

Evaluation Utility Metrics (EUMs) in Reflective Practice

Ralph Renger

Just Evaluation Services, LLC, Arizona, United States

Jessica Renger

Claremont Graduate University, California, United States

Richard N. Van Eck and Marc D. Basson

University of North Dakota, North Dakota, United States

Jirina Renger

Walden University, Minnesota, United States

Abstract: *The article proposes three evaluation utility metrics to assist evaluators in evaluating the quality of their evaluation. After an overview of reflective practice in evaluation, the different ways in which evaluators can hold themselves accountable are discussed. It is argued that reflective practice requires evaluators to go beyond evaluation quality (i.e., technical quality and methodological rigor) when assessing evaluation practice to include an evaluation of evaluation utility (i.e., specific actions taken in response to evaluation recommendations). Three Evaluation Utility Metrics (EUMs) are proposed to evaluate utility: whether recommendations are considered (EUM_c), adopted (EUM_a), and (if adopted) the level of influence of recommendations (EUM_i). The authors then reflect on their experience in using the EUMs, noting the importance of managing expectations through negotiation to ensure that EUM data are collected and the need to consider contextual nuances (e.g., adoption and influence of recommendations are influenced by multiple factors beyond the control of the evaluators). Recommendations for increasing EUM rates by paying attention to the frequency and timing of recommendations are also shared. Results of implementing these EUMs in a real-world evaluation provide evidence of their potential value: practice tips led to an EUM_c of 100% and EUM_a of over 80%. Methods for considering and applying all three EUMs together to facilitate practice improvement are also discussed.*

Keywords: *evaluation, evaluation recommendations, metrics, reflective practice, utility*

Résumé : *L'article propose trois mesures d'utilité en évaluation pour aider les évaluateurs et évaluatrices à évaluer la qualité de leurs évaluations. Après un survol de la pratique réflexive en évaluation, il est question des différentes façons par lesquelles*

Corresponding Author: Ralph Renger, 14777 E. Circle M Ranch Place, Vail, Arizona, United States, 85641; ralph@justevaluation.com

les évaluateurs et évaluateurs peuvent se responsabiliser vis-à-vis de leur pratique. On avance que la pratique réflexive demande aux évaluateurs et aux évaluateurs d'aller au-delà de la qualité de l'évaluation (par exemple, la qualité technique et la rigueur méthodologique) au moment d'évaluer la pratique d'évaluation pour tenir aussi compte de l'utilité de l'évaluation (par exemple, des gestes précis posés en réponse à des recommandations en évaluation). Trois mesures d'utilité en évaluation (EUM) sont proposées pour évaluer l'utilité : les recommandations considérées (EUM_c), adoptées (EUM_a) et (si adoptées) le niveau d'influence des recommandations (EUM_i). Les auteurs et autrices parlent ensuite de leurs expériences d'utilisation des EUM, notamment l'importance de gérer les attentes par négociation pour veiller à ce que les données d'EUM soient recueillies et le besoin de considérer les nuances contextuelles (par exemple, l'adoption et l'influence des recommandations sont soumises à de multiples facteurs dont le contrôle échappe aux évaluateurs et évaluateurs). Des suggestions liées à l'augmentation des taux d'EUM sont faites, qui portent notamment sur la nécessité de prêter attention à la fréquence et au moment des recommandations. Les résultats de la mise en œuvre de ces EUM dans de vraies évaluations montrent leur valeur potentielle : des conseils quant à la pratique ont mené à une EUM_c de 100 % et à une EUM_a de plus de 80 %. Il est aussi question de méthodes pour considérer et mettre en œuvre les trois EUM ensemble pour faciliter l'amélioration de la pratique.

Mots-clés : *évaluation, recommandations en évaluation, mesures, pratique réflexive, utilité*

EVALUATION UTILITY METRICS (EUMS) IN REFLECTIVE PRACTICE

The purpose of this article is to share metrics we use in our evaluation practice as a way of reflecting on the utility of our work and holding ourselves accountable. Reflective practice is the first and arguably most fundamental of the five evaluation competency domains (CES, 2018). The reflective practice domain consists of six competencies that “focus on the evaluator’s knowledge of evaluation theory and practice; application of evaluation standards, guidelines, and ethics; and awareness of self, including reflection on one’s practice and the need for continuous learning and professional growth” (p. 5).

