and the characteristics of the two obtained samples are representative of the overall AEA member population.

The questions asked of AEA members in the two survey questionnaires intentionally sought breadth rather than depth. Therefore, the findings reported very likely oversimplify complex concepts. The investigations, taken as a whole, therefore, should be considered exploratory.

The two studies produced conflicting results in terms of the frequency of performing the last step of the logic of evaluation. In the first investigation almost three-fourths of respondents indicated that they frequently conducted the synthesis procedure, whereas in the second study few reported constructing a singular evaluative conclusion about an evaluand in their regular practice. This irregularity could potentially be due to a variety of factors including the construction and wording of the question prompts or response options provided, for example.

Lastly, and perhaps most importantly, the study samples consisted only of AEA members and do not reflect the diversity of evaluation practice throughout the world, generally, or those of non-AEA evaluation practitioners, specifically. According to Schröter (2017), November), many, if not most, practicing federal program evaluators in the United States are not members of any professional evaluation association or society, for example. In her investigation, only 35% of respondents indicated being AEA members and 65% self-reported having no knowledge of evaluation theory whatsoever.

5.3. Future research

Both the general logic and various working logics provide numerous opportunities for systematic inquiry comparing and contrasting similarities and differences in practice across evaluation approaches as well as between individuals having differing characteristics (e.g., levels of education, disciplinary backgrounds, primary work settings, roles as evaluator [internal versus external]). In addition, such investigations could potentially provide insight into what is prescribed in theory versus what is actually practiced (e.g., Coryn, Noakes, Westine, & Schröter, 2011; Miller & Campbell, 2006).

6. Notes

1 Both survey questionnaires are available from the first author upon request.

2 In addition to requesting informed consent, participants also received the following information in the initial and subsequent e-mail notifications: "You are receiving this e-mail as a member of the American Evaluation Association (AEA). This research request was reviewed by a Research Request Task Force consisting of tenured AEA members. If you have concerns about the survey and would like to express them to the AEA leadership, please e-mail info@eval.org. Any concerns raised will be shared, confidentially, with the Executive Committee of the association. AEA allows its membership list to be used infrequently for research that focuses on the field of evaluation. If you would like to opt-out of AEA's research list, please send an e-mail request to infro@eval.org. Please note that we encourage you to consider remaining on the list as such research strengthens and furthers the field's knowledge base."

References

Altschuld, J. W., & Engle, M. (2015). Accreditation, certification, and credentialing: Relevant concerns for US evaluators. *New Directions for Evaluation, 145*, 1–113.

American Evaluation Association (AEA) (2018). *The 2018 AEA evaluator competencies.* Retrieved fromhttps://www.eval.org/page/competencies.

Canadian Evaluation Society (CES) (2018). *Competencies for Canadian evaluation practice.* Retrieved fromhttps://evaluationcanada.ca/sites/default/files/competencies_for_canadian_evaluation_practice_revised_edition_2018.pdf.

Christie, C. A. (2003). What guides evaluation? A study of how evaluation practice maps onto evaluation theory. *New Directions for Evaluation, 97*, 7–35.

Christie, C. A. (2011). Advancing empirical scholarship to further develop evaluation theory and practice. *Canadian Journal of Program Evaluation, 26*(1), 1–18.

Christie, C. A., Quiñones, P., & Fierro, L. (2014). Informing the discussion on evaluator training: A look at evaluators' course taking and professional practice. *American Journal of Evaluation, 35*(2), 274–290.

Cook, T. D. (1997). Lessons learned in evaluation over the past 25 years. In E. Chelimsky, & W. R. Shadish (Eds.). *Evaluation for the 21st century: A handbook* (pp. 30–52). Thousand Oaks, CA: Sage.

Coryn, C. L. S., Becho, L. W., Westine, C. D., Mateu, P. F., Abu-Obaid, R. N., Hobson, K. A., et al. (2019). Material incentives and other potential factors associated with response rates to Internet surveys of American Evaluation Association members: Findings from a randomized experiment. *American Journal of Evaluation*.

Coryn, C. L. S., Noakes, L. A., Westine, C. D., & Schröter, D. C. (2011). A systematic review of theory-driven evaluation practice from 1990 to 2009. *American Journal of Evaluation, 32*(2), 199–226.

Coryn, C. L. S., Ozeki, S., Wilson, L. N., Greenman, G. D., II, Schröter, D. C., Hobson, K. A., et al. (2016). Does research on evaluation matter? Findings from a survey of American Evaluation Association members and prominent evaluation theorists and scholars. *American Journal of Evaluation, 37*(2), 159–173.

