

## Perspectives on Quality

## Unpacking the black box of improvement

ROHIT RAMASWAMY<sup>1</sup>, JULIE REED<sup>2</sup>, NIGEL LIVESLEY<sup>3</sup>,  
VICTOR BOGUSLAVSKY<sup>4</sup>, EZEQUIEL GARCIA-ELORRIO<sup>5</sup>, SYLVIA SAX<sup>6</sup>,  
DIARRA HOULEYMATA<sup>7</sup>, LEIGHANN KIMBLE<sup>4</sup>, and GARETH PARRY<sup>8</sup>

<sup>1</sup>Gillings School of Global Public Health, University of North Carolina, Chapel Hill, USA, <sup>2</sup>Collaborations for Leadership in Applied Health Research and Care Northwest London (CLARHC NWL), <sup>3</sup>Institute for Healthcare Improvement, USA, <sup>4</sup>USAID Applying Science to Strengthen and Improve Systems (ASSIST) Project, Quality & Performance Institute, University Research Co., LLC, Chevy Chase, MD, USA, <sup>5</sup>Institute for Clinical Effectiveness and Health Policy, Buenos Aires, Argentina, <sup>6</sup>University of Heidelberg, Heidelberg, Germany, <sup>7</sup>Chief of Party USAID Applying Science to Strengthen and Improve Systems (ASSIST) Project, Mali, and <sup>8</sup>Institute of Healthcare Improvement, Cambridge, MA, USA

Address reprint requests to: Rohit Ramaswamy, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, USA. E-mail: ramaswam@email.unc.edu

Editorial Decision 29 November 2017; Accepted 25 January 2018

### Abstract

During the Salzburg Global Seminar Session 565—‘Better Health Care: How do we learn about improvement?’, participants discussed the need to unpack the ‘black box’ of improvement. The ‘black box’ refers to the fact that when quality improvement interventions are described or evaluated, there is a tendency to assume a simple, linear path between the intervention and the outcomes it yields. It is also assumed that it is enough to evaluate the results without understanding the process of by which the improvement took place. However, quality improvement interventions are complex, nonlinear and evolve in response to local settings. To accurately assess the effectiveness of quality improvement and disseminate the learning, there must be a greater understanding of the complexity of quality improvement work. To remain consistent with the language used in Salzburg, we refer to this as ‘unpacking the black box’ of improvement. To illustrate the complexity of improvement, this article introduces four quality improvement case studies. In unpacking the black box, we present and demonstrate how Cynefin framework from complexity theory can be used to categorize and evaluate quality improvement interventions. Many quality improvement projects are implemented in complex contexts, necessitating an approach defined as ‘probe-sense-respond’. In this approach, teams experiment, learn and adapt their changes to their local setting. Quality improvement professionals intuitively use the probe-sense-respond approach in their work but document and evaluate their projects using language for ‘simple’ or ‘complicated’ contexts, rather than the ‘complex’ contexts in which they work. As a result, evaluations tend to ask ‘How can we attribute outcomes to the intervention?’, rather than ‘What were the adaptations that took place?’. By unpacking the black box of improvement, improvers can more accurately document and describe their interventions, allowing evaluators to ask the right questions and more adequately evaluate quality improvement interventions.

**Key words:** improvement, evaluation complex systems, Cynefin framework

## Background

Scientific approaches to improve processes of care in other industries have gained popularity in healthcare. These approaches, labeled as ‘the science of improvement’, ‘improvement science’ or ‘quality improvement’, are being used by practitioners to improve the quality of care in facility or community settings. Quality improvement approaches provide guidance for how to apply systematic thinking and problem-solving techniques to improve a wide range of issues in healthcare. Researchers seeking to evaluate whether these approaches lead to improved outcomes have reported mixed results based on the research method used [1–4]. Researchers have tended to conduct literature reviews of published studies using quality improvement approaches to assess whether the studies report the level of rigor considered acceptable to draw a causal inference between the activity and the outcome. These studies include randomized control trials, quasi-experiments using comparison groups or other methods enabling the establishment of a counterfactual. Because quality improvement projects are conducted in complex contexts, commonly accepted rigorous study designs are often not used. For example, in a review by Moraros, 21 out of 22 selected papers used pre–post designs without a comparison or control group [3]. Because commonly accepted study designs are often not used in quality improvement, researchers have been cautious about attributing quality improvement approaches to improved outcomes. However, these study designs may not be adequate in understanding and learning from the complexity of quality improvement. Treating quality improvement like a ‘black box’ will not serve the needs of researchers or practitioners as it does not adequately describe the process of the quality improvement intervention. In the Salzburg Global Seminar, this was described as ‘unpacking the black box’, a term we will use in this study.

