

Perspectives on Quality

A framework for learning about improvement: embedded implementation and evaluation design to optimize learning

DANIKA BARRY¹, LEIGHANN E. KIMBLE², BEJOY NAMBIAR^{3,4},
GARETH PARRY⁵, ASHISH JHA⁶, VIJAY KUMAR CHATTU^{7,8},
M. RASHAD MASSOUD², and DON GOLDMANN⁵

¹Harvard Medical School, 25 Shattuck St, Boston, MA 02115, USA, ²USAID Applying Science to Strengthen and Improve Systems (ASSIST) Project, Quality & Performance Institute, University Research Co., LLC, 5404 Wisconsin Ave, Suite 800, Chevy Chase, MD 20815, USA, ³Institute for Global Health, UCL, 30, Guilford Street, London WC1N 1EH, UK, ⁴Academy of Medical Sciences, Malawi University of Science and Technology (MUST), 6Km Off Mugabe Highway, Along Goliati Road – Near Ndata Farm, P.O Box 5196, Limbe, Malawi, ⁵Institute for Healthcare Improvement, Cambridge, 53 State Street, 19th Floor, Boston, MA 02109, USA, ⁶Harvard T.H. Chan School of Public Health, 158 Longwood Ave, Boston, MA 02115-5810, USA, ⁷Public Health and Primary Care Unit, Faculty of Medical Sciences, The University of the West Indies, Mona, Kingston 7, Jamaica, and ⁸School of Global Health & Bioethics, EUCLID University, Trinidad and Tobago

Address reprint requests to: Leighann E. Kimble, USAID Applying Science to Strengthen and Improve Systems (ASSIST) Project, Quality & Performance Institute, University Research Co., LLC, 5404 Wisconsin Ave, Suite 800, Chevy Chase, MD 20815-3594, USA. Tel: +1-301-828-3541; Fax: +1-301-941-8427; Email: lkimble@urc-chs.com

Editorial Decision 5 January 2018

Abstract

Improving health care involves many actors, often working in complex adaptive systems. Interventions tend to be multi-factorial, implementation activities diverse, and contexts dynamic and complicated. This makes improvement initiatives challenging to describe and evaluate as matching evaluation and program designs can be difficult, requiring collaboration, trust and transparency. Collaboration is required to address important epidemiological principles of bias and confounding. If this does not take place, results may lack credibility because the association between interventions implemented and outcomes achieved is obscure and attribution uncertain. Moreover, lack of clarity about what was implemented, how it was implemented, and the context in which it was implemented often lead to disappointment or outright failure of spread and scale-up efforts. The input of skilled evaluators into the design and conduct of improvement initiatives can be helpful in mitigating these potential problems. While evaluation must be rigorous, if it is too rigid necessary adaptation and learning may be compromised. This article provides a framework and guidance on how improvers and evaluators can work together to design, implement and learn about improvement interventions more effectively.

Key words: improvement, learning, complex adaptive systems, implementation, delivery

Introduction

As a result of discussions between 60 participants from 22 countries during the Salzburg Global Seminar Session—565 ‘Better Health

Care: How do we learn about improvement?’ [1–3], the Framework for Learning about Improvement and the Evaluation Continuum were developed to guide improvement program design, implementation

and evaluation. This paper will introduce the Framework and the Evaluation Continuum that were developed by participants during the Salzburg Global Seminar—Session 565 and explain how it can facilitate collaboration between evaluators and improvers. We emphasize the importance of this collaboration through embedded evaluation design to create rigorous, adaptive evaluation and improvement program designs that are fit-for-purpose that maximize learning to better understand attribution and generalizability of results. In short, in this article, we will answer the following questions:

- How can we design improvement programs to make evaluation better? and
- How do we evaluate improvement programs to capture the learning and value of these programs?

The Evaluation Continuum

As the participants discovered through conversations, among the various evaluation models available [4, 5], there is a continuum of potential models specifying when and how improvers and evaluators interact and the degree to which evaluation is embedded within or independent of the improvement design and implementation (Fig. 1). Along this continuum, we refer to embedded designs as having evaluators within the improvement activity itself, collaborating, communicating and potentially collocating with the improvement team. The Evaluation Continuum shown in Fig. 1 was developed by Salzburg

Global Seminar—Session 565 participants to demonstrate the trade-offs between maximizing objectivity and enhancing learning.