Coryn, C. L. S., Wilson, L. N., Westine, C. D., Hobson, K. A., Ozeki, S., Fiekowsky, E. L., et al. (2017). A decade of research on evaluation: A systematic review of research on evaluation published between 2005 and 2014. *American Journal of Evaluation, 38*(3), 329–347.

Cullen, A. E., Coryn, C. L. S., & Rugh, J. (2011). The politics and consequences of including stakeholders in international development evaluation. *American Journal of Evaluation, 32*(3), 345–361.

Davidson, E. J. (2007). Editorial: Unlearning some of our social scientist habits. *Journal of MultiDisciplinary Evaluation, 4*(8), iii–vi.

Davidson, E. J. (2005). *Evaluation methodology basics: The nuts and bolts of sound evaluation.* Thousand Oaks, CA: Sage.

Davies, R., & MacKay, K. (2014). Evaluator training: Content and topic validation in university evaluation courses. *American Journal of Evaluation, 35*(3), 419–429.

Datta, L. (2003). Important questions, intriguing method, incomplete answers. *New Directions for Evaluation, 97*, 37–46.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method* (3rd ed.). Hoboken, NJ: Wiley.

House, E. R., & Howe, K. (1999). *Values in evaluation and social research.* Thousand Oaks, CA: Sage.

Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2011). *Program evaluation: Alternative approaches and practical guidelines* (4th ed.). Upper Saddle River, NJ: Pearson.

Fournier, D. M. (1995). Establishing evaluative conclusions: A distinction between general and working logic. *New Directions for Evaluation, 68*, 15–32.

Grob, G. F. (2018). Evaluation practice: Proof, truth, client relationships, and professional growth. *American Journal of Evaluation, 39*(1), 123–132.

Henry, G. T., & Mark, M. M. (2003). Toward an agenda for research on evaluation. *New Directions for Evaluation, 97*, 69–80.

Jones, S. C., & Worthen, B. R. (1999). AEA members' opinions concerning evaluator certification. *American Journal of Evaluation, 20*(3), 495–506.

King, J., McKegg, K., Oakden, J., & Wehipeihana, N. (2013). Evaluative rubrics: A method for surfacing values and improving the credibility of evaluation. *Journal of MultiDisciplinary Evaluation, 9*(21), 11–20.

LaVelle, J. M. (2018). *2018 directory of evaluator education programs in the United States.* Retrieved fromhttps://conservancy.umn.edu/bitstream/handle/11299/200790/Directory_full.pdf?sequence=5&isAllowed=y.

LaVelle, J. M. (2014). *An analysis of evaluation education programs and evaluator skills across the world (doctoral dissertation).* Claremont, CA: Claremont Graduate University.

LaVelle, J. M. (2011). Planning for evaluation's future: Undergraduate students' knowledge of and interest in program evaluation. *American Journal of Evaluation, 32*(3), 362–375.

LaVelle, J. M., & Donaldson, S. I. (2015). The state of preparing evaluators. *New Directions for Evaluation, 145*, 39–52.

LaVelle, J. M., & Donaldson, S. I. (2010). University-based evaluation training programs in the United States 1980–2008: An empirical examination. *American Journal of Evaluation, 31*(1), 9–23.
Love, A. J. (1994). Should evaluators be certified? *New Directions for Program Evaluation, 62*, 29–40.
Mark, M. 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. R. Brandon (Eds.). *Fundamental issues in evaluation* (pp. 111–134). New York, NY: Guilford.
Mertens, D. M., & Wilson, A. T. (2018). *Evaluation theory and practice: A comprehensive guide* (2nd ed.). New York, NY: Guilford.
Miller, R. L. (2010). Developing standards for empirical examinations of evaluation theory. *American Journal of Evaluation, 31*(3), 390–399.
Miller, R. L., & Campbell, R. (2006). Taking stock of empowerment evaluation: An empirical review. *American Journal of Evaluation, 27*(3), 296–319.
Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Thousand Oaks, CA: Sage.
Schröter, D. C. (2017). Evaluation competencies among public administrators: Learning from others. *Paper presented at the annual meeting of the American Evaluation Association*.
Schwandt, T. (2015). *Evaluation foundations revisited: Cultivating a life of the mind for practice.* Stanford, CA: Stanford University Press.
Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Vol. Eds.), *Perspectives on curriculum evaluation: Vol. I*, (pp. 39–83). Chicago, IL: Rand McNally.
Scriven, M. (1980). *The logic of evaluation.* Inverness, CA: Edgepress.
Scriven, M. (1986). New frontiers of evaluation. *Evaluation Practice, 7*(1), 7–44.
Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Newbury Park, CA: Sage.
Scriven, M., et al. (2007). The logic of evaluation. In H. V. Hansen (Ed.). *Dissensus and the search for common ground, CD-ROM* (pp. 1–16). Windsor, ON: OSSA.
Scriven, M. (2012). The logic of valuing. *New Directions for Evaluation, 133*, 17–28.
Scriven, M. (2013). *Key evaluation checklist (KEC).* Retrieved from http://www.michaelscriven.info/images/KEC_3.22.2013.pdf.
Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice.* Newbury Park, CA: Sage.
Shadish, W. R., & Epstein, R. (1987). Patterns of program evaluation practice among members of the evaluation research society and evaluation network. *Evaluation Review, 11*(5), 555–590.
Stufflebeam, D. L., & Coryn, C. L. S. (2014). *Evaluation theory, models, & applications* (2nd ed.). San Francisco, CA: Jossey-Bass.
Szanyi, M., Azzam, T., & Galen, M. (2013). Research on evaluation: A needs assessment. *Canadian Journal of Program Evaluation, 27*(1), 39–64.

