
Program Evaluation Standards for Utility Facilitate Stakeholder Internalization of Evaluative Thinking in the West Virginia Clinical Translational Science Institute

Reagan Curtis

West Virginia University

Abhik Roy

West Virginia University

Nikki Lewis

West Virginia University

Evana Nusrat Dooty

West Virginia University

Taylor Mikalik

West Virginia University

Background: The program evaluation standards (PES) can be considered established criteria for high-quality evaluations. We emphasize PES Utility standards and evaluation capacity building as we strive for meaningful application of our work in the real world.

Purpose: We focused our methodology on understanding how stakeholders discussed utility and how their perceptions related to our evaluation work aligned with the Utility domain of the program evaluation standards.

Setting: The West Virginia Clinical Translational Science Institute (WVCTSI), a statewide multi-institutional entity for which we have conducted tracking and evaluation since 2012.

Intervention: Sustained collaborative engagement of evaluation stakeholders with the goal of increasing their utilization of evaluation products and evaluative thinking.

Research Design: Case study.

Data Collection and Analysis: We interviewed five key stakeholders. We used themes developed from coding of interview data to inform document analyses. We used interview and document analyses to develop additional themes and illustrative examples, as well as to develop and describe a five-level evaluation uptake scale.

Findings: We describe shifts in initiation, use, and internalization of evaluative thinking by non-evaluation personnel. These shifts prompted our development and application of an evaluation uptake scale to capture increased evaluation capacity among stakeholders over time. We discuss how focus on the PES Utility standards and evaluation capacity building facilitated these shifts, and their implications for maximizing utility of evaluation activity in large, complex programmatic evaluations.

Keywords: *program evaluation standards; evaluation utility; evaluation capacity building.*

Journal of MultiDisciplinary Evaluation
Volume 19, Issue 43, 2023

JMDE
Journal of MultiDisciplinary Evaluation

ISSN 1556-8180
<http://www.jmde.com>

Introduction

We present our application of the Joint Committee on Standards for Educational Evaluation (JCSEE) program evaluation standards (PES) in the development and evaluation of the West Virginia Clinical Translational Science Institute (WVCTSI), which is a statewide multi-institutional entity for which we have conducted tracking and evaluation since 2012. We agree with Yarbrough et al. (2010) that the various PES are co-equal, in the sense that none can be ignored, but we emphasize utility as a necessary condition for the meaningful application of our work in the real world, because “accurate but unused evaluations have little if any actual worth” (p. xxxii). In that sense, utility coincides with evaluation capacity building as stakeholders take up and use evaluation products and apply evaluative thinking.

Within the context of large-scale federally funded programs, evaluations are not only essential but, in many circumstances, required components in oversight and reporting. Even with this mandate, there are numerous benefits to conducting high-quality program evaluations. Among these are the use of data for continual and incremental programmatic improvements and evaluation capacity building. Given this, while the results that come from high-quality program evaluations can be used to improve programs, resources to conduct program evaluations are frequently limited. This often pressures evaluators to shift from the ideal study of a program to something that is sufficient and feasible. In other words, a shift from a study that is *high quality* to one that is *good enough* can affect utility (e.g., Cooksy & Mark, 2012; Stevahn et al., 2020; White et al., 2017), especially in cases where stakeholders do not believe evaluations to be of high quality (Bundi et al., 2021).

Even with limited resources, there are quality checks that can guide evaluations; for example, the PES, hereafter denoted as “the standards,” outlined by the Joint Committee on Standards for Educational Evaluation (JCSEE; Yarbrough et al., 2011). The standards have been thoroughly investigated for their use in assessing program quality (e.g., Ruhe & Boudreau, 2013; Sanders, 1994; Westbrook et al., 2017; Wingate, 2009). The five domains (i.e., Utility, Feasibility, Propriety, Accuracy, and Evaluation Accountability) were developed to assess the quality of evaluation activities; each domain is outlined as an independent construct. We believe that evaluation capacity building, in the sense that stakeholders use evaluation findings and increase their own use of

evaluative thinking, should be a central goal for evaluators and, further, that such evaluation capacity building is supported by high-quality evaluations that address all five PES domains.

Literature Review

We focus our review of literature on how terms such as “evaluation,” “utility,” and “evaluation capacity building” have been understood and applied. Following King and Alkin (2019), we organize our discussion of the Utility standards by designating Standards U1, Evaluator Credibility; U2, Attention to Stakeholders; and U8, Concern for Consequences and Influence as criteria related to the *evaluator*. We then describe the remaining Utility standards as being associated with the *evaluation*: U3, Negotiated Purposes; U4, Explicit Values; U5, Relevant Information; U6, Meaningful Processes and Products; and U7, Timely and Appropriate Communicating and Reporting. Understanding the Utility standards as factors of the evaluator and the evaluation grounds our work.

Defining Program Evaluation and Utility

There remains a lack of consensus on what exactly constitutes a program evaluation. This disagreement has existed since the 1960s (Carter, 1971), if not earlier. Amongst academics, the description given in *Evaluation Thesaurus* (Scriven, 1991, p. 1) appears to be the most likely candidate for agreement (Picciotto, 2011), but is not adopted by all (Wanzer, 2021). For example, Wolf (1990) provided a procedural definition of evaluation as a process that often parallels the context of an evaluation, in particular as a systematic collection and interpretation of evidence, a judgment of values, and an orientation to action. Practitioners and the public at large have an even broader view of the term (King & Stevahn, 2013). As a result, the effort to define the term “evaluation utilization,” or simply “utility,” has had a long and complicated history and, as with the “program evaluation,” these terms exist without an agreement as to their meaning (Alkin & King, 2016).

Studies have indirectly (e.g., Coryn et al., 2011) and directly investigated frameworks for evaluation use, with varied outcomes (e.g., Fleischer & Christie, 2009; Olejniczak, 2017; Pattyn & Bouterse, 2020; Peterman & Gathings, 2019; Taut & Alkin, 2003; Turnbull, 1999). In three compilation studies pointed out by Alkin and King (2017), namely those from Cousins and Leithwood

(1986), Shulha and Cousins (1997), and Johnson et al. (2009), the common theme was an emphasis on approaches evaluators can use to increase the likelihood that their studies may be used to influence program or policy decisions.

Michael Kean once posited, “Similar to the conundrum, ‘If a tree falls in the forest and nobody hears it, did it make a sound?’, one may ask, ‘If a product/program/process is evaluated but the information is not considered, did an evaluation actually occur?’ ” (Alkin et al., 1990, p. 20). This idea that the usefulness of an evaluation is tied to its utility is commonplace, though it’s meaning has been debated.

The semantic debate surrounding the terms “use,” “utility,” and “utilization” has spanned decades. In terms of evaluation, there continues to be a lack of consensus regarding whether the terms are synonymous, much less an agreed upon definition of each (e.g., Alkin, 1982; Alkin & King, 2017; Braskamp et al., 1982; Braskamp, 1982; Campbell, 1971; Connolly & Porter, 1980; Dailak, 1982; Ginsburg & Rhett, 2003; Grasso, 2003; Heilman, 1983; Henry & Mark, 2003; King, 1988; Kirkhart, 2000; Riecken & Boruch, 1974; Rutman, 1982; Vedung, 2000/2017; Weiss, 1972).

We take a broad view of the idea, aligned mostly with Johnson et al. (2009): use, utility, and utilization are the application of evaluative processes, products, and/or findings to produce an impact or effect. Traditionally, the concept of feasibility has also been applied to focus on the context of an evaluation (e.g., Durning et al., 2007; Fashola, 1989; Gowda et al., 2019; Jephson, 1992; Stake, 2000), with a direct link between feasibility and utility, especially in cases where a process evaluation is needed (Chen, 2005; Wholey, 1994).

Evaluator Factors Influencing Utility and Evaluation Capacity Building

Here we focus on Utility PES considered to be directly connected to the evaluator, before addressing those connected to the evaluation in the following section. Evaluator Utility standards include U1, Evaluator Credibility; U2, Attention to Stakeholders; and U8, Concern for Consequences and Influence.