Some have suggested that evaluation itself is a form of reflective practice, where we as the evaluators hold up an external mirror for the client. We agree, and we would further suggest that reflective practice is also about holding up the mirror to *ourselves* (evaluators)—to hold ourselves accountable for the value and utility of our evaluation.

The program evaluation standards recommend three ways in which we can hold ourselves accountable (Yarbrough, et al., 2010). The first is by maintaining documentation on the “designs, procedures, data, and outcomes” (Yarbrough et al., 2010). Scrupulous documentation is a hallmark of scientific inquiry (Anderson, 2004). Having a record of what was done, when, and how is essential for

understanding corrective actions that may be needed to improve one's evaluation practice.

The program evaluation standards also recommend meta-evaluation as a tool for holding ourselves accountable. A meta-evaluation is an evaluation of evaluation. That is, the evaluation itself is the evaluand (Bustelo, 2002; Scriven, 1991, Stufflebeam, 1974). Scriven (1991) wrote that "meta-evaluation is the professional imperative of an evaluation: it represents the recognition that 'evaluation begins at home,' that evaluation is self-referent and not just something one does to others" (p.229). If we hold up the mirror for our clients, then we need to hold it up to ourselves.

There are two types of meta-evaluation: internal and external. Bustelo (2002, p. 5) has noted that "internal meta-evaluations are carried out by the evaluators by themselves: thus, the same people in charge of the evaluation do its meta-evaluation. External meta-evaluations are done by someone not involved in the assessed evaluation process or study." The findings from both types of meta-evaluation provide opportunity for practice reflection and improvement.

The criteria for meta-evaluations are well documented (e.g., Stufflebeam, 1974), and there are models to help guide the self-reflection process (Smith, et al., 2015). Further, many tools are available to assist with completing a meta-evaluation, such as the key evaluation checklist (KEC) (Scriven, 2007), the Performance Assessment Rating Tool (ETA, 2021), our guiding principles (AEA, 2021), and of course the program evaluation standards themselves (Yarbrough et al., 2010).

Both the meta-evaluation criteria and the tools to evaluate them place a heavy emphasis on technological and methodological rigor (e.g., internal validity, reliability, objectivity, etc.) (Bundi et al., 2021). McCormick (1997) noted that many evaluators believe that "technical quality and methodological rigour . . . should be the primary criteria by which evaluations are judged" (p. 1). Although it was written over 25 years ago, we argue that McCormick's observation is still the predominant perspective in meta-evaluation; many evaluators believe that their accountability begins and ends with designing and conducting a quality evaluation.

We argue, however, that while reflecting on the quality of an evaluation is certainly a necessary first step to establish that evaluation findings are trustworthy (Yarbrough et al., 2010), the necessary second step is to establish whether and to what extent the overarching *purpose* of evaluation was met. Put another way, it is necessary then to determine whether the evaluation was of value (Scriven, 1991).

Among the indicators of an evaluation's value is whether the findings are used and (when used) the extent to which they improve the evaluand. Evaluation use is defined as is "the intentional and serious consideration of evaluation information by an individual with the potential to act on it" (King & Pechman, 1982, p. 40). The factors influencing evaluation use are many and well documented (Johnson et al., 2009). For quite some time in our discipline it was generally accepted that an evaluation high in quality will increase evaluation use (Cousins & Leithwood, 1986). However, we agree with Bundi et al. (2021) that "the quality of an evaluation and evaluation use do not necessarily go hand in hand" (p. 1). Thus, it is

necessary to reflect on both evaluation quality and use; it would be an error to assume that our evaluations will be used even if the evaluation was designed and implemented with the highest quality.

While the literature on the relationship between evaluation use and quality may be equivocal, there is general agreement that evaluation utility is a necessary (if not sufficient) predictor of evaluation use (Patton, 2008). Evaluation utility is defined as the extent to which the evaluation has “value to someone or some institution” (Scriven, 1991, p. 368). McCormick (1997) insisted an evaluation “should not be done if there [is] no prospect for it being useful for some audience” (p. 2). We concur; weighing evaluation utility is of the utmost importance during our meta-evaluation and reflection processes.