## What is the ‘Black Box’ of Improvement and Why Unpack it?

There is a realization among researchers, evaluators and improvers that quality improvement interventions cannot be understood outside of the context in which they occur. The success of a quality improvement project depends on how the approach was tailored to solve the problem in the context. Context is defined by Øvretveit simply as all factors that are not part of the intervention [5]. The complex and inseparable relationship between an improvement intervention and context means that we cannot treat quality improvement as a ‘black box’ whose contents are mysterious. Quality improvement interventions are difficult to evaluate and even harder to generalize without unpacking the ‘black box’ of each intervention. Due to the relationship between improvement interventions and context, it is difficult to use traditional research designs to evaluate the effectiveness of quality improvement.

The black box can be unpacked through descriptions of the improvement intervention. Detailed information helps evaluators identify what elements of a quality improvement intervention are generalizable, versus those specific to the context. Unfortunately, while this level of detail in improvement documentation is necessary, it is uncommon [1].

A few studies have attempted to unpack the ‘black box’. For example, Dixon-Woods *et al.* [6] developed an ex-post theory to explain why a quality improvement project to reduce central venous catheter bloodstream infection was successfully implemented in the context of a cohort of hospitals in Michigan. In an unsuccessful

quality improvement project, Aveling compares the barriers to the implementation of WHO’s surgical safety checklist in hospitals in the UK and in an unnamed country [7]. In each example, to demonstrate how improvement worked, or did not work within the context, the ‘black box’ was unpacked to describe the relationship between the interventions and context. However, these studies are exceptions. There is no standard approach to describing what is inside the ‘black box’ of improvement.

In this article, we will describe the Cynefin framework [8] and demonstrate how it can be applied to unpack the ‘black box’ of improvement. We will reflect on the applicability of the framework to four case studies presented during the Salzburg Global Seminar on ‘Better Health Care: How Do We Learn About Improvement’ [9].

## Exploring Complexity: Four Case Studies

During the Salzburg Global Seminar, a session titled ‘Unpacking the Black Box of Improvement’ was organized. In the session, a group of participants presented case studies on improvement projects they had designed, led or evaluated. Below is a description of each these case studies.

### Improving Household Water Quality: Ghana

This project aimed to improve the quality of stored household water in 400 communities in Ghana. Research showed that containers with small diameters were less likely to contain contaminated water because it was impossible to dip hands into the container. To implement narrow-mouthed containers in communities that were traditionally used to wide mouth containers, a quality improvement approach was used with iterative Plan-Do-Study-Act (PDSA) Cycles to test local adaptations of narrow-mouthed containers that would be culturally and operationally acceptable, as well as local manufacturing options to make these containers available, and cleaning protocols for decontamination.

The intervention’s effectiveness was tested using a stepped-wedge experiment and the water quality in narrow-mouthed containers was compared to that in control communities using traditional containers. By end of the trial, some members of the control communities acquired narrow-mouthed containers, and some members of the intervention communities had lost or given away theirs. For this reason, an intention-to-treat evaluation of outcomes revealed no significant difference between the intervention and control groups in the percent of households with contaminated water. However, an evaluation of households that had narrow-mouthed containers, irrespective of the group, showed significantly lower rates of contamination. In this case, the problem was not with the intervention, but with the complexity of its implementation.

### Introducing a checklist to improve hospital discharge: UK

A project conducted in a rural hospital in the UK aimed to improve the quality, timeliness and consistency of hospital discharge for adult patients by introducing a discharge checklist that was effective in reducing delayed discharges and length of stay in other hospitals. Over 20 members of staff were involved in reviewing and adapting the checklist to make it consistent with local policies, practices and language. The checklist was iteratively developed using PDSA tests with small numbers of patients. Based on the results of these tests, the checklist was assessed and revised.

