New Insights for Tracking and Reporting Milestones

Jessica Renger
Kaiser Permanente, Pomona, California, United States

Ralph Renger
Just Evaluation Services, LLC, Vail, Arizona, United States

Richard N. Van Eck and Marc D. Basson
School of Medicine and Health Sciences, University of North Dakota, Grand Forks, North Dakota, United States

Jirina Renger
Walden University, Minneapolis, Minnesota, United States

Abstract: *Despite being a common evaluation metric, milestones are poorly defined, and guidance for their use is underdeveloped in the evaluation literature. This practice note attempts to address this oversight by providing evaluation-specific guidance for tracking and reporting milestones. Using a case study, we share our initial approach to tracking and reporting milestones, as well as how critical reflection guided by the evaluation standards provided us with key insights that allowed us to make improvements to our initial process. We hope this practice note acts as a springboard for elevating the critical dialogue surrounding milestones in the evaluation literature.*

Keywords: *milestones, monitoring, evaluation standards*

Résumé : *Bien qu'il s'agisse d'une mesure d'évaluation commune, les jalons sont mal définis et leur cadre d'utilisation est sous-développé dans la littérature en matière d'évaluation. Cette note de pratique tente de combler cette lacune en offrant des lignes directrices particulièrement conçues pour l'évaluation, en vue de suivre et de présenter les jalons. À l'aide d'une étude de cas, nous faisons part de notre approche initiale au suivi et à la présentation des jalons, de même que de la façon dont une réflexion critique orientée par les normes d'évaluation nous a permis d'arriver à des conclusions clés pour l'amélioration de notre processus initial. Nous espérons que cette note de pratique servira de tremplin pour élever le dialogue critique entourant les jalons dans la littérature en matière d'évaluation.*

Mots clés : *jalons, surveillance, normes d'évaluation*

Corresponding author: Jessica Renger, 1340 Hillcrest Drive, Pomona, CA, United States, 91768; Jessica.A.Renger@kp.org

 Canadian Journal of Program Evaluation / La Revue canadienne d'évaluation de programme
37.2 (Fall / Automne), 273–282 doi: 10.3138/cjpe.74487

INTRODUCTION

The purpose of this practice note is to share the authors' insights regarding tracking and reporting one of the most common, but surprisingly ill-defined, evaluation metrics—the milestone. *Merriam-Webster* (2021) defines a milestone as "an important point in the progress of something." If that seems too vague to be of value, we agree. Unfortunately, in the evaluation literature, there is not much more clarity, with milestones defined variously as steps along the way to goals, objectives, or aims (Green & Kreuter, 2005). In some cases, milestones are conflated with goals, objectives, or aims themselves. Even the most comprehensive repository of evaluation terminology, The Evaluation Thesaurus (Scriven 1991), fails to mention or define the term *milestone*.

The literature offers guidance pertaining to writing, tracking, and reporting milestones in documents provided by those who commission evaluations (e.g., Edgar, McLean, Hogan, Hamstra, & Holmboe, 2020). However, there is considerable variability in how milestones are defined, and the definitions often lack precision. For example, milestones have been defined variously as "the articulation of the level of performance expected" (Accreditation Council for Graduate Medical Education [ACGME], 2021, p. 3), "assurance that [staff] attain appropriate educational goals" (ACGME, 2021, p.3), "distinct points in time . . . to determine if the program is ready to progress to the next steps" (Vaneman, Carlson, & Wolfgeher, 2019, p. 18), and "an essential component of program evaluation . . . to identify trends within a program, between programs within an institution, and between institutions . . . for ongoing program improvement" (Heath, Edgar, & Guralnick, 2021, p.122).

The lack of consensus regarding the definition of milestones can cause confusion among evaluators about what to track, when to track it, and how to best to report information gathered, which in turn limits evaluators' ability to track and report milestones. This lack of clarity also forces evaluators using the milestone as an evaluation metric to constantly redefine milestones and their associated criteria in the context of their projects, creating unnecessary redundancy and wasted time.