Satoshi Ozeki is a recent graduate of the Interdisciplinary Ph.D. in Evaluation (IDPE) program at Western Michigan University (WMU).

Chris L. Coryn is the Director of the Interdisciplinary Ph.D. in Evaluation (IDPE) program and a Professor of Evaluation in the Graduate College at Western Michigan University (WMU). He received a B.A. in Psychology from Indiana University (IU) in 2002 and a M.A. in Social and Experimental Psychology in 2004, also from IU. He earned his Ph.D. in Evaluation in 2007 at WMU. He has published more than 50 scholarly, peer-reviewed papers in journals such as the Albanian Journal of Agricultural Sciences, American Journal of Evaluation, Canadian Journal of Program Evaluation, Child & Adolescent Social Work Journal, Chronic Illness, Computers & Education, Current Research in Social Psychology, Energy Efficiency, Evaluation & The Health Professions, Evaluation and Program Planning, Evaluation Journal of Australasia, International Criminal Justice Review, Journal of Evaluation in Clinical Practice, Journal of Materials Education, Journal of Mixed Methods Research, Journal of Psychoeducational Assessment, New Directions for Evaluation, The Journal of Social Psychology, The Qualitative Report, World Medical & Health Policy, and has also authored and/or edited several books, book chapters, and monographs, and authored or coauthored nearly 100 commissioned evaluation and research reports. He is currently the Executive Editor of the Journal of MultiDisciplinary Evaluation and also serves on the Editorial Advisory Boards of the American Journal of Evaluation and Evaluation and Program Planning. He is also the coauthor (with Daniel L. Stufflebeam) of the second edition of Evaluation Theory, Models, & Applications and coeditor (with Carl D. Westine) of the four-volume Contemporary Trends in Evaluation Research. He has been involved in and led numerous research studies and evaluations, funded by the Department of Justice (DOJ), National Science Foundation (NSF), National Institutes for Health (NIH), nonprofits, and many others, across several substantive domains, including research and evaluation in education, science and technology, science policy, health and medicine, community and international development, and social and human services. Since obtaining his first graduate degree in 2004, he has served as the Principal Investigator, Co-Principal Investigator, or Methodologist for numerous research and evaluation grants and contracts totaling more than $3,000,000. He has given numerous lectures, speeches, and workshops, both nationally and internationally, including Albania, Austria, Canada, France, Germany, India, Japan, Kosovo, Luxembourg, Macedonia, Nepal, New Zealand, Russia, Switzerland, Thailand, United Kingdom, and United States. From 2010 to 2011 he provided instruction at The Evaluators' Institute (TEI), held annually in Chicago, Ottawa, San Francisco, and Washington. Additionally, he has worked and published with many of the field's leading scholars including Tom D. Cook, J. Bradley Cousins, E. Jane Davidson, John A. Hattie, Patrick McKnight, Robin L. Miller, Michael Q. Patton, Jim Rugh, James R. Sanders, Michael Scriven, and Daniel L. Stufflebeam, among others.

Daniela C. Schröter, a German native, holds a Ph.D. in interdisciplinary evaluation and a Master's degree in intercultural business communications and German and American studies. She is very active in research and evaluation studies on policy and strategy, food safety, networks and partnerships, evidence-based practice, and education. Her evaluation portfolio includes the United States Departments of Education and Agriculture, CDC, and national science and research foundations in Luxembourg, Russia, Switzerland, and the United States. In 2015 and 2016, Dr. Schroeter was honored for Excellence in Discovery at WMU. In 2013, she received the national Marcia Guttentag Award of the American Evaluation Association. Dr. Schroeter serves as an associate editor for the Journal of MultiDisciplinary Evaluation, which she helped found. Her work has truly international reach covering China, Germany, Kenya, Nepal, New Zealand, Norway, Russia, South Africa, Slovenia, Switzerland, and Thailand.