Following introduction into practice, challenges emerged. The quality improvement team attempted to resolve challenges with subsequent PDSA cycles. The discharge checklist was intended for use within multi-disciplinary discussions to support patient review and care planning which revealed problems with ward rounds processes involving doctors and nurses doing rounds in isolation and poor communication between staff about patient care plans. Attempts to improve ward rounds processes identified dysfunctional relationships between professional groups that also needed to be addressed.

Changes in ward rounds structures influenced other parts of the system, including the availability of doctors for outpatient clinics, requiring care negotiation. Other issues to be resolved included streamlining documentation requirements for nursing staff and improving radiology and pathology ordering and prioritization processes to ensure timely feedback for discharge decision making. After 18 months, the project achieved a reduction in length of stay and the percentage of patients experiencing delayed discharge.

**Identifying and managing high-risk women during ANC: India**

Zonal Hospital, Mandi, is a tertiary care, 300-bed hospital in a northern state in India. The antenatal clinic was a single room off the hospital’s outpatient department open one half-day a week and staffed by five nurses, who saw an average of 50 patients a day. The nurses were not happy with the care they were providing. Only 1.4% of women who came to the clinic were identified as having a high-risk condition (literature suggests that 10–15% of pregnant women will have a high-risk condition requiring additional investigations or care). On average, only 69 and 42% of women had their blood pressure and hemoglobin measured respectively and this varied considerably week by week. All nurses were to carry out all antenatal care (ANC) activities. There was no queue system, which meant that women had to wait a long time for care and complained about waiting. In response, nurses would sometimes skip different aspects of care.

To address the wait times and ensure that all aspects of care were carried out, the improvement team decided they would assign tasks to specific staff and set up a queuing system. In the process, the team realized that some nurses were unfamiliar with some of the ANC tasks and needed to be retrained. Post-training trials of the queuing system revealed variations in the level of utilization of various stations which required multiple reallocations of tasks across workstations. After these reallocations, the process stabilized.

**Anemia control and prevention in under-5 children: Mali**

This project aimed to improve the delivery of evidence-based interventions to reduce the prevalence of anemia in pregnant women in a district in Mali. The project involved the testing and implementing interventions which included educating providers and staff; creating job-aids; identifying a dedicated medical officer for triage; educating patients on danger signs for anemia; and public service messages and advocacy at the community level.

The interventions listed were implemented and carefully tested in three stages at the facility and community level. After each set of changes was introduced, data were regularly collected and analyzed based on agreed upon indicators. Interacting components of the package significantly improved the percentage of women tested for anemia from 15 to 88%. Educational interventions aimed at service providers on the importance of testing pregnant women’s hemoglobin levels resulted in a 6-fold improvement in this practice from May to October 2013. The progress of the intervention was temporarily impeded by the unavailability of hemoglobin test strips from October to December 2013.

**Describing Complex Contexts: The Cynefin Framework**

David Snowden, a researcher in complexity theory, developed the Cynefin framework (Cynefin ‘is a Welsh word that signifies the multiple factors in our environment and our experience that influence us in ways we can never understand’ (Snowden and Moore 2007)) to characterize contexts. The framework, shown in Fig. 1, has been used to help managers assess their organizational environment and choose appropriate decision-making strategies based on this assessment. It has also been adapted by Michael Quinn Patton, an evaluation expert, for use as an evaluation tool for developmental evaluation [10].

The framework has been used to evaluate a global gender training program for a large international development agency [11], and for evaluating mine and utility company safety [12]. The Cynefin framework is typically traversed counterclockwise starting from the simple context in the lower right quadrant. Simple contexts have repeating patterns and consistent events with clearly defined cause and effect relationships. The decision-making model in simple contexts is described as sense-categorize-respond. If an evaluation senses that the desired outcome is not achieved, the causes are confined to one of few categories and the response can be based on a set of a few simple rules.

The next quadrant in the Cynefin framework, as displayed in the top right of Fig. 1, describes a complicated context. In complicated contexts, cause and effect relationships exist but are not obvious. The decision-making model in a complicated context is sense-analyze-response. Multiple pathways may exist linking cause and effect, and if an evaluation senses that the desired effect size is not achieved, some additional analysis by an expert can lead to the appropriate response.

