Continuous Improvement Takes Well-Designed Comprehensive Program Evaluation

Ray·Taylor
AND ASSOCIATES

Rossi Ray-Taylor, PhD - Ray.Taylor and Associates, LLC

Continuous Improvement Takes Well-Designed Comprehensive Program Evaluation by Rossi Ray-Taylor, PhD, Ray.Taylor and Associates, LLC

Over the past three decades leaders in the business and social sectors, including education, have focused on reform, transformation and improvement. Improvement science has demonstrated that continuous improvement is essential to success. Leaders in this effort are Anthony Bryk and Louis Gomez and the work of the Carnegie Foundation for the Advancement of Teaching. Bryk and Gomez ground their work in decades of research in the reform efforts of Chicago schools. They, among others, ask "Why have so many reform efforts yielded so little improvement?"

This article describes the juncture between continuous improvement and well-designed comprehensive program evaluation and their power to produce positive impact. Although written from the perspective of K-12 education, the concepts also apply to other social innovation / social impact initiatives. And while much can be said, this article is intended to be a brief practitioners' overview.

Use Evaluation to Learn to Improve

Funders, and state and local education leaders are looking closely at how to accelerate the rates of success and sustain improvement in school reform. Researchers and policy makers have questioned the return on investment of funds and time, along with other resources, and the opportunity costs that are the result of the poor outcomes experienced by many transformation initiatives. Increasingly, social sectors have turned attention to improvement science to focus on learning to improve. Bryk, Gomez, Dylan Wiliam, and Michael Fullen are among leaders in the field.

Popular articles in business journals describe data and analytics needed for improvement. While data are needed they are not sufficient. Strong, well-designed and well-implemented program evaluation is needed to draw insight and direction for improvement from data.

The purpose of this paper is to offer a practitioner centered guide that can inform practice to move beyond compliance to intentionality and continuous improvement in design, process and results. We apply improvement strategies to program evaluation to produce actionable information and insight.

> "Evaluation is to help projects become even better than they planned to be… First and foremost evaluation should support the project…"
>
> W. K. Kellogg Foundation Evaluation Approach, 1997

This paper links improvement to comprehensive and well-designed program evaluation.

Transformation, reform and improvement of organizations are best grounded in the information provided by credible, rigorous, and comprehensive program evaluation. Independent external evaluation can focus lens on an initiative that may not be clear when viewed internally. The strategies of process and impact evaluation feed the information needs and build the connection to practice needed for continuous improvement. In ***Learning to Improve: How America's Schools Can Get Better at Getting Better*** (2015), Anthony Bryk, Louis Gomez, Alicia Grunow, and Paul LeMahieu offer six principles for improving how schools and districts learn to improve. Their work is based in large part on their experiences chronicling Chicago schools' attempts at reform over the past couple of decades.

The six improvement principles from Bryk and Gomez, et al, and the Carnegie Foundation for the Advancement of Teaching are:

1. *Make the work problem-specific and user-centered*.

2. *Focus on variation in performance*.

3. *See the system that produces the current outcomes*.

4. *We cannot improve at scale what we cannot measure*.

5. *Use disciplined inquiry to drive improvement*.

6. *Accelerate learning through networked communities*.

We apply these six principles to design and conduct evaluation, and outline the connection between principles for improvement and evaluation.

Problem-specific and user-centered – The contextual stages of an evaluation call for clear articulation of the needs, theory of the case, and understanding of the local culture and participants' voice and experiences. *Focus*

> The work and precision put into this phase can shape the evaluation. Consider how needs are identified and described. What needs is the initiative designed to address? What problem is to be solved? What is the focus of needs, who owns the issue, who describes the matter, with what language?

Variations in performance – Process evaluation assists in understanding variations in implementation, resources and outcomes. Variation in performance can also come from the disparate impact of the intervention upon subsets of the participants, or differences in performance over time. *Fidelity and disparate impact*

There is a difference between intent and the actual exercise of an initiative. Even with the best intensions the actual implementation may miss the mark. The conditions, resources, or implementation may not be as planned.

See the system – The systemic view is achieved through the contextual analysis acquired in the process evaluation stage. Through process analysis connections can be made among various features of an initiative and the full system it inhabits. *Systems thinking – design thinking*

The connections that make up a system are essential to understanding the full impact.