In the latter third of the 20th century, an evaluator’s credibility was a subject of some concern and was strongly connected to the idea of utility. Without a formal definition or set of criteria, evaluator credibility was predicated on the views of stakeholders (e.g., Braskamp et al., 1978; Ripley, 1985). Though mostly absent in the literature, since the publication of the PES there have been some

attempts to further operationalize and detail exactly what constitutes credibility and how practitioners can establish and maintain it. For example, Russ-Eft et al. (2008) described the credibility of evaluators as being contingent on their ability to realize the needs of an organization, its people, and its intended purpose, and noted that practitioners must find a way to install and preserve it. They outlined eight standards to consider when assessing credibility, namely that evaluators (a) maintain professionalism; (b) demonstrate relevant knowledge associated with an evaluand; (c) stay informed on modern views, methods, and pertinent technologies; (d) update their skill set accordingly; (e) participate in professional organizations and events associated with evaluation; (f) engage in evaluation capacity building; (g) be proficient in effective written, oral, and visual communication for disseminating work; and (h) build a network of professionals in the field (Russ-Eft et al., 2008).

Hopson and Horsford (2015) independently expanded on Russ-Eft et al. (2008) by providing additional guidance on approaches to increase credibility and thereby utility. Hopson and Horsford explained that evaluators should build trust within an organization’s community, address outcomes from an evaluation that may be perceived negatively, recognize any unique cultural artifacts that are inherent to those within the context of a study, and frequently describe both the importance and the process of any work being performed for context as well as stakeholder acceptance and support.

Bryson et al. (2011) defined stakeholders as those “individuals, groups, or organizations that can affect or are affected by an evaluation process and/or its findings” (p. 1). Furthermore, the *Encyclopedia of Evaluation* outlined how these individuals can be classified into multiple groups, namely those who are (a) primary users with the authority to make decisions about a program, b) directly responsible for the program and its operations, (c) primary and secondary beneficiaries of said program, and (d) affected negatively by the program (Greene, 2005). While the latter criteria are certainly important, the need to build and maintain a strong and ethical relationship with decision makers is a key in the acceptance of evaluation outcomes and utility of findings (Johnson et al., 2009). To identify and work with primary users in a participatory setting, Bryson and Patton (2015) recommended that evaluators build skills in facilitating relationships, building evaluation capacity, finding connected people who occupy central positions within a network, understanding that these individuals may change at

any given time, engaging in high-quality exchanges with stakeholders, developing a culture of evaluation, and exhibiting an understanding of cultural sensitivity.

U8, Concern for Consequences and Influence, is intimately tied to U2, Attention to Stakeholders, and to utility. “Use,” “utility,” and “utilization” are inherently benign terms, as each lacks a negative or positive connotation. Even the assumed definition mentioned earlier lacks any description of *how* evaluations are used. Specifically, “effect” and “impact” could refer to either helpful or adverse outcomes, or both. This perspective is supported by Cousins (2004), who notes that while the ethical behaviors of an evaluator are important when addressing the misuse of an evaluation, assessing ethical behaviors of an evaluator without considering actions of non-evaluator users misses an important part of the picture. While monitoring misuse is sometimes infeasible (Patton, 2005), assessing non-use is more difficult. In fact, of the four categorizations of use and misuse of evaluation findings given by Cousins (i.e., ideal use, misuse, unjustified non-use, and justified non-use), without direct knowledge of user intent, the dimension associated with non-use is the most difficult to assess, due to factors such as an organization’s resistance to change (Weiss, 1972), additional considerations not within the scope of an evaluation (Luukkonen-Gronow, 1989; Weiss, 1972), adopting favorable outcomes (Leviton & Hughes, 1981), or simply ignoring findings altogether (Alkin & Coyle, 1988; Alkin & King, 2017; Christie & Alkin, 1999).

As a result, evaluators with limited resources may choose to concentrate on ways to address evaluation use. A comprehensive synthesis of the numerous methods and techniques that have been reported for improving use of evaluation findings is beyond the scope of this paper (e.g., Alkin, 1982; Bundi et al., 2021; Bundi & Trein, 2022; Cousins & Leithwood, 1986; Donnelly et al., 2014; Johnson et al., 2009; King & Stevahn, 2013; Kirkhart, 2000; Patton, 2012; Peck & Gorzalski, 2009). However, Rogers (2018a, 2018b) identified seven broad approaches that are feasible and appear to address many findings within the literature: (a) recognizing the users and uses early in the evaluative process, (b) detecting obstacles that may hinder a study, (c) identifying when analyses and reporting are needed, (d) selecting modes of reporting that are appropriate and accessible, (e) pursuing additional information about an evaluand after the conclusion of a study, (f) creating knowledge products from what was learned, and (g) disseminating findings per a plan.

The needs of a stakeholder often change as a program matures. The involvement of multiple stakeholders reflects a diversity of values within an organization. This yields the opportunity for increased evaluation use (Alkin & King, 2017). In turn, the evaluative process benefits from an evaluator playing both the role of a trusted friend who is not afraid to tell the truth (Rallis & Rossman, 2000) and the role of a teacher who passes judgment (Schwandt, 2001), though they most certainly serve in many additional roles (e.g., Hirsch & Quartaroli, 2009; Mathison, 1991; Morabito, 2002).

Evaluation Factors Influencing Utility and Evaluation Capacity Building

Having discussed Utility standards connected to the evaluator, we now focus on those connected to the evaluation. Evaluation Utility standards include U3, Negotiated Purposes; U4, Explicit Values; U5, Relevant Information; U6, Meaningful Processes and Products; and U7, Timely and Appropriate Communicating and Reporting.

The AEA guiding principles include a focus on utility that overlaps with the PES, and utility is a central tenet of many proposed evaluator competencies (e.g., American Evaluation Association, 2018; Stevahn et al., 2005, 2020; Stufflebeam, 2003; Zorzi et al., 2002). However, utility in those instances is broadly defined. Following the taxonomy established by Cousins et al. (2004), we focus on approaches to evaluation capacity building where stakeholders learn about the evaluative process by active participation.

This view, also known as *process use*, affords individuals the opportunity to both conduct evaluations and utilize the results (e.g., Chouinard, 2013; Chouinard & Cousins, 2012; Cousins et al., 2014; Cullen & Coryn, 2011; Grack Nelson & Schreiber, 2009; Odera, 2021). Patton (1998) reported seven types of process use: (a) enhancing shared understanding; (b) increasing engagement, self-determination, and ownership; (c) infusing evaluative thinking into organizational culture; (d) instrumentation effects and reactivity; (e) supporting and reinforcing program intervention; and (f) program and organizational development. However, Forss et al. (2002) described somewhat differing types, namely with respect to participants (a) boosting morale, (b) creating shared understanding, (c) developing networks, (d) learning to learn, and (e) strengthening the project. Patton described purposes affiliated with long-term goals, whereas those denoted by Forss et al. appear to be more immediate.

Bezzi (2006) argued that the merit of evaluation can be found in “discovering unknown meanings, which help stakeholders to develop a new self-awareness, and in implementing new connections between people, actions, and thoughts.” (p. 67). This aligns with our view that evaluators should take a participatory approach to build an evaluation community who serve not only as participants, but also as partners who themselves become agents of change. The utility to stakeholders of being involved in the evaluative process not only builds capacity, but also bolsters trust between stakeholders and evaluators.

Our Perspective, Context, and Evidence Base

We believe that empowering stakeholders to find utility in an evaluation, such that they use evaluative products and thinking to effect change, should be the central goal of all evaluation work, and that the other PES domains function to further that goal. The program evaluation standards focus evaluators on addressing context, stakeholder values, and constituent needs as they strive to conduct evaluations with utility, feasibility, propriety, accuracy, and evaluation accountability. What follows is a description of our use of the PES in the context of 10 years of evaluation work for and with WVCTSI, and how that work has facilitated shifts in initiation, use, and internalization of evaluative thinking by WVCTSI personnel. We draw on 10 years of evaluation data and experiences, focusing primarily on funded proposals (both initial and renewed) detailing evaluation plans; collaborative logic models, which have evolved over time; quarterly evaluation reports and associated steering committee discussions; targeted evaluation reports on various subcomponents; and participant-check interviews with key stakeholders, focused on their evaluation needs and utilization of our evaluation recommendations.