The top left corner of the Cynefin framework is the most relevant for quality improvement. Complex contexts are those in which linkages between cause and effect may dynamically emerge through learning and adaption, and these may vary by setting. The decision model is probe-sense-respond. In complex contexts, a fixed or static evaluation will not allow us to sense the system—experimentation is needed to probe the system and learn from it before an appropriate

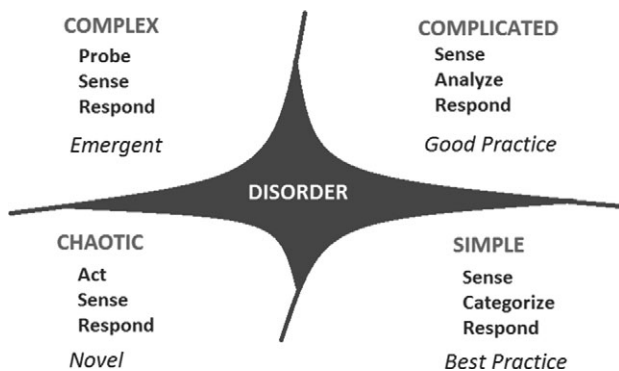


Figure 1 The Cynefin framework for decision making.

response can be generated. When properly conducted, the PDSA cycle is an example of this model. In PDSA cycles, the Plan and the Do steps probe the system in a structured way, the Study step senses the results of the probe and the Act step formulates the response.

The bottom left quadrant of the Cynefin framework is a chaotic system. In a chaotic system, the link between cause and effect is unknown and underlying patterns cannot be easily determined through experimentation. An example of a chaotic system could be a response to a disaster such as the recent Ebola outbreak in West Africa [13]. In chaotic systems, the model is act-sense-respond. Action needs to be taken first, and learning sensed from these actions can be used to formulate a response that hopefully moves the system back towards structure and stability.

## Unpacking the Black Box: Discussion and Themes

A structured discussion followed the case study presentations at the Salzburg Global Seminar, revealing several considerations when unpacking the ‘black box’ of improvement. As demonstrated by the case studies, achieving the goal in an improvement project requires the project to adapt and respond to contextual influences and resolve issues as they emerge. Contextual influences and issues are not always fully anticipated or understood at the outset of the project. Recognizing the ‘black box’ as the complex relationship between improvement success and context, the following themes should be described and documented.

### a) Community and organizational culture matter

The role of culture is apparent in all four case studies. In the water quality example, communities in other sites, with their own water storage habits, cultural norms and modes of social engagement could have engaged in behaviors different from what was observed, in which case, the control group may not have obtained the narrow-mouthed containers. What can be learned from this example is that it is important to learn from and document how context influences results. In this case, context influenced the spread of narrow-mouthed containers to the control group. In the Mali case, community mobilization strategies to encourage referral of women to the ANC during the first trimester of pregnancy were tailored to the local socio-cultural context by incorporating both community and facility involvement in the introduced interventions. In the UK example, changes to work practices involved ensuring accurate documentation of patient care plans. Ensuring accurate documentation ran into larger systems constraints, which needed to be changed to yield improved results. What began as a simple intervention revealed deeper contextual issues of workforce policies and resource utilization.

### b) Changes in one part of the system lead to hard-to-predict changes in other parts of the system

In the UK example, what was initially seen as a simple intervention (a discharge checklist) required changes to work policies and practices in multiple departments whose overall impact is difficult to assess. In India, changes in the workflow to reduce patient wait times had effects on other processes. A change in the workflow to reduce wait times, therefore, required changes to task allocation and load balancing across multiple process steps. In Mali, while not evaluated, it is plausible that increased testing enhanced demand for testing strips as increased demand for testing strips led to a temporary

stock out of testing strips that was not balanced by corresponding increases in supply.

### c) People in the system learn and adapt

In the Ghana example, community members learned about the potential benefits of narrow-mouthed containers and either acquired some for themselves or gave them away. Understanding how this took place and how the implementers responded are more important for implementation than the outcome evaluation conducted. The role of people and their responses to improvement implementation activities play a large role in how an improvement intervention works. In this case, the role of implementers and their responses to implementation were not documented as part of the improvement evaluation.

## Implications for Describing the Black Box of Improvement

The themes described above reflect the complexity of the ‘black box’ and reinforce the need for documentation describing the relationship between context and the improvement intervention itself. However, there is no standard terminology or taxonomy to do so, making it difficult to describe the black box of improvement. Applying the Cynefin framework can help with determining the appropriate category within which an improvement intervention fits and for determining the appropriate evaluation approach.