Measure – Rigorous measurement of implementation, short term outcomes, and long term impact are hallmarks of a well-designed and well-implemented evaluation. The true challenge is to avoid measures of opportunity – measures that are easily obtained and charted but may have little relationship to the true impact and goals. *Measurable – valid, reliable*

Use disciplined inquiry – Guidance by the inquiry and analysis practices of a well-designed evaluation gives the disciplined inquiry needed to produce reliable, valid and high-impact evaluation results. Evaluation standards have been chronicled by the American Evaluation Association, ESSA and other research based entities. *Purposeful program evaluation*

"Evaluation to Prove / Evaluation to Improve"

Networked communities for improvement – Sharing results throughout the evaluation process with stakeholders and others in the implementation network builds the knowledge base and capacity for improvement. Sharing also offers the opportunity for evaluators to gain insight from practitioners, initiative leaders, and service recipients that can give depth and nuance to the evaluation. *Professional collaboration*

Why evaluate?

Program evaluation can be precipitated by any number of means. Why it is being conducted helps to define the design, methodology and distribution of the results. Here are five common reasons for program evaluation.

- Continuous improvement – in some organizational cultures evaluation is a continuous part of assessing effectiveness of the organization's strategies.

- Corrective feedback needed to achieve excellence

- Accountability – evaluation for accountability can be employed as policy or funding decisions are being made

- Funder mandates – most grant funders require some level of evaluation

- Decision maker mandates – Board of Education members or other decision makers may mandate periodic evaluation of key initiatives, or in response to certain conditions

How to Evaluate – more than just the data (numbers)

Once we have determined *WHY* evaluate we turn next to outline *HOW* to organize and conduct an evaluation. If we are led by the two key purposes for evaluation – *to prove* and *to improve* we note that:

- Improvement can be informed by process, impact, and local culture
- Impact can be comprehensive, measurable and sustainable. However, impact can also be unplanned, ephemeral, off-track and transient.

Timeline – Contrary to what for some is common practice, comprehensive program evaluation should begin at the initiative design phase and continue throughout the initiative and beyond. Evaluation considerations like plans for data collection and schedules are ideally designed into the initiative from the start. Analysis and feedback discussions among the evaluation team, the project management team and in some cases service recipients, over the course of the evaluation can yield useful information to continuously improve.

We begin with process evaluation to paint a full picture of the initiative.

Process Evaluation

Process evaluation informs the project team and funders about how the project is actually implemented. Process Indicators:

- Give information that the project is on track to success.

- Inform evaluators about the factors that impact project outcomes. Including:

    o   Ways the implementation of the initiative differed from plans.

    o   How incidents impacted the implementation or outcomes.

    o   Any non-alignment of purpose, resources, methodology or assumptions that may impact the outcomes.

A clear assessment of the intent of the initiative is critical. The *why*, *when*, *how*, and *for whom* are important to a clear, complete and informative evaluation. A succinct statement of intent – what is hoped to be achieved – the mission and goals of an initiative, is central to process analysis.

Process evaluation asks and answers - How did you do it; Why this way; With what fidelity; To what extent were changes made or needed; To what extent was the initiative adaptable?

A clear statement of the case for the initiative is important to an evaluation. The research or experiential basis for the project can be tested through the evaluation process.

Process evaluation can be informative to fully understand an initiative, including its fidelity and how it did or did not adapt to a changing environment or resources. Process analysis gives meaning to impact, sustainability, and replicability. And as noted earlier, it is central to questions of improvement. In the

case of analysis of methodology, process evaluation insight can help discover if and how assumptions about targets, methodology and resources fit.

Context and local culture – understanding context and local culture can give information about the "air" surrounding an initiative. Local traditions, taboo, infractions, power relationships, historical events, beliefs, and allegiances can be part of how an initiative is experienced by those who are its champions or targets.

Contextual analysis can provide powerful insight about a program and can yield information useful for targeting improvement. Factors like prior experience of the project participants, recent initiatives, and competing demands, can form cultures supportive of, or resistant to the initiative. Local history and expectations can shape an initiative and how it is perceived.

Context Matters – Consider an initiative begun at the direction of a superintendent who strongly supported site based decision making. Following his departure from the district the new superintendent implemented a centralized decision making regime and added initiatives competing with the former superintendent's plans. The result – processes and outcomes were no longer achieved and the initiative was abandoned. Contextual factors may have led to poor outcomes more than the merits of the initiative itself.

Projects exist in an evolving information climate and changing expectations; where what is evaluated at point "A" may be different than at point "B". Benchmarks, niche analysis and environmental scan can be useful methods to capture information about changing context, resources, and expectations.

Indicators of systemic health (e.g., financial, student test outcomes, community support, enrollment stability, leadership stability) can foreshadow initiative success. By way of example, an initiative begun in a climate of pending layoffs due to financial instability may face resistance and be shaped by unforeseen budget cuts.