Pure external evaluation with a fixed, unchanging improvement program design exists on one end of the Evaluation Continuum, as shown in Example 1 in Fig. 1. In the interest of maintaining the objective, unbiased evaluations, there has been a tendency for evaluators to follow Example 1, developing and executing their evaluation plan in isolation from the improvement design and implementation team [4–7]. Although the pure external evaluation is more objective, the trade-off is a lack of understanding of the iterative and adaptive nature of the improvement program. While the evaluator in Example 1 may be able to determine whether the improvement worked, they often lack the in-depth understanding of the interaction between theory of change and context of the improvement program, as well as any changes that may have been made during the improvement program implementation. Lacking the information needed to determine how and why the improvement did or did not work, evaluators are limited in the feedback they can provide to improvers.

Moving along the Evaluation Continuum in Examples 2–4, evaluators increasingly collaborate and communicate more with improvers, providing advice and insights while capturing adaptations to the protocol and implementation plan and modifying their evaluation accordingly. In an evaluation model with external evaluators as in Example 1, improvers often voice concern that evaluators, ‘didn’t measure what we actually did’. In Examples 2–4, an adaptive improvement design is used. The difference along the spectrum in

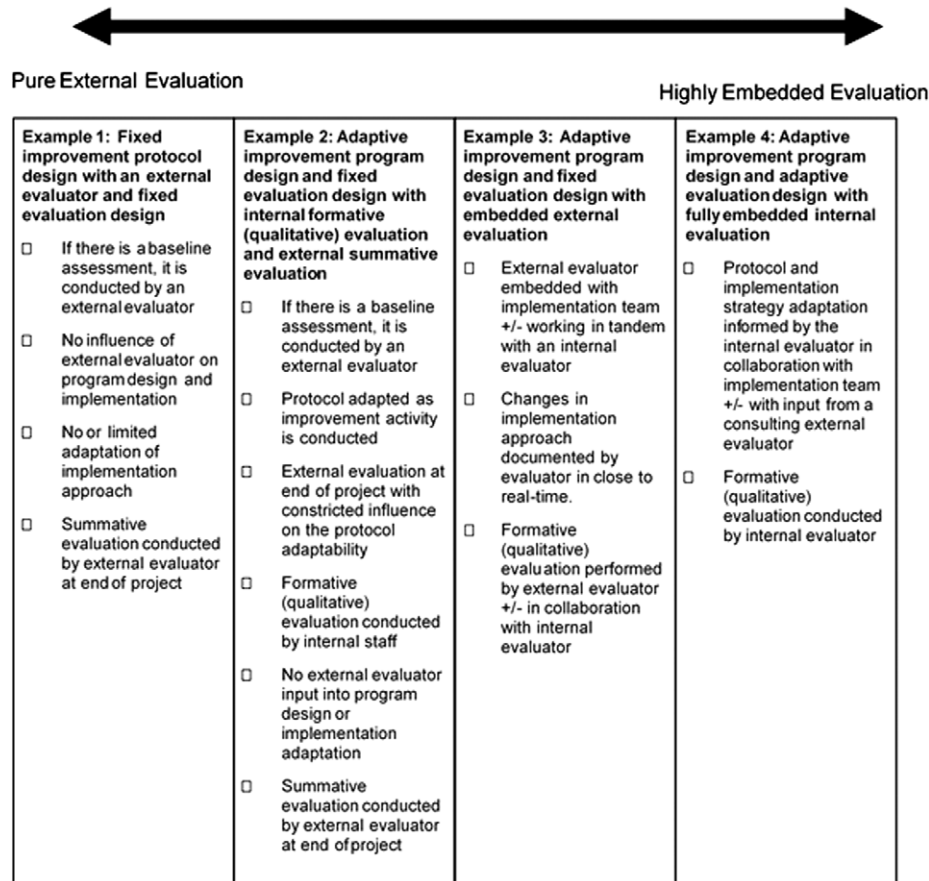


Figure 1. The evaluation continuum.