As in the case studies, improvers typically adapt their interventions to account for the complex contextual environment by changing strategies, methods and tools, as needed. To create generalizable knowledge about how an improvement intervention works within context, improvers must unpack the black box to translate tacit knowledge into descriptions that capture the context of the improvement. Operating in complex contexts, quality improvement interventions usually follow the probe-sense-respond model of the Cynefin framework, employing multiple PDSA cycles and tests of change. Even when the intervention is documented, there is a tendency to describe complex contexts as though they are simple or complicated. The Cynefin framework can be used to describe the internal workings of a quality improvement intervention, thereby unpacking the ‘black box’.

When using the Cynefin framework, it is important to ask guided questions about the context of the intervention. Questions may include, but are not limited to:

- How did the implementers use probes to learn and adapt their intervention?
- How was this learning and adaptation of the intervention used to formulate a response?
- What elements of context influenced this work?
- What aspects of the learning and adaptation in this improvement intervention are generalizable to other contexts?

Asking such questions enhances our understanding of generalizable elements and context-specific characteristics of improvement interventions.

## Conclusion: Furthering the Salzburg Conversation

Many applications of quality improvement in healthcare involve settings that are complex, adaptive and evolve over time. However, we

often describe quality improvement interventions as though they function in simple or complicated systems. By failing to address quality improvement interventions as operating in complex systems, we have not identified evaluation methods that capture the complexity and context of improvement interventions. It is time to use the energy from the Salzburg Global Seminar to bring researchers, evaluators and improvers together to create approaches that describe and document quality improvement in a way that allows us to learn what happens deep within the 'black box' of improvement.

## References

1. Shojania KG, Grimshaw JM. Evidence-based quality improvement: the state of the science. *Health Aff* 2005;24:138–50.
2. Glasgow JM, Scott-Caziewell JR, Kaboli PJ. Guiding inpatient quality improvement: a systematic review of lean and six sigma. *Jt Comm J Qual Patient Saf* 2010;36:533–40.
3. Moraros J, Lemstra M, Nwankwo C. Lean interventions in healthcare: do they actually work? A systematic literature review. *Int J Qual Health Care* 2016;28:150–65.
4. DelliFraine JL, Langabeer JR, Nembhard IM. Assessing the evidence of six sigma and lean in the health care industry. *Qual Manag Health Care* 2010;19:211–25.
5. Ovreteit J. Understanding the conditions for improvement: research to discover which context influences affect improvement success. *BMJ Qual Saf* 2011;20:i18–23.
6. Dixon-Woods M, Leslie M, Tarrant C *et al.* Explaining Matching Michigan: an ethnographic study of a patient safety program. *Implement Sci* 2013;8:70.
7. Aveling E, McCulloch P, Dixon-Woods M. A qualitative study comparing experiences of the surgical safety checklist in hospitals in high-income and low-income countries. *BMJ Open* 2013;3:e003039.
8. Snowden DJ, Boone ME. A leader's framework for decision making. *Harv Bus Rev* 2007;85:69–76.
9. Salzburg Global Seminar. (2016). *Better Health Care: How do we Learn About Improvement? Salzburg Global Health and Health Care Innovation Series*. [https://www.usaidassist.org/sites/assist/files/salzburgglobal\\_report\\_565\\_online.pdf](https://www.usaidassist.org/sites/assist/files/salzburgglobal_report_565_online.pdf).
10. Patton MQ. *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. New York: Guilford Press, 2011.
11. Britt H (2011). *Using the Cynefin Framework in Evaluation Planning: A Case Example*. Retrieved from [https://heatherbritt.files.wordpress.com/2013/05/cynefin\\_case\\_example\\_2011-2-7-1.pdf](https://heatherbritt.files.wordpress.com/2013/05/cynefin_case_example_2011-2-7-1.pdf) on 11 November 2017.
12. Sardone G, Wong GS (2010). *Making Sense of Safety: A Complexity-Based Approach to Safety Interventions*. Retrieved from <http://cognitive-edge.com/articles/making-sense-of-safety-a-complexity-based-approach-to-safety-interventions/> on 11 November 2017.
13. Mangiarotti S, Peyre M, Huc M. A chaotic model for the epidemic of Ebola virus disease in West Africa (2013–2016). *Chaos* 2016;26:113112.