We further argue that evaluation use is itself still not a sufficient metric for meta-evaluation. Evaluation findings may be used but have little or no influence on the evaluand. Of course, the further we get down this logic chain, the less control we have. For example, an evaluation finding may not be used for any number of reasons, including communication barriers, lack of perceived value, or simple forgetfulness on the part of the evaluand. While context and nuance must be accounted for when considering evaluation utility during meta-evaluation, use and influence of findings remain important elements of the process.

Like many evaluators, however, we have found evaluating the utility of our evaluations to be challenging. The program evaluation utility standard recommends consideration of eight dimensions of utility, many of which are highly subjective and difficult to operationalize (Yarbrough et al., 2010). Patton (2013) developed a Utility Focused Evaluation Checklist (UFEC), consisting of 17 steps to guide a meta-evaluation of utility, and steps 14–16 focus on the use of evaluation findings. While the UFEC makes important contributions to improving evaluation utility practices, we found it to be difficult to implement in our evaluation practice. Instead, we propose three overarching metrics for evaluation utility that are compatible with the reflective practice model in meta-evaluation and are objective and practical to implement in the context of real-world evaluation.

CONSIDERATIONS IN DEVELOPING EVALUATION UTILITY METRICS (EUMS)

One way to assess whether stakeholders value evaluation processes and products is through stakeholder *actions* (Bundi et al., 2021; Yarbrough, 2017). Process actions are those immediate program improvements that occur as a result of participating in the evaluation process. For example, in our ongoing evaluation of a Center for Translational Research (CTR), we observed stakeholders making immediate changes to their standard operating procedures during the process of defining them (Renger et al., 2021). With respect to products, Yarbrough (2017) notes that stakeholder value is evidenced by the degree to which they *act* on the evaluation report to make program changes.

A common way in which evaluators promote action on evaluation findings is by providing recommendations. Recommendations “go beyond plain evaluation

conclusions [findings]” and are taken to mean “suggestions for specific appropriate actions” (Scriven, 1991, p. 303). Evaluation recommendations integrate evaluation findings and *context* to offer a suggested course of action. We argue that this makes recommendation use a valid metric for evaluation utility; if high-quality recommendations are made, are communicated effectively, and are practical to implement, the extent to which they result in action and the influence of those actions on improving the evaluand provide meaningful measures of evaluation efficacy.

Of course, there are some evaluators who argue that we should not be making recommendations at all (Iriti et al., 2005). Scriven (1991) believes that this reluctance stems from reductionist thinking and the error in reasoning that one cannot “infer conclusions . . . from observations” (p. 374). Reductionists abide by the value-free doctrine, arguing they should not make recommendations because to do so requires integration of their own values in interpreting context, a phenomenon Scriven (1991) coined “valuephobia.” The value-free doctrine has been summarily dismissed by evaluators as impossible for decades (Glass & Ellett, 1980; Scriven, 1991).

Whether a report should include recommendations depends on the negotiated evaluation scope and purpose (Yarbrough et al., 2010). For example, an evaluation whose purpose is knowledge development using a randomized control trial (Henry et al., 2000) should not provide recommendations but instead should conclude by stating whether study hypotheses were supported, leaving the decisions of how to act on those conclusions to others. We are concerned here with those evaluations that do, and should, include recommendations as part of their report. It is our position that evaluators who have planned and executed a high-quality evaluation that produced trustworthy findings should then provide recommendations based on their findings. By extension, then, it is also our position that the reflective practice process should include some assessment as to whether our stakeholders took action regarding our evaluation recommendations. This assessment must, of course, be sensitive to the complex and dynamic factors that influence whether a recommendation is adopted. Failure to adopt a recommendation is not always a reflection on the evaluator or the recommendation itself.