Examples 2–4 is in whether the improvement design, evaluation design or both, are adaptive.

In Example 2, while the improvement program design adapts throughout the course of the improvement, the evaluation design remains fixed. As mentioned previously, a fixed external evaluation design provides a more objective evaluation. However, fixed evaluation design does not necessarily account for changes in the improvement program design. In Example 2, because the improvement program design is adaptable, a fixed evaluation design would not account for the changes made to the improvement program design.

In Example 3, the improvement program design is again adaptable but includes an internal evaluator who is embedded in the project, working with the external evaluator. With closer collaboration, the internal evaluator can document and communicate information about the improvement program, and any changes made to it, to the external evaluator. The external evaluator in this example may or may not decide to consider the adaptations made in the improvement design in making changes to their own evaluation design.

On the other end of the Evaluation Continuum, as shown in Example 4, is highly embedded evaluation. Highly embedded evaluation includes close collaboration and communication between evaluators and the improvement team, with the evaluator potentially collocated with improvers. Example 4 uses an adaptable improvement program design and an adaptable evaluation design, which feedback into one another as changes are made in the improvement program.

Evaluations on the ‘highly embedded evaluation’ end of the spectrum are less objective than those in Examples 1–3 because the evaluator is internal to the improvement program. However, the learning and understanding about the improvement in a highly embedded scenario are increased, allowing for the evaluator to provide ongoing information and feedback to the improvement. The feedback from evaluators can then be used by improvers to adapt the improvement program design to yield better results.

Embedded evaluation design requires evaluators to interact with improvers at the design level of the improvement initiative, as well as with improvers who are on the ground coaching teams and facilities in implementing the improvement. There is no single ‘correct’ model; the design should be fit-for-purpose and reflect the expectations of the audience for the evaluation [7–9]. For example, a governmental agency may demand a model that maximizes objectivity and assesses quantitatively whether the intervention ‘worked’. Health care delivery system leaders may favor an evaluation that maximizes qualitative learning and adaptation to context so that they will be confident that the intervention can be more widely implemented and scaled up.

Overall, there was consensus among the seminar participants that the field of improvement should move towards more highly embedded models for evaluation and improvement design. Embedded evaluation allows for closer communication, coordination and potential co-location of implementers and evaluators. Embedded evaluations tell us not just whether an improvement program ‘worked’ but also provides information about why an improvement program did or did not work in consideration of the context and theory of change of the improvement program.

Understanding how and why an improvement program worked or did not work generates knowledge that can be shared for other improvement programs. A transition towards embedded evaluations would transform the way we think about the roles of implementers and evaluators—shifting the paradigm towards a ‘marriage’ between evaluation and implementation. We believe that increased collaboration, feedback, communication, transparency and trust between

improvers and evaluators will improve both program and evaluation design and optimize learning, credibility and impact.

Introducing the Framework for Learning About Improvement

As shown in Figure 2, improvement efforts start with a specific objective for the benefit of stakeholders. The rationale for choosing this objective should be described—for example, the existence of gap between evidence and practice for hypertension screening and control [10]. The objective needs to be achieved in a specified time-period within the constraints of a specific context at a given point in time. The theory of change to achieve the objective includes the scientific basis for ‘what’ is to be done together with the practical implementation or ‘how’ it is to be done.

Once the theory of change is developed, with the most detail possible, improvers and evaluators must decide where they want to position themselves on the continuum portrayed Fig. 1. In a more highly embedded and adaptive evaluation design as in Examples 3 and 4, improvers can develop a program design that feeds into the evaluation design [11]. With a clear understanding of the context and theory of change of the improvement program design, evaluators can design an evaluation that both determines whether the improvement worked and provides an understanding of the relationship between the context and the theory of change of the improvement. Evaluators can then provide this feedback to improvers so that the improvement design can be further adapted to obtain better results. The bi-directional arrows of the Framework for Learning about Improvement (Fig. 2) are meant to denote the iterative nature of the interaction between the parts of the framework.