Why is such a common concept in our field so poorly defined in the evaluation literature? Why have we not studied milestones more carefully? Perhaps some consider milestones a form of monitoring rather than evaluation (Scriven, 1991). As such, evaluators may feel that milestones are defined "enough" and can be monitored solely as a quality-control measure rather than a measure of the health and function of the program or system itself. However, to do so ignores a potentially critical role for milestones in evaluation. Milestones are distinct from related concepts such as indicators or outputs because they exist in a continuum, are time bound, and provide a path forward for project progression. Due to this distinction, milestones serve a unique evaluation purpose and can play an important role if they are aligned carefully with the evaluation process. Milestone guidance written by evaluators, for evaluators is sorely needed.

In this practice note, we provide evaluation-specific guidance for tracking and reporting milestones—not as a definitive solution, but to illustrate how such a solution could elevate the value of milestones in evaluation. Through doing so, we

 doi: 10.3138/cjpe.74487

argue that milestones should be viewed as falling along a continuum of evaluation purposes serving primarily an oversight and compliance function (Mark, Henry, & Julnes, 2000). We utilize our evaluation of a Center for Translational Research (CTR) as a case study to frame the discussion—first explaining our approach to tracking and reporting milestones in this context and then highlighting how reflective practice informed by the evaluation standards provided us with numerous key insights and lessons learned (Yarbrough, Shulha, Hopson, & Caruthers, 2011). It is important to note that while our experiences with milestones occurred in a real-world context, in adhering to the Transparency and Disclosure proprietary standard (P5), we have substituted actual data with fictitious data (Yarbrough et al. 2011).

CASE STUDY: THE EVALUATION OF A CTR

Case context

A CTR is a National Institutes of Health–funded initiative with two overarching goals: to ensure CTR investigators have the support they need to obtain additional funding for clinical and translational research (National Institutes of Health, 2017, p. 2) and to support CTR investigators in translating findings from basic research to clinical practice, effectively moving along the translational science continuum (University of North Dakota, 2021). To achieve these two goals, a CTR is composed of smaller operating bodies, or cores, each tasked with helping investigators develop research skills and accomplish tasks necessary to advance their projects. A typical CTR is made up of seven cores: administrative, professional development, biostatistics epidemiology and research design, pilot projects, community engagement and outreach, clinical research resources and facilities, and tracking and evaluation.

As members of the tracking and evaluation core, we were responsible for assessing the efficiency with which CTR cores interacted when providing support to investigators, as well as the extent to which each CTR core was progressing towards their stated goals. To do this, we asked each core to create a list of milestones and share that information with us so we could periodically assess their progress. Milestones were chosen as the evaluation metric because the term *milestone* was used by the funding organization and milestone data was required for reporting purposes. Each core's milestone progression was shared with leadership annually to make adjustments as needed. This information was collected for an initial grant cycle and again for a later grant renewal. We initially reported on 82 milestones across cores for the first grant cycle and 67 milestones for the grant renewal. Our initial strategy for gathering and reporting milestones for the first grant cycle as well as our reflections and revisions to this process for the grant renewal are discussed in the following.

Initial milestone tracking and reporting process

One challenge was that despite being a required evaluation element, the funding agency provided no guidance pertaining to defining or tracking milestones. Given this absence of milestone guidance, and the lack of information on milestones in the evaluation literature, we turned to the program evaluation standards to help direct us in tracking and reporting milestones (Yarbrough et al., 2011). We began by tracking milestones in a straightforward manner in line with the Practical Procedures feasibility standard (F2), which states "evaluation procedures should be practical and responsive to the way the program operates" (Yarbrough et al., 2011, p. 87). After we received the milestone list from each core, we checked in annually to determine whether the stated milestones were met. We did this using a simple checklist format to ascertain whether a milestone was met (Gawande, 2009).

To report milestones, we once again relied on the Practical Procedures feasibility standard (F2) and presented a simple annual report that listed the milestones for each core and indicated whether each was completed, in progress, or had yet to be started. We presented a snapshot of this information in a pie chart to visualize and compare each core's progress toward their milestones. If a milestone was not completed as expected, we then followed up with the leadership personnel in each core and engaged in root cause analysis (Coşkun, Akande, & Renger, 2012) to determine why progress was halted and work to resolve underlying issues.