In our literature search, we found no data on the extent to which evaluation recommendations were used to inform decision making. There are countless examples of evaluation reports describing specific recommendations that led to specific program improvements. However, within any single evaluation report it is often unclear how *many* recommendations were made, how many were *considered*, how many were *adopted*, and their *level of influence*. To use a statistical analogy, many evaluation reports provide occasional reference to the numerator (i.e., a recommendation that was adopted), but in the absence of a denominator (i.e., the total number of recommendations made) it is difficult to arrive at conclusions regarding overall evaluation utility. For example, if an evaluation made five recommendations, all were adopted, and all led to actionable changes in program processes or products, then this might be considered evidence of strong evaluation utility. Such findings may serve to reinforce the chosen evaluation approach.

On the other hand, if the same evaluation made 50 recommendations and only five were acted upon, then this should serve as an impetus to reflect on how to improve the evaluation approach to increase utility.

EVALUATION UTILITY METRICS (EUMS)

Based on the aforementioned considerations, we now propose three evaluation utility metrics (EUMs). These are not intended to be definitive or prescriptive but rather a place from which to begin discussions over the need to start developing and using EUMs as a way to reflect on how to improve evaluation practice.

Recommendations considered (EUM_c)

EUM_c refers to the percentage of evaluation recommendations considered and is simply

$$EUM_c = \frac{\text{Evaluation Recommendations Considered}}{\text{Evaluation Recommendations Made}} * 100$$

By “considered” we mean evidence that our stakeholders read and deliberated about the evaluation recommendation. Considering a recommendation is a prerequisite to its being adopted. A high EUM_c provides some initial confidence that the evaluation is on the right track. Associated practices in establishing EUM_c include following up to ensure that recommendations were received and soliciting evaluand response regarding intended action/inaction and related rationale.

Recommendations adopted (EUM_a)

EUM_a refers to the percentage of evaluation recommendations adopted. Adopted is defined as action: changes made as a result of the evaluation process or product (i.e., recommendation), taken by the individual who has the potential to act on it.

$$EUM_a = \frac{\text{Evaluation Recommendations Adopted}}{\text{Evaluation Recommendations Made}} * 100$$

The EUM_a provides direct insight as to whether [Yarbrough et al. \(2010\)](#)’s criterion for stakeholder action was met. As an alternative, we suggest providing decision makers with three response options: not adopted, partially adopted, and fully adopted. In such a case, the multiplier might be adjusted (e.g., 0.0 for not adopted, 0.60 for partially adopted, and 1.0 for fully adopted). Associated practices for establishing EUM_a include follow-up with the evaluand regarding intended actions established during EUM_c. Recommendations that are not adopted must be examined carefully before being included as a reflection of evaluation utility. For example, a recommendation that is not adopted for reasons of resource cost,

timing, or philosophical beliefs may not be valid measures of the utility of the evaluation itself.

Recommendation level of influence (EUM_r)

The EUM_i metric builds on the debate in our field that evaluations of evaluation utility are incomplete without considering the level of evaluation influence (Henry, 2000, 2003; Henry & Mark, 2003; Mark & Henry, 2004). In the words of Alkin and Taut (2002), “the concept of influence adds to the concept of use in instances in which an evaluation has unaware/unintended impacts” (p. 9).

We found evaluating the influence of a recommendation especially challenging. However, the work by Rickinson et al. (2020) provided a clue as to how the level of influence might be tracked and quantified. These authors suggest that findings can be interpreted only in the context of the systems in which they were generated. Therefore, we grounded EUM_i in systems and systems thinking.

Meadows (2008) wrote, “There are no separate systems. The world is a continuum. Where to draw a boundary around a system depends on the purpose of the discussion” (p. 97). For evaluation purposes, we often draw the boundaries at the program level (Wolski, 2020). Therefore, to understand a recommendation’s influence it may be essential to also understand the extent to which that recommendation impacts broader boundaries, that is, other higher system levels. For example, many programs exist within higher-level societal and governmental systems (Friedman, 2009). One such program we evaluated was a Center for Translational Research (CTR). The purpose of a CTR is to provide infrastructure support, in the form of cores, to cancer researchers. Many of our evaluation recommendations focused on improving coordination within a support core (Renger et al., 2020). What we learned was that although a recommendation targeted a specific core, it had the potential to influence other CTR cores, the institution within which the CTR resides, other collaborating institutions within the CTR network, and potentially other CTRs or similar systems elsewhere.