Designing Improvements to Make Evaluation Better and Promote Learning

Improvement design should document and describe the following: (i) the existing processes of care; (ii) the shortfalls of the existing processes of care; (iii) the interaction between the shortfalls of the

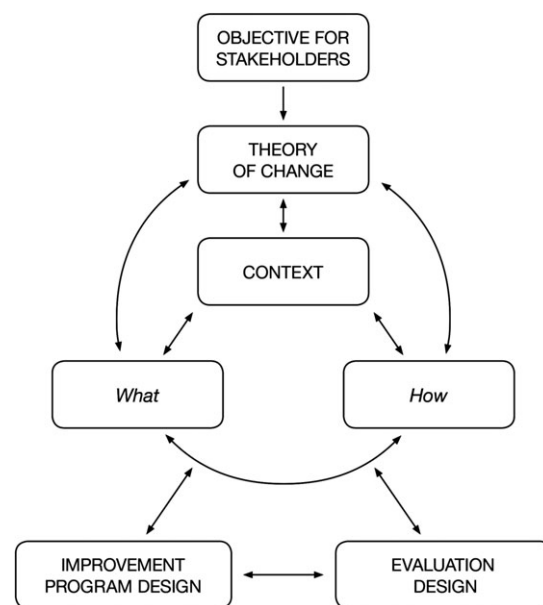


Figure 2. Framework for learning about improvement.

existing processes of care with the context; (iv) the improvement approach; and (v) specific changes made to the processes of care through the improvement approach [12].

The strategy for scale-up and spread of the program can also be included in the theory of change, as well as strategies to increase the likelihood that improvement will be sustainable, along with a follow-up plan to ensure that implementation activities can be maintained and that improvements in key processes and outcomes are sustained. The costs and resources involved in implementation should be captured as well as estimates of the costs and resources required to sustain and spread the interventions.

Evaluations That Enhance Learning and Facilitate Improvement

In addition to designing evaluations to enhance learning and facilitate improvement, general best practices must be followed in designing an evaluation of an improvement program. A key guiding principle of evaluations is to measure as close to the outcome level as possible rather than measuring at the process level alone. However, this may be difficult because the anticipated outcomes are relatively rare (i.e. it is not feasible to power the program to detect a significant change in the outcome), occur beyond the period of study, or are difficult to capture (e.g. patient reported outcomes). Process measurement is easier to justify when abundant evidence suggests that the processes being measured are tightly linked to the outcomes of interest. For example, reliable administration of medications to control blood sugar, lipids and blood pressure (a process measure) almost certainly will have an impact on complications of diabetes (outcome measures)—improvements in outcomes that will not be seen in a short time frame. The evaluation plan also should include an examination of heterogeneity (variation) in improvement among the implementation sites (wards, clinics, hospitals, health systems). Learning, and possibly generalizability, can be enhanced by understanding what worked where and why. For example, contextual factors may have resulted in disappointing results in some places, whereas mitigation of such factors in other sites facilitated improvement.

As with any program that collects data, the evaluation must have a plan for reliable data validation. Methods accounting for time and strength of causal inference should be used. This includes displaying data in a time-series manner. For causal inference, it is helpful to have a comparison group. The comparison group used can either be similar to the program group or a randomized control group to account for secular trends and counterfactuals.

Evaluators must understand the complexity and iterative nature of improvement and the contextual factors that accelerate or hinder progress. To evaluate an improvement program as it occurs, evaluators should work with improvers in the program design phase to ensure that there is a shared understanding of the theory of change and improvement program design. This dialog should continue throughout the program to capture any modifications in the theory of change or program design that occur. Given that improvement takes place in complex adaptive systems, unless the improvement program design is fixed, as in Example 1 of Fig. 1, the improvement program design is likely to change. In external evaluation designs, communication between the improvement team and evaluators does not likely occur, which results in missing communication of any modifications made to the theory of change and program design.