Reflecting on our initial milestone tracking and reporting process

As we were preparing for the CTR grant renewal, we recognized that engaging in a systematic reflective practice process would be essential to ensure our evaluation approach was responsive to the CTR context and well positioned to support ongoing learning and improvement (Canadian Evaluation Society, 2010; Tovey & Skolits, 2021). Our reflective practice process involved (a) evaluation team members engaging in individual reflection on the strengths and weaknesses of our evaluation approach and (b) holding weekly team meetings to discuss insights gleaned from individual reflections, compare our processes against the guidelines offered by the Program Evaluation Standards, and discuss pathways toward improvement. As a result of our reflective process, we identified six key insights regarding tracking and reporting milestones that were then integrated into our approach for grant renewal. We share those insights and the resulting changes here.

Insight #1: Milestone writing should not be taken for granted

At the project onset, we assumed that milestone generation would be a rather straightforward process that would only require tracking with minimal oversight. Thus, as an evaluation team, we did not engage in a thorough quality assurance process to examine each milestone in depth. As a result, there was no standardization of the initial milestones generated. This resulted in substantial variation in the nature and quality of the milestones written by stakeholders, reflecting the variation found in the literature regarding milestones. These variations have a significant impact on the system and the evaluation, as a result (Renger, Foltysova, Ienuso, Renger, & Booze, 2017). For example, many of the milestones were ongoing project activities, or larger goals, that did not include specified end points that could be tracked and reported (e.g., "solicit proposals for funding"). The absence of a clearly defined end point made it impossible to determine whether such a milestone was on track. In addition, milestone formats varied greatly by core, making it difficult to compare progress across the CTR. Thus, we found that for our milestone reporting to meet the Reliable Information accuracy standard (A3), we had to provide our cores with more specific definitions, criteria, and guidance regarding the milestone-generation process for the grant renewal.

Insight #2: Milestones must include or suggest a schedule for completion

Funding was for a 5-year period. The success of the CTR hinged on establishing the critical infrastructure (i.e., establishing partnerships, hiring staff, creating a web interface, etc.) in year 1. Thus, the chronology and sequencing of milestones were especially germane but lacking in several of our milestones. In the initial grant cycle, this information was missing for 41 of our 82 initial milestones (50%). Thus, in the grant renewal, we worked more closely with our staff using the SMART (Specific, Measurable, Attainable, Relevant, Time-bound) criteria (Bjerke & Renger, 2017) and honing in on the "T" time-bound criterion. Our reflective practice approach also revealed that milestones were heavily focused on monitoring infrastructure progress and that there were significant lacunae in milestones related to moving forward the CTR mission (e.g., timely feedback to awardees regarding their proposals).

Insight #3: Consider milestone pacing

In our reflection, we realized a distinguishing feature of milestones is that they exist in a continuum; every milestone is connected to other related milestones. In essence, milestones can be conceptualized as stepping stones across a river, leading to a final goal or destination. Two necessary components of this continuum are sequencing and pacing. We found that project stakeholders sequenced their milestones correctly—that is, their milestones were well ordered and followed a logical progression. Thus, we focused our efforts more heavily on the pacing of milestones, examining not only *when* a milestone was due but when it was due *relative* to other milestones.

The original purpose of milestones was to help travelers stay on schedule by allowing them to estimate the time to their destination by extrapolating from the time it took to travel between milestones. Extrapolation was possible because the distance between milestones was at fixed intervals. In our CTR, however, we noticed the pacing of milestones was variable, rather than fixed. Some of the milestones occurred closer in time, making it easier to stay on track and identify potential roadblocks or bottlenecks in a timely fashion. In other cases, however, milestones occurred farther apart in time, making it harder to identify potential problems or threats to achieving them. Thus, for the grant renewal, we needed to encourage program staff to consider the timing between milestones and provide pacing recommendations as appropriate to ensure each milestone had the greatest chance of being successfully completed.