We use a simple scoring system to capture the level of influence: 0 for no influence (recommendation not considered and not adopted), 1 to represent influence within the immediate program boundaries, 2 for the next system level, 3 for the next system level, and so forth. The number of system levels being considered is idiosyncratic to each evaluation and how the boundaries are defined.

USING THE EUMS FOR REFLECTION

There is a common-sense and progressive contingency between the three EUMs. That is, one would track whether a recommendation was adopted only if it was first considered. Similarly, one would track the level of influence only for recommendations that were at least partially adopted.

The EUMs are overarching indicators of evaluation utility and can be used separately or jointly to guide the evaluator in practice reflection. For example, if EUM_c is low, perhaps we need to reflect on our credibility (U1). On the other

hand, if EUM_c is high but EUM_a is low, then we can have greater confidence that the evaluation is on the right track, but perhaps we need to reflect on whether we are accurately capturing the context, such as the individual or cultural stakeholder values (U4). Put more broadly, these three EUMs can serve as a kind of return on investment (ROI) metric for the evaluation itself, thus holding ourselves accountable while at the same time conveying the value of the evaluation to the stakeholders.

REFLECTIONS AND LESSONS LEARNED IN USING EUMS

Negotiating EUM data collection

The data needed to complete EUM metrics comes from the consumers of the evaluation recommendations; the decision makers need to tell us whether they considered a recommendation, adopted it, and its level of influence. The willingness of decision makers to participate in collecting EUM data is itself an indicator that they value the evaluation process (Yarbrough, 2017). However, in our experience, asking decision makers about actions they did or didn't take falls outside of their normal evaluation experience and can make them feel uncomfortable. Therefore, we recommend managing these expectations by negotiating the EUM data-collection process during the evaluation-planning phase (Yarbrough et al., 2010). For example, in the ongoing evaluation of the aforementioned CTR, the principal investigator (PI) agreed to provide feedback via email within two weeks regarding whether a recommendation was adopted. With rare exception, the PI provides recommendation and adoption data within 24–48 hours. In our evaluation of the COVID-19 response, we negotiated similar agreements with public health leadership and received a 100% compliance rate.

Once we learn of the intention to adopt a recommendation, we establish the timeline for its adoption. We then monitor to see whether the recommendation was adopted as intended.

Improving EUMs through recommendation timing

Recommendations are a form of feedback. Two important criteria to consider when providing feedback are timing and frequency (McShane & Von Glinow, 2009). When we began collecting EUM data, we presented our recommendations together in an annual summary report format. Although this is a popular evaluation report format, we have learned several reasons that it is impractical for collecting EUM data and improving our EUMs. First, decision makers simply did not have a large enough block of time to review and detail the actions they took on numerous recommendations. Second, some decision makers were not willing to re-engage in a process they viewed as terminated with the submission of a final report. Third, the time-bound nature of recommendations is susceptible to all the documented challenges involved in a retrospective review (Weinger et al., 2003). Additionally, a retrospective review of recommendations can be uncomfortable.

With the benefit of hindsight, it may be evident that some recommendations that were not acted upon should have been. This may reflect poorly on leadership. Conversely, it may be that some recommendations that were adopted didn't lead to the expected changes, reflecting poorly on the evaluator.

Our solution to these problems was to alter the timing of our recommendations. We have had good success in getting decision makers to provide EUM data when recommendations are presented individually and sequentially. To do this, we use near-real-time evaluation (RTE). RTE has been used primarily in the emergency response sector (Brusset et al., 2010), where waiting until the end of an event or reporting period to present recommendations can have disastrous consequences. The advantage of making near-RTE recommendations is that the decision makers' focus is on a single recommendation when it is contextually and temporally relevant. It is our observation that recommendations made in proximity to their antecedent conditions are more likely to be considered and acted upon. Our observation mirrors psychological research on temporal proximity and behavior change, although whether this is the underlying cause for action requires further study (Bashir et al., 2014). Additionally, responding to any single request takes only a few minutes, decreasing the perceived evaluation burden. If not done thoughtfully, then near-RTE can result in recommendations based on partial data or transient system states that are not borne out over time, presenting a threat to resources and creating "evaluation fatigue." We have found that it is therefore important to be highly selective and deliberate in how often we make recommendations.