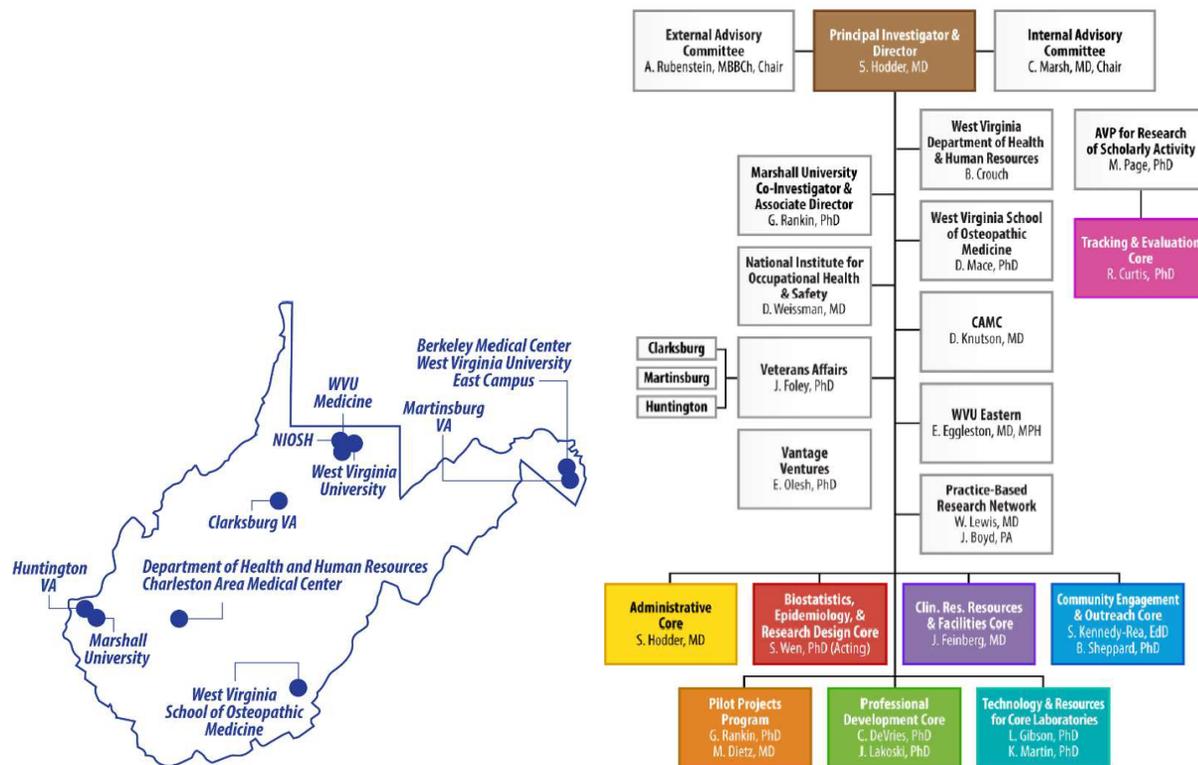
Context: West Virginia Clinical Translational Science Institute (WVCTSI)

Funded by the National Institute of General Medical Sciences Clinical and Translational Research Awards, WVCTSI requires robust programmatic evaluation. WVCTSI's overarching goal is to serve as a transformational force, leading statewide clinical and translational collaborative

research initiatives to catalyze and accelerate delivery of solutions that address WV health problems. Critical to this goal is the development of well-trained investigators, committed to working with rural populations and developing trust among local communities, such that research results are believed. This goal aligns well with West Virginia University's mission statement, which states “a commitment to a diverse and inclusive culture that advances education, healthcare, and prosperity for all by providing access and opportunity, by advancing high-impact research, and by leading transformation in West Virginia.”

WVCTSI is a statewide partnership network encompassing WV's major academic medical centers as well as the state's osteopathic medical school, the Veteran's Administration facilities, and the National Institute for Occupational Safety and Health (see Figure 1). This large and complex network for clinical and translational science trains, supports, and enables clinical translational researchers to make positive impacts on health and health care, with a focus on priority health disparities in addiction and resulting emerging epidemics, cancer, cardiovascular disease, chronic lung disease, and most recently COVID-19. The complexity of the organization of WVCTSI includes eight key component areas (KCAs), often referred to as “cores”, through which eight institutional partners coordinate their work, and this complexity presents programmatic evaluation challenges and opportunities. We lead the tracking and evaluation (TE) core, situated externally to all other core leadership to facilitate our ability to provide external perspective and evaluative input.

Figure 1. WVCTSI Partners and Organizational Structure



Our WVCTSI tracking and evaluation activity is designed to drive data-based decision-making and to ensure that all goals are achieved. This programmatic evaluation system is grounded in the PES and includes robust systems and processes to provide leadership and personnel with objective performance data to assess progress toward WVCTSI aims and core objectives. TE Core uses collaborative capacity building evaluation and accountability frameworks to broadly disseminate evaluation data and recommendations, maintain focus on state-level health outcomes, assist all cores in their evaluation and data-based decision-making processes, and empower WVCTSI personnel to recognize and address challenges to productivity and desired outcomes. Our primary focus is on providing actionable evaluation products and empowering WVCTSI personnel to utilize evaluative thinking for continuous quality improvement. This, in our view, is how utility should be defined.

We have seen an increase in WVCTSI personnel initiating requests for more in-depth evaluation support rather than TE Core initiating such interactions. That, coupled with TE Core's expertise

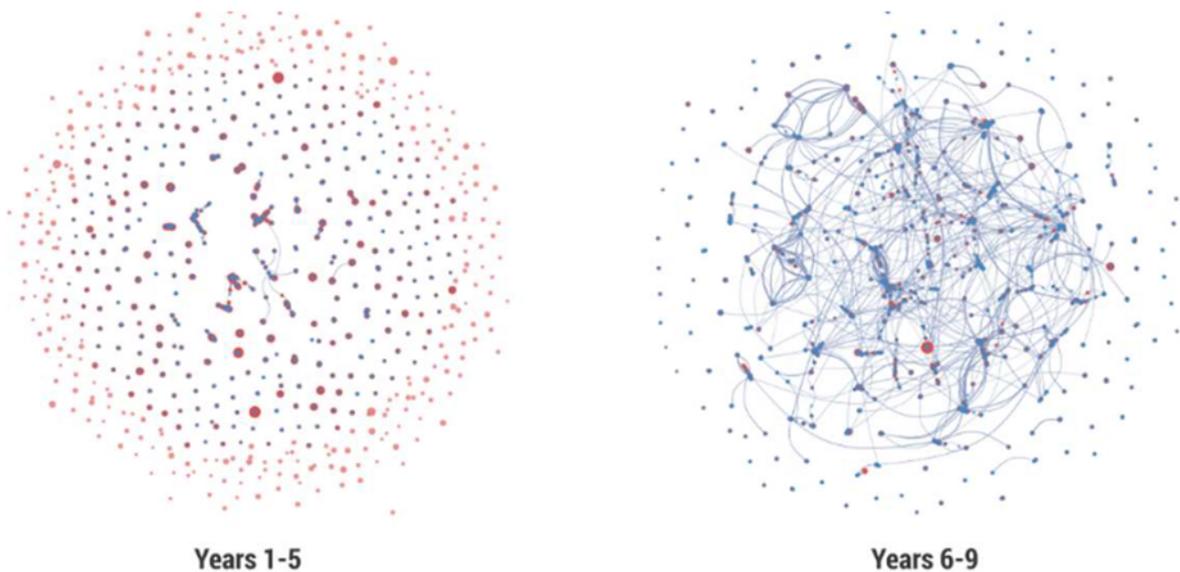
in data science, allows us to create and support utilization of data-science resources. We have piloted and made publicly available some such resources (e.g., WVCTSI Linked Publications Hub at <https://percwv.com/datatables/pubs/>) and continue to develop others. These resources integrate data collection with automated data visualization to provide near-real-time evaluation data to WVCTSI personnel, allowing them to rapidly translate implications from that data to inform their core practices. TE Core support helps WVCTSI personnel increase their proficiency at utilizing these resources and focus their evaluative thinking to identify evaluation recommendations and translate them into practice. TE Core personnel continually look for opportunities to create additional resources to address emerging evaluation data and training needs.