Insight #4: Milestones should be iterative

With the conception of milestones as "steps along the way" toward a goal, when one goal and its associated milestones are complete, it logically follows that another goal with milestones should be created to maintain project growth. However, this is not how milestones are discussed in the literature, and it was our experience that the project stakeholders did not view milestones as ongoing and iterative, but rather as summative and fixed. For them, milestones were outputs, something to be done cyclically (annually), often without reference or connection to prior goals and milestones. Therefore, when a milestone was met, it was seldom replaced with a new milestone, and new milestones sometimes lacked continuity with prior work. For example, many milestones pertained to hiring new CTR staff members. If a staff member was hired in the first year, then the milestone sequence was terminated. No additional milestones were created pertaining to the tasks the new staff member would undertake, how the staff member would contribute to expanding capacity in desired areas, or other potentially relevant undertakings that would continue the milestone sequence.

To avoid milestone stagnation for the grant renewal, our evaluation team needed to encourage project staff to set new goals and associated milestones when all the milestones of a goal were met. Encouraging program staff to engage in the iterative generation of milestones is in line with the Meaningful Processes and Projects utility standard (U6), which states that "evaluations should construct activities, descriptions, and judgements in ways that encourage participants to rediscover, reinterpret, or revise their understandings and behaviors" (Yarbrough et al., 2011, p. 51).

Insight #5 Milestones should be tracked and reported in a timely manner

As is the case in many evaluations, our initial approach was to report on milestone achievements annually. However, in keeping with the Timely and Appropriate Communicating and Reporting utility standard (U7), we realized that the yearly reporting interval was not sensitive enough to be able to provide decision makers with timely information. For example, a milestone completed 11 months behind schedule (due in the first month of the project report period but completed in the 11th month) was viewed as equally successful as a milestone completed on time. Our reporting strategy had to mirror the milestone schedule, but we also realized the need to strike a balance between providing timely and sufficiently frequent feedback and overreporting and overwhelming (Lurie & Swaminathan, 2009). Tracking and reporting milestones in real-time was too costly and violated the Practical Procedures feasibility standard (F2). Thus, we settled on a near-real-time (NRT) tracking and reporting strategy, agreeing to provide updates on milestones within four weeks of their scheduled completion date. Of note, the four-week interval was chosen in our case as it coincided with regularly planned leadership meetings. Adjustments to this interval should be made based on the nature of the evaluation project and to reduce reporting burden to key stakeholders.

Insight #6: Milestones should be reported with explicit reference to their schedule for completion

To better report milestones, we referenced three utility standards: U5 (Relevant Information), U6 (Meaningful Processes and Products), and U7 (Timely and Appropriate Communicating and Reporting) (Yarbrough et al., 2011). Due to the variable nature of milestone pacing, we realized that for reporting purposes, the relevant and meaningful information was not simply whether a milestone was completed but whether it was completed according to schedule (Harned, Schmidt, Korslund, & Gallop, 2021). Therefore, we needed to plot the planned milestone schedule against the actual milestone completion timeline.

As an example, numerous milestones were implemented to ensure the timely hiring of new CTR staff. We first plotted the planned milestone schedule for a project hire (shown as the solid line in Figure 1). Figure 1 shows that there are eight milestones covering a 12-month hiring period, the first of which is “identify a hiring need” and the last of which is “finalize hiring.” As the figure illustrates, the timing between milestones is variable, with some milestones occurring in closer succession (i.e., approval to hire and job descriptions) than others (i.e., advertise position and review applications).

Next, we plotted when the milestones were actually completed. This is shown as a dotted line in Figure 1. From the figure we can see that some milestones were completed on schedule (i.e., identify hiring need, approval to hire, advertise position) or ahead of schedule (i.e., develop a job description), whereas others took longer to complete than scheduled (i.e., review applicants, interview candidates), pushing the timeline back and delaying the hiring process.