DISCUSSION AND CONCLUSION

Holding ourselves accountable is central to our integrity as evaluators (AEA, 2021). This article has focused on one way in which we can hold ourselves accountable: by reflecting on the utility of our evaluation and making necessary changes to improve our practice based on that reflection. As the saying goes, "the proof of the pudding is in the eating," and the ultimate measure of utility is whether stakeholders took concrete actions in response to our recommendations (Yarbrough, 2017).

We have shared three metrics that we found useful in our reflective practice. When we first started tracking EUMs, about two-thirds of our recommendations were considered and less than half were adopted. This caused us to reflect upon our methods and upon *the way* in which we provided recommendations. By negotiating a feedback policy with decision makers, we increased our EUM_c to 100%. By making adjustments to our evaluation approach, for example by better matching the method to the problem coupled with RTE, our EUM_a rose to 81% (Renger et al., 2020).

Certainly, evaluators can engage in reflection in many ways (Smith et al., 2015; van Draanen, 2017). We are not attempting to be prescriptive in what is often more art than science. However, we contend that reflection is incomplete

if we do not consider what stakeholder actions were taken as a result of our recommendations (Yarbrough, 2017). The EUM metrics serve as a starting point for such reflection. Low EUMs might signal the need for a complete overhaul of the evaluation strategy. When we encounter a low EUM, we engage in root cause analysis using the eight utility criteria to guide our inquiry. On the other hand, a high EUM might be cause to replicate the evaluation approaches in the future.

As one reviewer of this article noted, “metrics are imperfect, but they help us understand and further reflect on our practice.” We concur, and we recognize that our EUMs do not consider the relative importance of evaluation recommendations. Should adopting a recommendation to hire an additional staff member be weighted as heavily as one recommending a leadership regime change? Nor do our suggested EUMs factor in adopted recommendations that didn’t lead to the expected changes. However, we need to begin somewhere, and regardless of the metric specifics, if we are to be true to the standards to which we claim to abide, then we need to start evaluating the actions taken in response to our recommendations.

Our implementation of the EUMs also suggests that perhaps the relationship between evaluation use and utility isn’t unidirectional. The conventional wisdom is that utility is a prerequisite to use: If decision makers find the recommendation useful, then they will use it. However, we found that negotiating consideration of the recommendations paves the way, and indeed indirectly forces, the decision makers to make a determination of a recommendation’s utility.

We limited our suggested EUMs to actions taken in response to products, specifically in response to evaluation recommendations. However, building on the work of Yarbrough (2017), we should also consider expanding the EUMs to consider actions taken by being involved in the evaluation process itself.

Yanow (2009) wrote that reflective practice requires “passionate humility.” Beginning to track EUMs can be a humbling experience. It certainly was for us. It isn’t easy to face the reality that your client doesn’t consider or adopt your recommendations, that your recommendations have little or no influence, or that they are even viewed as “a waste of time” (Friedman, 2009). Making changes based on reflection on these metrics requires high evaluator self-efficacy because such reflection, if sincere and mature, will often require that evaluators leave their comfort zone to find better ways of conducting their evaluation.

ACKNOWLEDGEMENT

This practice note was funded by National Institutes of Health Award Number U54GM128729.

REFERENCES

American Evaluation Association (AEA). (2021). *Guiding principles*. https://www.eval.org/Portals/0/Docs/AEA_289398-18_GuidingPrinciples_Brochure_2.pdf