As a project's complexity increases, so do its challenges in adhering to the PES, such as attending to increasingly diverse stakeholders and considering concerns for consequences and influence in more complicated networks. WVCTSI's complexity has increased over time, which is partially revealed through changes in its measures

of productivity and collaboration, measures that also help delineate the context of the case study and evaluative approach we describe here. There are a total of 1,325 publications linked to WVCTSI during the current 5-year funding cycle, nearly doubling the number of linked publications (698) in the first 5-year funding cycle. The high level of publication productivity has built a strong foundation for successful external funding applications. The external funding resulting from WVCTSI services totals \$159.3 million to date during the current funding cycle, more than triple the linked external

funding of Years 1 through 5 (\$48.9 million). As another indicator of increasing complexity, a network analysis of investigators with publications relating to WVCTSI-enabled projects is shown in Figure 2. Years 1 through 5 show 696 total linked publications by 2,418 unique authors. The same analysis conducted for Years 6 through 9 illustrates a dramatic increase in collaboration and team science, showing an increase of 1,151 publications and 4,297 authors with 1,255 more authors contributing to multiple authorship teams compared to baseline.

Figure 2. Co-Authorship Networks Show Increase in Collaboration Over Time



Note: Blue dot = sole author publication. Blue with red dot = multi-author publication. Curved blue line = co-author connections across multiple publications.

Methodology: Evidence Base and Analytic Approach

Aligned with our perspective that utility should be the primary goal of evaluation, we focused our methodology on understanding how stakeholders discussed utility and how their perceptions related to our evaluation work aligned with the PES Utility domain. We interviewed five key stakeholders who have worked with WVCTSI since 2013 in a variety of leadership roles across multiple cores (not including TE). The semi-structured interview protocol is provided in Table 1. Interviews were audio-recorded, transcribed, and coded by two independent raters. Raters used an a priori coding system to categorize all text related to each PES

Utility standard. The description of each Utility standard provided below was taken as a definition of the code, and the few discrepancies between raters were resolved through consensus. The Utility standards used for our coding of interview transcripts are as follows:

- **U1 Evaluator Credibility:** Evaluations should be conducted by qualified people who establish and maintain credibility in the evaluation context.
- **U2 Attention to Stakeholders:** Evaluations should devote attention to the full range of individuals and groups invested in the program and affected by its evaluation.

- U3 Negotiated Purposes: Evaluation purposes should be identified and continually negotiated based on the needs of stakeholders.
- U4 Explicit Values: Evaluations should clarify and specify the individual and cultural values underpinning purposes, processes, and judgments.
- U5 Relevant Information: Evaluation information should serve the identified and emergent needs of stakeholders.
- U6 Meaningful Processes and Products: Evaluations should construct activities, descriptions, and judgments in ways that encourage participants to rediscover, reinterpret, or revise their understandings and behaviors.
- U7 Timely and Appropriate Communicating and Reporting: Evaluations should attend to the continuing information needs of their multiple audiences.
- U8 Concern for Consequences and Influence: Evaluations should promote responsible and adaptive use while guarding against unintended negative consequences and misuse. (<https://evaluationstandards.org/program/>)

Table 1. WVCTSI Key Stakeholder Interview Protocol

Interview Protocol

1. Please tell me about the history of your association with WVCTSI.
 - a. When did you first associate with WVCTSI and in what role?
 - b. How, if at all, has your role with WVCTSI changed over time?
2. When did you first become aware of the tracking and evaluation core?
 - a. What was your initial understanding of tracking and evaluation's role in WVCTSI?
 - b. What specific tracking and evaluation activities impacted your work with WVCTSI? In what ways did those activities impact your work?
3. Think back for a moment and when you are ready, please describe in detail the earliest experience you remember involving tracking and evaluation core personnel or activity.
 - a. Prompt as needed for details re: what TE core actually did.
 - b. Why do you think that experience was memorable for you?
 - c. How, if at all, did that experience impact you?
4. Now please describe a more recent memorable experience with tracking and evaluation core personnel or core activity. (same prompts as #3)
5. How, if at all, has your perception of WVCTSI tracking and evaluation changed over time? [Reflect back to respondent X categories of changes; repeat prompts for each X]
 - a. Please describe a specific example related to changes in X.
 - b. Can you think of any other specific examples related to those changes?
6. How, if at all, have you used tracking and evaluation data, reports, or services?
 - a. How, if at all, has your use of tracking and evaluation data, reports, or services changed over time?
7. Is there anything else you would like to tell me, or any additional questions you think I should have asked?

Themes developed from analysis of PES Utility standard coding of interview data informed analysis of initial (years 1 through 5) and renewal (years 6 through 10) funded proposals (including detailed evaluation plans); collaborative logic models for each core, which have evolved over time; 10 years of quarterly evaluation reports (review of metrics connected to logic models, plus evaluative recommendations) and associated steering committee discussions (focused on implications of quarterly evaluation reports and other evaluation products); and targeted evaluation reports on various subcomponents. Interview and document analysis were used to develop the themes and

illustrative examples provided below, as well as to develop and describe a five-level evaluation uptake scale that we believe illuminates evaluation capacity building toward maximizing utility. The levels we developed are as follows and will be described more fully with illustrative examples in the Findings and Discussion section below:

- Level 0: No evaluation products
- Level 1: Evaluation products documented
- Level 2: Utilization of evaluation products by stakeholders documented (uptake)
- Level 3: Evidence of evaluative thinking engaged in by stakeholders (ownership)

- Level 4: Stakeholders initiate new evaluative efforts (transcendence)

Findings and Discussion

We present findings and discussion of the application of two analytic frames, consisting of coding stakeholder interview data by PES Utility standard and coding evaluative activity by level of evaluative uptake. We believe this focus also helps reveal how other PES domains were supportive of utility. Our Utility standard analysis revealed a dual-pronged theme regarding the interrelatedness of the Utility standards, and our evaluative uptake analysis revealed change across time as we focused on evaluation capacity building.

Utility Standard Analysis

Analysis of coded interview transcripts revealed that interviewed stakeholders spoke most about U5, Relevant Information, and U6, Meaningful

Processes and Products. In fact, more than half the time, when statements were coded as any other standard, those statements were also coded as U5 or U6 (see Table 2). This co-occurrence signaled a two-pronged theme common across all stakeholder interviews. This theme suggested both that (a) U5 and U6 were most salient for interviewees as critically important aspects of what the evaluation team did for them, and (b) the other Utility standards were importantly supportive of successfully applying U5 and U6. While we did not explicitly code for other PES domains beyond Utility, we did recognize that much of the discussion coded as U5 or U6 and overlapping with other Utility standards also would overlap with standards in other PES domains. We consider the lack of statements coded as U4, Explicit Values, or U8, Concern for Consequences and Influence, as a fruitful area for renewed focus in our evaluative work with WVCTSI going forward and a clear indication of the importance of interrogating our work utilizing the standards as a frame of reference.

Table 2. Utility Standard Coding Co-Occurrence

	Total	w/U2	w/U3	w/U5	w/U6	w/U7
U1	13	3	0	12	9	1
U2	16	-	1	13	11	2
U3	9	-	-	6	6	0
U5	46	-	-	-	24	3
U6	33	-	-	-	-	4
U7	5	-	-	-	-	-

Note: No statements were coded as U4 or U8. Numbers do not sum to total due to multiple codes per statement.

The following quote, which is representative of the first part of this theme, illuminates the importance of the standards' focus on being relevant and on encouraging stakeholders to rediscover, reinterpret, or revise their understandings and behavior:

Like, for the [specific component] evaluation, the initial focus group information and stuff like that, we've used it in our [recent] publication. We use it in our reporting that we give back to stakeholders. We use that information with the tracking or the quarterly reports for [our core] as in a larger—it's really helped us guide where we're not doing as much as we thought that we were, and need to beef that up, and other places where we were surprised like, "Wow, this is taken off. We

didn't really expect that." I think just always looking back at the data and just tracking and seeing the ebbs and flows of things are two examples that we definitely use it for in our core.