Involving evaluators in the early stages of improvement design as well as throughout the implementation process allows evaluators to

better understand the contextual interactions that yield individual changes in the implementation. Increased interaction between evaluators and improvers requires a transition from Example 1 on the Evaluation Continuum towards a more embedded relationship that includes increased coordination, communication and feedback as described in Examples 2–4. Engagement of evaluators throughout improvement programs not only allows for more informed evaluations but allows evaluators greater opportunity to provide feedback to improvers throughout the implementation process. Evaluator feedback is useful in informing improvers as to how and why their results are being achieved. From evaluator feedback, improvers can learn from their improvement, making necessary changes in program implementation and to their theory of change, if needed, in the effort to improve their results.

Conclusion

We propose that the design and evaluation of improvement programs should be mutually informing through a collegial and collaborative ‘marriage’ between improvers and evaluators. We favor more highly embedded evaluation to facilitate shared learning between implementers and evaluators so that the theory of change and associated activities can be amended in close to real-time, thus accelerating adaptation and improvement. In this article, we addressed two questions:

- How do we design improvement programs to make evaluation better? and
- How do we evaluate better to capture the learning and value of programs?

The Framework for Learning about Improvement and the Evaluation Continuum make the case for improvers to fully describe their theory of change and the contextual interactions associated with the changes they are making. The improvement and evaluation options we describe should promote more rigorous program design, faster, more flexible implementation, stronger evaluation, and more credible and generalizable results. Working together, improvers, implementers and evaluators will be in a better position to describe how improvement occurred (or why it did not) and whether the changes are context-specific or are generalizable and can be adapted for spread, and scale-up.

Acknowledgements

The authors would like to acknowledge Nancy Zions for her contributions to the Salzburg Global Seminar. We would also like to acknowledge Lisa Maniscalco, Rhea Bright, James Heiby and Kate Lairmore of USAID for their review and suggestions for this article.

References

1. Chowfla A. *Salzburg Global Seminar Session Report 565—Better Health Care: How Do We Learn About Improvement?* 2016. <http://www.salzburgglobal.org/topics/article/report-now-online-better-health-care-how-do-we-learn-about-improvement.html>.
2. Salzburg Global Seminar Session 565—Better Health Care: How Do We Learn About Improvement? <http://www.salzburgglobal.org/calendar/2010-2019/2016/session-565.html>.
3. Massoud MR, Barry D, Murphy A *et al.* How do we learn about improving health care: a call for a new epistemological paradigm. *Int J Qual Healthcare* 2016;28:420–4. doi:10.1093/intqhc/mzw039.

4. Patton MQ. Principles-focused. In: *Evaluation: The Guide*. New York: The Guilford Press, 2018.
5. Coly A, Parry G. *Evaluating Complex Health Interventions: A Guide to Rigorous Research Designs*. Washington, DC: Academy Health, 2017.
6. Lamont T, Barber N, de Pury J *et al.* New approaches to evaluating complex health and care systems. *Br Med J* 2016; 352. doi:<https://doi.org/10.1136/bmj.i154>.
7. Marshall M, Pagel C, French C *et al.* Moving improvement research closer to practice: the Researcher-in-Residence model. *BMJ Qual Saf* 2014; 23:801–5. doi:10.1136/bmjqs-2013-002779.
8. Parry GJ, Carson-Stevens A, Luff DF *et al.* Recommendations for evaluation of health care improvement initiatives. *Acad Pediatr* 2013;13:S23–30.
9. Portela MC, Pronovost PJ, Woodcock T *et al.* How to study improvement interventions: a brief overview of possible study types. *Postgrad Med J* 2015;91:343–54.
10. Stevens 2016. SQUIRE 2.0 (Standards for Quality Improvement Reporting Excellence): revised publication guidelines from a detailed consensus process. *BMJ Qual Saf* 2016;25:986–992. doi:10.1136/bmjqs-2015-004411.
11. Davidoff F, Dixon-Woods M, Leviton L *et al.* Demystifying theory and its use in improvement. *BMJ Qual Saf* 2015;24:228–38.
12. Ovretveit J. Understanding the conditions for improvement: research to discover which context influences affect improvement success. *BMJ Qual Saf* 2011;20:i18–23.