- Alkin, M. C., & Taut, S. M. (2002). Unbundling evaluation use. *Studies in Educational Evaluation*, 29(1), 1–12. [https://doi.org/10.1016/s0191-491x\(03\)90001-0](https://doi.org/10.1016/s0191-491x(03)90001-0)
- Anderson, R. (2004). Intuitive inquiry: An epistemology of the heart for scientific inquiry. *The Humanistic Psychologist*, 32(4), 307–341. <https://doi.org/10.1080/08873267.2004.9961758>
- Bashir, N. Y., Wilson, A. E., Lockwood, P., Chasteen, A. L., & Alisat, S. (2014). The time for action is now: Subjective temporal proximity enhances pursuit of remote-future goals. *Social Cognition*, 32(1), 83–93. <https://doi.org/10.1521/soco.2014.32.1.83>
- Brusset, E., Cosgrave, J., & MacDonald, W. (2010). Real-time evaluation in humanitarian emergencies. *New Directions for Evaluation*, 2010(126), 9–20. <https://doi.org/10.1002/ev.326>
- 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>
- Bustelo, M. (2002). *Metaevaluation as a tool for the improvement and development of the evaluation function in public administrations*. https://evaluationcanada.ca/distribution/20021010_bustelo_maria.pdf
- Canadian Evaluation Society (CES). (2018, October 19). *Competencies for Canadian evaluation practice*. https://evaluationcanada.ca/txt/2_competencies_cdn_evaluation_practice_2018.pdf
- 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>
- Employment and Training Administration (ETA). (2021). *Government performance and results act*. <https://www.dol.gov/agencies/eta/performance/goals/gpra>
- Friedman, M. (2009). *Video introduction to result based accountability*. <https://youtu.be/RtBC89F3Xi4>
- Glass, G. V., & Ellett, F. S., Jr. (1980). Evaluation research. *Annual Review of Psychology*, 31(1), 211–228. <https://doi.org/10.1146/annurev.ps.31.020180.001235>
- Henry, G. T. (2000). Why not use? *New Directions for Evaluation*, 2000(88), 85–98. <https://doi.org/10.1002/ev.1193>
- Henry, G. T. (2003). Influential evaluations. *American Journal of Evaluation*, 24(4), 515–524. <https://doi.org/10.1016/j.ameval.2003.10.001>
- 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>
- Henry, G. T., Mark, M. M., & Julnes, G. (2000). *Evaluation: An integrated framework for understanding, guiding, and improving public and nonprofit policies and programs*. Wiley.
- Iriti, J. E., Bickel, W. E., & Nelson, C. A. (2005). Using recommendations in evaluation: A decision-making framework for evaluators. *American Journal of Evaluation*, 26(4), 464–479. <https://doi.org/10.1177/1098214005281444>
- King, J. A., & Pechman, E. M. (1982). *The process of evaluation use in local school settings (Final Report of NIE Grant 81-0900)*. Orleans Parish School Board (ERIC Document Reproduction Service No. ED 233 037).