Our analysis of interview data, especially the co-occurrence of codes, made it clear that our interviewees believed the Utility standards focused on relevance and meaningfulness were most salient for them, but also that those aspects of utility were enabled by aligning our work with other Utility standards. The following quote, which is representative of the second part of this theme, illustrates how Utility standards focused on evaluator credibility and attention to stakeholders contributed:

I think the investigator tracking, that is a little bit different for [our core], because most of our investigators are not on campus or even at the other partner campuses. They're out in the community, so that working with the tracking and evaluation core about how to navigate that, and get them hooked up to services, and making sure we're tracking them, and stuff like that, that's been a big thing. Here recently, working with them on the [specific component] evaluation, we've got a big evaluation going on currently that started with [TE Core personnel]. For me, I just think the understanding and expertise in that group is a lot more than I knew when I first started, and then just starting to integrate more on different projects has really taken off in the past couple of years.

The following quote is representative of how U3, Negotiated Purposes; U5; and U6 co-occurred, such that negotiated purposes facilitated relevant and meaningful utility:

It was probably right at the beginning of COVID, but there was an evaluation meeting at WVCTSI as a whole, where we really had a whole discussion about how we wanted to capture those practice and policy changes and what they really mean and can we standardize that language. Because that's something [our core] had done in the [specific component] and some of our other community partners, but other cores really didn't have that connection. I think that was really helpful, that instead of— We had had conversations with the evaluation core prior to that and had made some progress, but being able to talk about it across all cores, and that was really [TE core personnel's] idea to talk about it in that setting. I think we were able to really come to a better understanding there. We still use those definitions we came up with that day.

Relatedly, this next quote focuses also on U7, Timely and Appropriate Communicating and Reporting, supportive of relevant and meaningful utility:

The data that we're getting is important. We're not wasting time on things that aren't really important or will make a difference in how the WVCTSI is run. Now the data, I think, really reflects what we need and what we're interested in that can either course correct or make sure that we're on the right track with the WVCTSI.

Evaluative Uptake Analysis

Our analysis of stakeholder interviews, evaluation products, and related documents (e.g., quarterly evaluation reports and related steering committee minutes), in the context of our dual focus on the standards and evaluation capacity building, led us to develop and apply a scale to categorize and describe uptake, ownership, and transcendence of evaluative thinking by stakeholders. We developed our evaluation uptake scale after considering analyses described in the prior section and reanalyzing interview data and evaluation products through emergent qualitative content analysis coding, with a focus on use of evaluation products and evidence of evaluative thinking. Our emergent coding of evidence of evaluative thinking uncovered distinctions we describe below as the various levels in this scale, which we then used to analyze how that evidence shifted over time across the first 9 years of our WVCTSI evaluation work.

We believe the upper levels of this scale are descriptive of how utility should be defined. Said another way, we believe utility comes about when the standards and evaluative practice enable stakeholders who are not evaluation personnel not only to use evaluative products but to apply and initiate evaluative activity directed toward continuous quality improvement as part of their daily activity. We next describe our evaluation uptake scale and give illustrative examples. Table 3 presents evidence of change over time in the level of evaluative uptake by non-evaluation personnel in WVCTSI.

Table 3. Instances of Evaluative Uptake Over Time

	Year 1–2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Year 9
Level 1	11	5						
Level 2	4	9	7	6	6	6	8	5
Level 3				1		11	3	1
Level 4								4

Note. Each instance is an evaluation product that was disseminated by TE core, or used, contributed to, or initiated by non-evaluation personnel. Level 0 has been omitted from this table, because instances of a lack of products cannot be counted. Level 0 was the situation for the first six months of Year 1.

The scale we developed consists of five levels, the first of which describes the situation before evaluation products have been developed and shared with the intention of those products being used by stakeholders. Evaluation products are primarily evaluation reports describing evaluation processes, data, findings, and recommendations, but also include data-science tools that allow stakeholders to directly interact with and download evaluation-relevant data. Analogous to the approach given by many introductory manuals (e.g., Davidson, 2022), Level 0 implies no evaluation products and is the situation that generally exists prior to a project getting off the ground and may continue to obtain as initial groundwork is done to develop evaluation products. It is important that evaluative activity in this stage attend to the PES to optimize the potential for later utility and evaluation capacity building.

As an example, in 2012, when WVCTSI was initially funded, we worked to position the evaluation team as external (i.e., not reporting administratively to other project leadership, which is atypical for this funding mechanism) and to establish our credibility as experts with demonstrated experience on large and complex programmatic evaluations. We solicited input from all WVCTSI core and partner leaders regarding their familiarity with logic models and, based on a low level of initial expertise from them, developed a logic model training, where core personnel collaboratively revised initial logic models we built from the funded proposal. These logic models eventually described how specific core and partner activities were meant to facilitate core and overall specific aims, as well as specific metrics to document activity and reveal progress toward those aims. We worked closely and collaboratively with multiple personnel from each core to iteratively improve these logic models. These activities were related to Utility standards focused on evaluator credibility and attention to stakeholders but did not involve documented evaluation products until after

logic models were disseminated and utilized to structure quarterly evaluation reports. We believe our focus on instantiating the standards at this stage set a foundation for our later work and developing relationships that have been critical to our success as an evaluative team working with WVCTSI.

Level 1 is the first level where evaluation product dissemination is documented; it continues until there is evidence of utilization, or uptake, of those documents by stakeholders. It remains important throughout this and subsequent levels that evaluators continue to attend to the standards and to evaluation capacity building to help stakeholders progress up the levels in this scale. For instance, if evaluators have not established credibility, attended to stakeholders, and made their evaluative products relevant, meaningful, and timely, then it is unlikely the evaluation products will be used.

Echoing Lawrenz, Gullickson, and Toal (2007), we believe that it is crucial in practice to recognize that the standards describe continua rather than dichotomies. That is, rather than being met or unmet, they are met by degree. For example, it is not the case that evaluation is either relevant or not, but instead that somewhat relevant evaluation products can iteratively be improved through continued attention to stakeholders who see evaluators as credible (alongside attention to other standards). Our experience has shown that evaluation activity that attends to moving along continua described by the standards importantly facilitates evaluation capacity building. For WVCTSI, for example, we disseminated quarterly reports and logic models during years 1 through 2 with little, if any, evidence that these were utilized by stakeholders. We then focused on meeting with individual core personnel and revising quarterly reports and logic models based on their feedback while also describing how we thought about, and hoped they would think about, the utility of these evaluation products. By years 3 and 4, we saw more consistent evidence of the uptake and use of these

evaluation products. That said, we continue to work to make them more relevant, meaningful, and timely to facilitate continued movement toward increased utility.

Level 2 is defined by evidence that stakeholders use an evaluation product initiated and disseminated by evaluators. This happens, for instance, when stakeholder control is low (Jacobson & Azzam, 2018) and the evaluation reports on specific components (e.g., a workshop evaluation focused on satisfaction and recommendations for improvement) are used to advocate for continued or increased resources for that component. Level 2 is, in a sense, the foundational level at which any evaluation needs to arrive. It is at this point that the work of evaluators begins to have potential to impact change that can be documented, as stakeholders use evaluation products to facilitate their work. This constitutes uptake, or the instrumental use of findings (Vo, 2015), but not ownership or transcendence as described in subsequent levels.

Level 3, evidence of evaluative thinking engaged in by stakeholders (ownership), requires not only that evaluative products have been disseminated and used by stakeholders, but also that stakeholders are applying evaluative thinking as they utilize those products or help develop other products (Brandon & Fukunaga, 2014). This is more than passing along evaluation findings or using them to advocate for their program. It requires stakeholders to have internalized evaluative thinking and make contributions to the evaluation somewhat independently of evaluation personnel. We say “somewhat independently” because, in our experience, this most often happens collaboratively with evaluation personnel, but the important aspect is that some of the evaluative ideas originate with non-evaluation personnel. We believe our dual focus on standards (such as attention to stakeholders, relevance, and meaningfulness) and evaluation capacity building facilitated our stakeholders’ beginning to make this shift from evaluation consumer to evaluation contributor.