Reporting in this format can greatly increase the utility of milestones in two ways. First, it allows the reader to easily see which milestones are causing a delay and make revisions as necessary to ensure similar milestones in the future progress according to schedule. Second, this reporting format is easily scalable and can be used to report

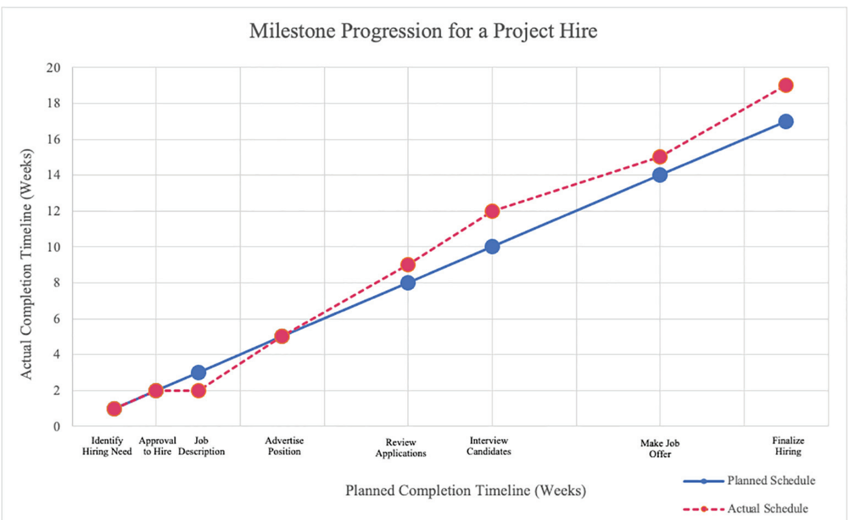

Figure 1. Milestone progression for a project hire
Source: Author’s illustration.

milestones at the smallest and largest levels of a program or system, making it an ideal choice for those wishing to use a standard reporting method in multiple contexts.

CONCLUSION

This practice note is intended as a first step toward a broader discussion of the use, tracking, and reporting of milestones in the evaluation literature. Through our experiences evaluating a CTR and engaging in reflective practice informed by the evaluation standards, we shared key insights regarding tracking and reporting milestones that we hope will be useful for other evaluators engaging in this work. The insights shared are not the totality of what can be learned from the milestone tracking and reporting process but do reflect key points of consideration we feel have been relatively underdeveloped in the current evaluation literature. The great number of insights generated from this work also serves to highlight the importance of evaluators engaging in regular and systematic reflective practice and sharing their findings with the broader evaluation community.

Through these insights we suggest milestones can be defined as "steps along the way to a goal that exist in a continuum, are iterative, and are time bound." Although this definition is largely informed by our experiences with this case, we believe it has potential for broader applicability within the field of evaluation and can act as a springboard for a more meaningful conversation about how the field of evaluation should articulate and define milestones. It is our hope that this note will serve as a jumping-off point for evaluators to more carefully consider how they define, track, and report milestones as a means of contributing to improved evaluation practice.

REFERENCES

Accreditation Council for Graduate Medical Education (ACGME). (2021). *The next steps in the outcomes-based accreditation project*. Retrieved from https://www.acgme.org/Portals/0/PFAssets/bulletin/bulletin5_08.pdf

Bjerke, M. B., & Renger, R. (2017). Being smart about writing SMART objectives. *Evaluation and program planning*, *61*, 125–127.

Canadian Evaluation Society. (2010). *Competencies for Canadian evaluation practice*. Retrieved from https://evaluationcanada.ca/txt/2_competencies_cdn_evaluation_practice_2010.pdf

Coşkun, R., Akande, A., & Renger, R. (2012). Using root cause analysis for evaluating program improvement. *Evaluation Journal of Australasia*, *12*(2), 4–14.

Edgar, L., McLean, S., Hogan, S. O., Hamstra, S., & Holmboe, E. S. (2020). *The milestones guidebook* (version 2020). Retrieved from Accreditation Council for Graduate Medical Education website: https://www.acgme.org/globalassets/milestonesguidebook.pdf

Gawande, Atul. (2009) *The checklist manifesto: How to get things right*. Metropolitan Books. New York, NY.

Green, L., & Kreuter, M. W. (2005). *Health program planning: An educational and ecological approach* (4th ed.). New York: McGraw Hill.