- Johnson, K., Greenseid, L. O., Toal, S. A., King, J. A., Lawrenz, F., & Bolkov, B. (2009). Research on evaluation use: A review of the empirical literature from 1986–2005. *American Journal of Evaluation*, 30(3), 377–410. <https://doi.org/10.1177/1098214009341660>
- Mark, M. M., & Henry, G. T. (2004). The mechanisms and outcomes of evaluation influence. *Evaluation*, 10(1), 35–57. <https://doi.org/10.1177/1356389004042326>
- McCormick, E. R. (1997). *Factors influencing the results of evaluation use* (Unpublished Master's thesis). University of Minnesota.
- McShane, S., & Von Glinow, M. (2009). *Organizational behavior: Emerging knowledge and practice for the real world* (5th ed.). McGraw Hill.
- Meadows, D. (2008). *Thinking in systems: A primer*. Chelsea Green.
- Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Sage.
- Patton, M. Q. (2013). *Utilization-Focused Evaluation (UFE) checklist*. https://wmich.edu/sites/default/files/attachments/u350/2014/UFE_checklist_2013.pdf
- Renger, R., Basson, M. D., Hart, G., Van Eck, R., Souvannasacd, E., Renger, J., & Foltsova, J. (2020). Lessons learned in evaluating the infrastructure of a Centre for Translational Research. *Evaluation Journal of Australasia*, 20(1), 6–22. <https://doi.org/10.1177/1035719x20909910>. Medline:34163122
- Renger, R., Renger, J., Basson, M. D., Van Eck, R. N., Renger, J., Souvannasacd, E., & Hart, G. (2021). Using the Homeland Security and Exercise Evaluation Program (HSEEP) building block approach to implement System Evaluation Theory (SET). *American Journal of Evaluation*, 42(4), 586–601. <https://doi.org/10.1177/1098214020986619>. Medline:34966242
- Rickinson, M., Sharples, J., & Lovell, O. (2020). Towards an understanding of quality of evidence use. In S. Gorard (Ed.), *Getting evidence into education: Evaluating the routes to policy and practice* (pp. 218–233). Routledge. <https://doi.org/10.4324/9780429290343>
- Scriven, M. (1991). *The evaluation thesaurus* (4th ed.). Sage.
- Scriven, M. (2007). *Key evaluation checklist (KEC)*. https://evaluationcanada.ca/distribution/20060523_scriven_michael.pdf
- Smith, T. L., Barlow, P. B., Peters, J. M., & Skolits, G. J. (2015). Demystifying reflective practice: Using the DATA model to enhance evaluators' professional activities. *Evaluation and Program Planning*, 52, 142–147. <https://doi.org/10.1016/j.evalprogplan.2015.04.004>. Medline:26051793
- Stufflebeam, D. L. (1974). Meta-evaluation. *Journal of MultiDisciplinary Evaluation*, 7(15), 99–158.
- van Draanen, J. (2017). Introducing reflexivity to evaluation practice: An in-depth case study. *American Journal of Evaluation*, 38(3), 360–375. <https://doi.org/10.1177/1098214016668401>
- Weinger, M. B., Slagle, J., Jain, S., & Ordonez, N. (2003). Retrospective data collection and analytical techniques for patient safety studies. *Journal of Biomedical Informatics*, 36(1–2), 106–119. <https://doi.org/10.1016/j.jbi.2003.08.002>. Medline:14552852
- Wolski, M. (2020). *System requirements: The context and boundary of a system*. <https://michael.wolski.pro/2018/01/system-requirements-the-context-and-boundary-of-the-system/>

Yanow, D. (2009). Ways of knowing: Passionate humility and reflective practice in research and management. *The American Review of Public Administration*, 39(6), 579–601. <https://doi.org/10.1177/0275074009340049>

Yarbrough, D. B. (2017). Developing the program utility standards: Scholarly foundations and collaborative processes. *Canadian Journal of Program Evaluation*, 31(3), 284–304. <https://doi.org/10.3138/cjpe.349>

Yarbrough, D. B., Shula, L. M., Hopson, R. K., & Caruthers, F. A. (2010). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Corwin Press.

AUTHOR INFORMATION

Ralph Renger received his PHD from the University of Calgary. He was a professor for 23 years at the University of Arizona and University of North Dakota, where he taught program and system evaluation. He left academia to pursue his interest in system evaluation. He has over 70 peer-reviewed publications and is widely recognized as an international expert in systems evaluation.

Jessica Renger is a current PhD student in the Evaluation and Applied Research Methods program at Claremont Graduate University. She has extensive evaluation experience both internally in collaboration with the Claremont Evaluation Center and externally as an independent consultant. She has worked on projects nationally and internationally in the health, education, and leadership sectors. She specializes in theory-driven program evaluation, effective interpersonal communication, and health promotion.

Richard N. Van Eck is the founding Dr. David and Lola Rognlie Monson Endowed Chair in Medical Education and Associate Dean for Teaching and Learning at the University of North Dakota School of Medicine and Health Sciences (SMHS), where he plays a leadership role in curriculum evaluation, education scholarship, and faculty development. He is the PI on the SMHS's AMA Innovation grant and on the AMA Accelerating Change in Medical Education Consortium projects on interprofessional telemedicine and gamification in medical education. He is a frequent keynote speaker, presenter, and author on serious games, including TEDx Manitoba and SXSW.

Marc D. Basson received his MD from the University of Michigan, his PhD in Experimental Pathology from Yale, and his MBA from Wayne State University. He is the Senior Associate Dean for Medicine and Research at the University of North Dakota and the Director of the DaCCoTA.

Jirina Renger is a graduate from the University of Arizona, College of Public Health. She currently serves as a core faculty member in the PhD Public Health program in the College of Health Professions at Walden University. Her area of specialization is theory-driven evaluation and program theory reconstruction.