We saw several examples of this (Level 3) emerging in our second 5-year funding period with WVCTSI. Among other examples, these included non-evaluation personnel contributions to improving quarterly reporting data collection measures, personnel in one core collaborating on the evaluation plan for an external funding proposal, and collaborative development of a community member focus group protocol. In each of these cases, it was noticeable that stakeholders who were not evaluators took the lead in some

aspects of collaboratively developing and iteratively improving evaluation efforts.

Expanding on findings from Rodriguez-Campos (2011), we suggest a Level 4 where stakeholders initiate new evaluative efforts (transcendence). Evaluation uptake at this level has only recently emerged in our work with WVCTSI, and we believe it has largely been enabled by our developing a context where Level 3 was encouraged and valued. Though the reality of evaluation use is not well known in these types of large complex multi-site studies (Olejniczak et al., 2016), it is reasonable that this be considered an extension of the Utility standards, especially U3, Negotiated Purposes, and U6, Meaningful Processes and Products, as stakeholders move beyond (i.e., transcend) contributing evaluative thinking to requesting and initiating new evaluative efforts.

We believe our focus on Utility standards and evaluation capacity building enabled a climate where non-evaluation personnel felt they could learn from evaluators and then begin to apply what they learned. Often, in our experience, non-evaluative personnel begin with questions related to PES domains other than Utility. Questions related to Feasibility, Propriety, Accuracy, and Evaluation Accountability standards predominate these conversations. We align our efforts with guidance from the PES and seek to share our knowledge with non-evaluation personnel through individual meetings around evaluation products and through group workshops to facilitate evaluation processes and evaluation tool use. Our evaluative work with non-evaluation personnel helps stakeholders come to understand how and why they should use evaluative thinking and products, thus driving toward Utility. Over time, as non-evaluative personnel became more familiar and comfortable with evaluative thinking, this climate and our continued focus on honoring and extending their contributions leads to increasingly robust contributions from personnel across WVCTSI. This has included personnel outside of evaluation asking for help and then contributing to designing, building, and deploying new evaluation instruments and, in one case, taking the lead on an evaluation-relevant manuscript for publication. As uptake at this level has newly emerged, we have relatively few examples, but we are hopeful that more will emerge, and facilitating that emergence is a continuing focus of our evaluation work with WVCTSI.

Conclusion

In conclusion, we have found great value in aligning our work with the PES, with a specific goal to continuously improve utility and evaluation capacity building. Our early work was focused on building feasibility-, propriety-, accuracy-, and evaluation accountability-aligned evaluation processes and products. We recognize this early work as setting a foundation such that stakeholders value our evaluation work and therefore want to use evaluation products and engage in evaluative thinking. We have seen a shift over time as our focus on aligning our work with the PES helped build and strengthen our credibility and working relationships with stakeholders. We had seen a steady improvement in the utility of our work through stakeholders' sharing their perspectives anecdotally with us, and the improvement was substantiated by the stakeholders interviewed here. Our analysis of evaluation products and related evidence of stakeholder uptake, ownership, and transcendence showed relatively consistent improvement over time, with transcendence emerging only recently. This progression highlights the importance of sustained long-term focus on increasing utility and evaluation capacity building. We hope that the evaluation uptake scale described here will be a valuable tool for others to apply in their own evaluative contexts, and that such application will focus evaluators on supporting stakeholders not only to value and utilize evaluation products but also to contribute evaluative thinking for continuous quality improvement.

Author Note

Reagan Curtis: <https://orcid.org/0000-0003-1873-1744>
 Abhik Roy: <https://orcid.org/0000-0002-7085-8964>
 Nikki Lewis: <https://orcid.org/0000-0002-7765-8965>
 Evana Nusrat Dooty: <https://orcid.org/0000-0002-2081-4840>

We have no conflicts of interest to disclose. Correspondence concerning this article should be addressed to Reagan Curtis, 355 Oakland Street, Morgantown, WV 26506. Email: Reagan.Curtis@mail.wvu.edu.

References

Alkin, M. C. (1982). Introduction: Parameters of evaluation utilization/use. *Studies in Educational Evaluation*, 8(2), 153–155. [https://doi.org/10.1016/0191-491X\(82\)90006-2](https://doi.org/10.1016/0191-491X(82)90006-2)

Alkin, M. C., & Coyle, K. (1988). Thoughts on evaluation utilization, misutilization and non-utilization. *Studies in Educational Evaluation*, 14(3), 331–340. [https://doi.org/10.1016/0191-491X\(88\)90027-2](https://doi.org/10.1016/0191-491X(88)90027-2)

Alkin, M. C., & King, J. A. (2016). The historical development of evaluation use. *American Journal of Evaluation*, 37(4), 568–579. <https://doi.org/10.1177/1098214016665164>

Alkin, M. C., & King, J. A. (2017). Definitions of evaluation use and misuse, evaluation influence, and factors affecting use. *American Journal of Evaluation*, 38(3), 434–450. <https://doi.org/10.1177/1098214017717015>

American Evaluation Association. (2018). *AEA Evaluator Competencies*. <https://www.eval.org/About/Competencies-Standards/AEA-Evaluator-Competencies>

Bezzi, C. (2006). Evaluation pragmatics. *Evaluation*, 12(1), 56–76. <https://doi.org/10.1177/1356389006064189>

Brandon, P. R., & Fukunaga, L. L. (2014). The state of the empirical research literature on stakeholder involvement in program evaluation. *American Journal of Evaluation*, 35(1), 26–44. <https://doi.org/10.1177/1098214013503699>

Braskamp, L. A. (1982). A definition of use. *Studies in Educational Evaluation*, 8(2), 169–174. [https://doi.org/10.1016/0191-491X\(82\)90009-8](https://doi.org/10.1016/0191-491X(82)90009-8)

Braskamp, L. A., Brown, R. D., & Newman, D. L. (1978). The credibility of a local educational program evaluation report: Author source and client audience characteristics. *American Educational Research Journal*, 15(3), 441–450. <https://doi.org/10.2307/1162497>

Braskamp, L. A., Brown, R. D., & Newman, D. L. (1982). Studying evaluation utilization through simulations. *Evaluation Review*, 6(1), 114–126. <https://doi.org/10.1177/0193841X8200600108>

Bryson, J. M., & Patton, M. Q. (2015). Analyzing and engaging stakeholders. In K. E. Newcomer, H. P. Hatry, & J. S. Wholey (Eds.), *Handbook of practical program evaluation* (4th ed.). Jossey-Bass & Pfeiffer Imprints, Wiley.

Bryson, J. M., Patton, M. Q., & Bowman, R. A. (2011). Working with evaluation stakeholders: A rationale, step-wise approach and toolkit. *Evaluation and Program Planning*, 34(1), 1–12.