Harned, M. S., Schmidt, S. C., Korslund, K. E., & Gallop, R. J. (2021). Does adding the Dialectical Behavior Therapy Prolonged Exposure (DBT PE) protocol for PTSD to DBT improve outcomes in public mental health settings? A pilot nonrandomized effectiveness trial with benchmarking. *Behavior therapy*, *52*(3), 639–655.

Heath, J. K., Edgar, L., & Guralnick, S. (2021). Assessment of learning, for learning: Operationalizing milestones data for program level improvement. *Journal of Graduate Medical Education*, *13*(Suppl. 2), 120–123. https://doi.org/10.4300/JGME-D-20-00849.1

Lurie, N. H., & Swaminathan, J. M. (2009). Is timely information always better? The effect of feedback frequency on decision making. *Organizational Behavior and Human Decision Processes*, *108*(2), 315–329.

Mark, M. M., Henry, G. T., & Julnes, G. (2000). *Evaluation: An integrated framework for understanding, guiding, and improving policies and programs*. San Francisco, CA: Jossey-Bass.

Merriam Webster (2021). *Milestone*. Retrieved from https://www.merriam-webster.com/dictionary/milestone

National Institutes of Health. (2017). *Institutional Development Award (IDeA) program infra- structure for clinical and translational research (IDeA-CTR)* (PAR-17-304). Retrieved from https://grants. nih.gov/grants/guide/pa-files/par-17-304.html

Renger, R., Foltysova, J., Ienuso, S., Renger, J., & Booze, W. (2017). Evaluating system cascading failures. *Evaluation Journal of Australasia*, *17*(2), 29–36.

Scriven, M. (1991). *Evaluation Thesaurus* (4th ed.). SAGE. Newbury Park, CA.

Tovey, T. L. S, & Skolits, G. J. (2021, December). Conceptualizing and engaging in reflective practice: Experienced evaluators' perspectives. *American Journal of Evaluation*. https://doi.org/10.1177/1098214020983926

University of North Dakota (2021). *DaCCoTA: What is Clinical Translational Research?* Retrieved from https://med.und.edu/daccota/

Vaneman, W., Carlson, R., & Wolfgeher, C. (2019). *Defining a model-based systems engineering approach for milestone technical reviews*. Acquisition Research Program (NPS-SE-20-016). Retrieved from https://dair.nps.edu/handle/123456789/4168

Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). SAGE. Thousand Oaks, CA.

AUTHOR INFORMATION

Jessica Renger, MA, works as an evaluator in the National Office of Community Health for Kaiser Permanente and is an advanced PhD student and Adjunct Professor in the Evaluation and Applied Research Methods program at Claremont Graduate University. She specializes in theory driven program evaluation, systems thinking, effective interpersonal communication, and health promotion.

Ralph Renger, PhD, was a tenured professor at the University of Arizona and the University of North Dakota where he taught graduate-level program and system evaluation. He has over 80 peer-reviewed publications focused primarily on bridging theory to practice. He is now an independent consultant focused mainly on evaluating complex interventions

operating and functioning as systems. His book on the same topic is scheduled to be released in the Fall of 2022.

Richard N. Van Eck, PhD, 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. He has served as Co-PI, evaluator, game design consultant, and/or advisory board member for 12 grants from the American Medical Association, NSF, DoE, HRSA, and DARPA and has written and co-authored two book chapters on evaluation, including "Evaluating Coaching Programs" and "A tale of inherent integration: Process methodology in support of a shared purpose to investigate videogame programming in libraries," an invited article for a special issue on methodology in *Research in the Schools*, and four articles on evaluation of the DaCCoTA Center for Translational Research.

Marc Basson, MD, PhD, is Senior Associate Dean for Medicine and Research and Professor of Surgery, Biomedical Sciences, and Pathology at the University of North Dakota School of Medicine and Health Sciences as well as Principal Investigator for the DaCCoTA Center for Clinical and Translational Research.

Jirina Renger, DrPH, is a graduate from 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 Sciences and Public Policy at Walden University. Her area of specialization is Theory-Driven Evaluation and Program Theory Reconstruction.