- <https://doi.org/10.1016/j.evalprogplan.2010.07.001>
- Bundi, P., Frey, K., & Widmer, T. (2021). Does evaluation quality enhance evaluation use? *Evidence & Policy: A Journal of Research, Debate and Practice*, 17(4), 661–687. <https://doi.org/10.1332/174426421X16141794148067>
- Bundi, P., & Trein, P. (2022). Evaluation use and learning in public policy. *Policy Sciences*, 55(2), 283–309. <https://doi.org/10.1007/s11077-022-09462-6>
- Campbell, D. T. (1971). Reforms as experiments. *Urban Affairs Quarterly*, 7(2), 133–171. <https://doi.org/10.1177/107808747100700202>
- Carter, R. K. (1971). Clients' resistance to negative findings and the latent conservative function of evaluation studies. *The American Sociologist*, 6(2), 118–124.
- Chen, H. (2005). *Practical program evaluation: Assessing and improving planning, implementation, and effectiveness*. Sage.
- Chouinard, J. A. (2013). The case for participatory evaluation in an era of accountability. *American Journal of Evaluation*, 34(2), 237–253. <https://doi.org/10.1177/1098214013478142>
- Chouinard, J. A., & Cousins, J. B. (2012). *Participatory evaluation up close: A review and integration of research-based knowledge*. Information Age Publishing.
- Christie, C. A., & Alkin, M. C. (1999). Further reflections on evaluation misutilization. *Studies in Educational Evaluation*, 25(1), 1–10. [https://doi.org/10.1016/S0191-491X\(99\)00006-1](https://doi.org/10.1016/S0191-491X(99)00006-1)
- Connolly, T., & Porter, A. L. (1980). A user-focused model for the utilization of evaluation. *Evaluation and Program Planning*, 3(2), 131–140. [https://doi.org/10.1016/0149-7189\(80\)90061-0](https://doi.org/10.1016/0149-7189(80)90061-0)
- Cooksy, L. J., & Mark, M. M. (2012). Influences on evaluation quality. *American Journal of Evaluation*, 33(1), 79–84. <https://doi.org/10.1177/1098214011426470>
- 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. <https://doi.org/10.1177/1098214010389321>
- Cousins, J. B. (2004). Commentary: Minimizing evaluation misuse as principled practice. *American Journal of Evaluation*, 25(3), 391–397. <https://doi.org/10.1177/109821400402500311>
- Cousins, J. B., Goh, S., Clark, S., & Lee, L. (2004). Integrating evaluative inquiry into the organizational culture: A review and synthesis of the knowledge base. *The Canadian Journal of Program Evaluation*, 19(2), 99–141.
- Cousins, J. B., Goh, S. C., Elliott, C. J., & Bourgeois, I. (2014). Framing the capacity to do and use evaluation. *New Directions for Evaluation*, 141, 7–23. <https://doi.org/10.1002/ev.20076>
- Cousins, J. B., & Leithwood, K. A. (1986). Current empirical research on evaluation utilization. *Review of Educational Research*, 56(3), 331–364. <https://doi.org/10.3102/00346543056003331>
- Cullen, A. E., & Coryn, C. L. S. (2011). Forms and functions of participatory evaluation in international development: A review of the empirical and theoretical literature. *Journal of MultiDisciplinary Evaluation*, 7(16), 32–47.
- Daillak, R. H. (1982). What is evaluation utilization? *Studies in Educational Evaluation*, 8(2), 157–162. [https://doi.org/10.1016/0191-491X\(82\)90007-4](https://doi.org/10.1016/0191-491X(82)90007-4)
- Davidson, E. (2022). *Evaluation methodology basics: The nuts and bolts of sound evaluation*. Sage. <https://doi.org/10.4135/9781452230115>
- Donnelly, C., Letts, L., Klinger, D., & Shulha, L. (2014). Supporting knowledge translation through evaluation: Evaluator as knowledge broker. *The Canadian Journal of Program Evaluation*, 29(1).
- Durning, S. J., Hemmer, P., & Pangaro, L. N. (2007). The structure of program evaluation: An approach for evaluating a course, clerkship, or components of a residency or fellowship training program. *Teaching and Learning in Medicine*, 19(3), 308–318. <https://doi.org/10.1080/10401330701366796>
- Fashola, J. B. (1989). Evaluation, feasibility and relevance. *English for Specific Purposes*, 8(1), 65–73. [https://doi.org/10.1016/0889-4906\(89\)90007-0](https://doi.org/10.1016/0889-4906(89)90007-0)
- Fleischer, D. N., & Christie, C. A. (2009). Evaluation use: Results from a survey of U.S. American Evaluation Association members. *American Journal of Evaluation*, 30(2), 158–175. <https://doi.org/10.1177/1098214008331009>
- Forss, K., Rebién, C. C., & Carlsson, J. (2002). Process use of evaluations: Types of use that precede lessons learned and feedback. *Evaluation*, 8(1), 29–45. <https://doi.org/10.1177/13589020020080015>

- Ginsburg, A., & Rhett, N. (2003). Building a better body of evidence: New opportunities to strengthen evaluation utilization. *American Journal of Evaluation, 24*(4), 489–498. <https://doi.org/10.1177/109821400302400406>
- Gowda, D., Curran, T., Khedagi, A., Mangold, M., Jiwani, F., Desai, U., Charon, R., & Balmer, D. (2019). Implementing an interprofessional narrative medicine program in academic clinics: Feasibility and program evaluation. *Perspectives on Medical Education, 8*(1), 52–59. <https://doi.org/10.1007/s40037-019-0497-2>
- Grack Nelson, A., & Schreiber, R. C. (2009). Participatory evaluation: A case study of involving stakeholders in the evaluation process. *Visitor Studies, 12*(2), 199–213. <https://doi.org/10.1080/10645570903203521>
- Grasso, P. G. (2003). What makes an evaluation useful? Reflections from experience in large organizations. *American Journal of Evaluation, 24*(4), 507–514. <https://doi.org/10.1177/109821400302400408>
- Greene, J. C. (2005). Stakeholders. In S. Mathison (Ed.), *Encyclopedia of evaluation*. Sage.
- Heilman, J. G. (1983). Beyond the technical and bureaucratic theories of utilization: Some thoughts on synthesizing reviews and the knowledge base of the evaluation profession. *Evaluation Review, 7*(6), 707–728. <https://doi.org/10.1177/0193841X8300700601>
- Henry, G. T., & Mark, M. M. (2003). Beyond use: Understanding evaluation's influence on attitudes and actions. *American Journal of Evaluation, 24*(3), 293–314. <https://doi.org/10.1177/109821400302400302>
- Hirsch, M. L., & Quartaroli, T. A. (2009). Many hats: The methods and roles of the program evaluator. *Journal of Applied Social Science, 3*(2), 73–80. <https://doi.org/10.1177/193672440900300207>
- Hopson, R., & Horsford, S. (2015, September 7). *WE Week: Rodney Hopson and Sonya Horsford on "But can you do it" questions of evaluator credibility and organizational capacity: The nuances of evaluator credibility* [AEA365]. <https://aea365.org/blog/we-week-rodney-hopson-and-sonya-horsford-on-but-can-you-do-it-questions-of-evaluator-credibility-and-organizational-capacity-rodney-hopson-and-sonya-horsford-on-the-nuances-of-eval/>
- Jacobson, M. R., & Azzam, T. (2018). The effects of stakeholder involvement on perceptions of an evaluation's credibility. *Evaluation and Program Planning, 68*, 64–73. <https://doi.org/10.1016/j.evalprogplan.2018.02.006>
- Jephson, M. B. (1992). The purposes, importance, and feasibility of program evaluation in community-based early intervention programs. *Journal of Early Intervention, 16*(3), 252–261. <https://doi.org/10.1177/105381519201600305>
- Johnson, K., Greenesid, L. O., Toal, S. A., King, J. A., Lawrenz, F., & Volkov, B. (2009). Research on evaluation use: A review of the empirical literature from 1986 to 2005. *American Journal of Evaluation, 30*(3), 377–410. <https://doi.org/10.1177/1098214009341660>
- King, J. A. (1988). Research on evaluation use and its implications for evaluation research and practice. *Studies in Educational Evaluation, 14*(3), 285–299. [https://doi.org/10.1016/0191-491X\(88\)90025-9](https://doi.org/10.1016/0191-491X(88)90025-9)
- King, J. A., & Stevahn, L. (2013). *Interactive evaluation practice: Mastering the interpersonal dynamics of program evaluation*. SAGE.
- Kirkhart, K. E. (2000). Reconceptualizing evaluation use: An integrated theory of influence. *New Directions for Evaluation, 88*, 5–23. <https://doi.org/10.1002/ev.1188>
- Lawrenz, F., Gullickson, A., & Toal, S. (2007). Dissemination: Handmaiden to evaluation use. *American Journal of Evaluation, 28*(3), 275–289. <https://doi.org/10.1177/1098214007304131>
- Leviton, L. C., & Hughes, E. F. X. (1981). Research on the utilization of evaluations: A review and synthesis. *Evaluation Review, 5*(4), 525–548. <https://doi.org/10.1177/0193841X8100500405>
- Luukkonen-Gronow, T. (1989). The impact of evaluation data on policy determination. In D. Evered & S. Harnett (Eds.), *The evaluation of scientific research*. John Wiley and Sons.
- Mathison, S. (1991). Role conflicts for internal evaluators. *Evaluation and Program Planning, 14*(3), 173–179. [https://doi.org/10.1016/0149-7189\(91\)90053-J](https://doi.org/10.1016/0149-7189(91)90053-J)
- Morabito, S. M. (2002). Evaluator roles and strategies for expanding evaluation process influence. *American Journal of Evaluation, 23*(3), 321–330. <https://doi.org/10.1177/109821400202300307>

- Odera, E. L. (2021). Capturing the added value of participatory evaluation. *American Journal of Evaluation, 42*(2), 201–220. <https://doi.org/10.1177/1098214020910265>
- Olejniczak, K. (2017). The game of knowledge brokering: A new method for increasing evaluation use. *American Journal of Evaluation, 38*(4), 554–576. <https://doi.org/10.1177/1098214017716326>
- Olejniczak, K., Raimondo, E., & Kupiec, T. (2016). Evaluation units as knowledge brokers: Testing and calibrating an innovative framework. *Evaluation, 22*(2), 168–189. <https://doi.org/10.1177/1356389016638752>
- Patton, M. Q. (1998). Discovering process use. *Evaluation, 4*(2), 225–233. <https://doi.org/10.1177/13563899822208437>
- Patton, M. Q. (2005). *Misuse of evaluations*. In S. Mathison (Ed.), *Encyclopedia of evaluation*. Sage.
- Patton, M. Q. (2012). *Essentials of utilization-focused evaluation*. SAGE.
- Pattyn, V., & Bouterse, M. (2020). Explaining use and non-use of policy evaluations in a mature evaluation setting. *Humanities and Social Sciences Communications, 7*(1), 85. <https://doi.org/10.1057/s41599-020-00575-y>
- Peck, L. R., & Gorzalski, L. M. (2009). An evaluation use framework and empirical assessment. *Journal of MultiDisciplinary Evaluation, 6*(12), 139–156.
- Peterman, K., & Gathings, M. J. (2019). Using a community-created multisite evaluation to promote evaluation use across a sector. *Evaluation and Program Planning, 74*, 54–60. <https://doi.org/10.1016/j.evalprogplan.2019.02.014>
- Picciotto, R. (2011). The logic of evaluation professionalism. *Evaluation, 17*(2), 165–180. <https://doi.org/10.1177/1356389011403362>
- Rallis, S. F., & Rossman, G. B. (2000). Dialogue for learning: Evaluator as critical friend. *New Directions for Evaluation, 86*, 81–92. <https://doi.org/10.1002/ev.1174>
- Riecken, H. W., & Boruch, R. F. (1974). *Social experimentation: A method for planning and evaluating social intervention*. Academic Press.
- Ripley, W. K. (1985). Medium of presentation: Does it make a difference in the reception of evaluation information? *Educational Evaluation and Policy Analysis, 7*(4), 417–425. <https://doi.org/10.3102/01623737007004417>
- Rodriguez-Campos, L. (2011). Stakeholder involvement in evaluation: Three decades of the American Journal of Evaluation. *Journal of MultiDisciplinary Evaluation, 8*(17), 57–79.
- Rogers, P. (2018a, January 26). 7 Strategies to improve evaluation use and influence—Part 1. *Better Evaluation*. https://www.betterevaluation.org/en/blog/strategies_for_improving_evaluation_use_and_influence
- Rogers, P. (2018b, February). 7 Strategies to improve evaluation use and influence—Part 2. *Better Evaluation*. https://www.betterevaluation.org/blog/strategies_for_improving_evaluation_use_and_influence_part_2
- Ruhe, V., & Boudreau, J. D. (2013). The 2011 Program Evaluation Standards: A framework for quality in medical education programme evaluations. *Journal of Evaluation in Clinical Practice, 19*(5), 925–932.
- Russ-Eft, D. F., Bober, M. J., de la Teja, I., Foxon, M., & A. Koszalka, T. (2008). *Evaluator competencies: Standards for the practice of evaluation in organizations* (1st ed). Jossey-Bass.
- Rutman, L. (1982). Dimensions of utilization and types of evaluation approaches. *Studies in Educational Evaluation, 8*(2), 163–168. [https://doi.org/10.1016/0191-491X\(82\)90008-6](https://doi.org/10.1016/0191-491X(82)90008-6)
- Sanders, J. R. (1994). *The program evaluation standards: How to assess evaluations of educational programs*. Sage.
- Schwandt, T. A. (2001). Responsiveness and everyday life. *New Directions for Evaluation, 92*, 73–88. <https://doi.org/10.1002/ev.36>
- Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Sage.
- Shulha, L. M., & Cousins, J. B. (1997). Evaluation use: Theory, research, and practice since 1986. *Evaluation Practice, 18*(3), 195–208. <https://doi.org/10.1177/109821409701800302>
- Stake, R. E. (2000). Program evaluation, particularly responsive evaluation. In D. L. Stufflebeam, G. F. Madaus, & T. Kellaghan (Eds.), *Evaluation models: Viewpoints on educational and human services evaluation* (pp. 343–362). Springer Netherlands.
- Stevahn, L., Berger, D. E., Tucker, S. A., & Rodell, A. (2020). Using the 2018 AEA Evaluator Competencies for effective program evaluation practice. *New Directions for Evaluation, 168*, 75–97. <https://doi.org/10.1002/ev.20434>
- Stevahn, L., King, J. A., Ghery, G., & Minnema, J. (2005). Establishing essential competencies for program evaluators. *American Journal of*

- Evaluation*, 26(1), 43–59.
<https://doi.org/10.1177/1098214004273180>
- Stufflebeam, D. L. (2003). Professional standards and principles for evaluations. In T. Kellaghan & D. L. Stufflebeam (Eds.), *International handbook of educational evaluation* (pp. 279–302). Springer Netherlands.
https://doi.org/10.1007/978-94-010-0309-4_18
- Taut, S. M., & Alkin, M. C. (2003). Program staff perceptions of barriers to evaluation implementation. *American Journal of Evaluation*, 24(2), 213–226.
<https://doi.org/10.1177/109821400302400205>
- Turnbull, B. (1999). The mediating effect of participation efficacy on evaluation use. *Evaluation and Program Planning*, 22(2), 131–140. [https://doi.org/10.1016/S0149-7189\(99\)00012-9](https://doi.org/10.1016/S0149-7189(99)00012-9)
- Vedung, E. (2017). *Public policy and program evaluation*. Routledge. (Original work published 2000)
- Vo, A. (2015). Foreword. In C. Christie & A. Vo (Eds.), *Evaluation use and decision-making in society: A tribute to Marvin C. Alkin* (pp. vii–xvii). Information Age Publishing.
- Wanzer, D. L. (2021). What is evaluation? Perspectives of how evaluation differs (or not) from research. *American Journal of Evaluation*, 42(1), 28–46.
<https://doi.org/10.1177/1098214020920710>
- Weiss, C. H. (1972). *Evaluation research: Methods for assessing program effectiveness*. Prentice-Hall.
- Westbrook, T. R., Avellar, S. A., & Seftor, N. (2017). Reviewing the reviews: Examining similarities and differences between federally funded evidence reviews. *Evaluation Review*, 41(3), 183–211.
<https://doi.org/10.1177/0193841X16666463>
- White, J., National Academies of Sciences, Engineering, and Medicine (U.S.) (Eds.). (2017). *Principles and practices for federal program evaluation: Proceedings of a workshop*. The National Academies Press.
- Wholey, J. S. (1994). Handbook of practical evaluation. In J. S. Wholey, H. P. Hatry, & K. E. Newcomer (Eds.), *Assessing the feasibility and likely usefulness of evaluation* (pp. 15–39). Jossey-Bass Publisher.
- Wingate, L. A. (2009). *The program evaluation standards applied for metaevaluation purposes: Investigating interrater reliability and implications for use* [Doctoral dissertation, Western Michigan University].
- Wolf, R. M. (1990). A framework for evaluation. In H. J. Walberg & G. D. Haertel (Eds.), *The international encyclopedia of educational evaluation*. Pergamon Press.
- Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (Ed.). (2010). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). SAGE.
- Zorzi, R., Perrin, B., Mcguire, M., Long, B., & Lee, L. (2002). Defining the benefits, outputs, and knowledge elements of program evaluation. *The Canadian Journal of Program Evaluation*, 17(3), 143–150.