Impact Evaluation

Beyond best intentions, resources, and skilled implementation one must ask the impact of an initiative. To what extent did it achieve what it was intended to achieve. A common mistake of measuring impact is to focus on intent and implementation. In these cases the impact is described as merely implementation, not necessarily change. This takes us back to the need for clear statements of intent and measurable outcome statements and measures. One misstep made by those new to evaluation is to over-emphasize perceptions – "likes", or to rely too heavily on participant self-report of their change. Intent may not be the same as action. Post-initiative surveys, for instance, may show that while more students say that they intend to choose healthier foods in the cafeteria, actual review of selections may show no significant differences in student behavior.

Impact evaluation tells the project team, funders and other stakeholders about outcomes.

- What were the results of the initiative?

- To what extent did it meet planned goals? For whom?

- Were there unplanned outcomes and impacts? *Unintended consequences*

- Were the goal targets maintained after the project? *Sustainability*

Two considerations of an initiative's impact are magnitude and sustainability. Magnitude addresses the depth and breadth of impact. It questions the evidence-based social and environmental returns on investment – put bluntly, is the initiative worth the cost (e.g., in time, money, disruption, or distraction from other priorities)?

Sustainability explores the extent an intervention will last. Some reviews show that once the glow has ceased the results diminish. Sustainability can be compromised by any number of factors including attrition among the leadership, or champions of the initiatives, or those trained to implement it. Similarly, competition from the next project, the next perceived crisis, or diminished budgets may impact sustainability. Policy makers also play a role in sustainability. They, along with funders, can set a climate for improvement and evidence-based practices.

TMI – In this era of bountiful analytics there really is such a thing as *too much information*. A meaningful evaluation relies on credible, sound, reliable, timely and valid data; data that clarifies the problem to be addressed. All available data may not meet these qualifications.

While perception surveys like those distributed at the end of a workshop or training session may give a glimpse of the participants' views, they may not give enough information to describe an initiative's overall impact. And as seen in popular political campaigns, surveys are dependent on factors like the quality of the design and questions, timing, response rates, analysis, collection methods and platforms, and respondent motivation.

Look beyond intent - Take a look from the perspective of the end user – e.g., classroom teachers and principals, students, parents.

When viewed from these perspectives one may find that participants experience a tidal wave of innovation and new programs that may be under-resourced, not well conceived, or ineffectively implemented.

Three perspectives can be called on to guide work in the design and implementation of a comprehensive evaluation; **cultural competence**, **design thinking,** and **systemic impact**.

> Cultural Competence – Service providers implementing a community program in one city neighborhood found that parents were resistant to the parenting practices that were offered. When a team member dug deeper she learned that parents didn't trust what they viewed as *suburban parenting practices* of the initiative and viewed these practices as misguided, over protective and akin to missionary zeal.

The impact of **cultural competence** cannot be overestimated – beneath the surface value of an initiative and its relationships are the impact of matters like cultural history, language, values, local history, and class issues. These distinctions may or may not match commonly acknowledged groups, and may represent dynamics of groups-within-groups, or cross-group or cross-

Design Thinking – Teachers at one elementary school were trained to implement a new approach to teaching mathematics. Upon a closer look from the teachers' perspective project leaders found that during the same time the teachers were in their first year of implementing new literacy textbooks and were being evaluated using a new state mandated evaluation system. Teachers said that they were overloaded. Once all of these impacts from the teachers' perspective were considered the design of the mathematics initiative was altered considerably.

generational alliances and discord. What micro-aggressions or triggers may be present but not visible to all. For example, what may seem to be positive intent to an evaluator, initiative designer or service provider, may be viewed very differently by the service recipient.

Design thinking – Similarly, design thinking asks how those who are the targets of an initiative actually experience the initiative. For example, how do trainers, trainees, parents and students experience the intervention? Design thinking – asking the recipients of the service for their ideas, perceptions and recommendations may uncover, for instance, that there are strong feelings in the community that the initiative was "imposed from above" or that those in charge are *outsiders*, or that the remedies proposed do not value the local culture and are judgmental.

Systemic impact – Systemic impact casts light on the full impact, beyond the thresholds of the targeted audience. Identifying and addressing the full system impact on intervention, and vice versa, raises questions like – what works with the model, what works against it, what prerequisite conditions exist in the system? Systemic impact recognizes that the initiative may have ripples far beyond the targeted audience, and likewise, may be the recipient of waves of impact far beyond its intent.

On a final note – A well-designed and well-implemented program evaluation should reflect and inform the implementation and impact of an initiative, and provide insight needed for project leaders, participants, funders, and stakeholders to prove its impact and success, and to improve it.

Systemic Impact – In another example from the launch of teacher professional evaluation by principals, we found that teachers described hesitancy to coach, share lesson plans and help colleagues as a result of the new professional evaluation system. The teachers felt increased competition in the evaluation process. This was displayed throughout the system and impacted professional development, staffing decisions, new teacher induction, and other areas not foreseen in the professional evaluation initiation.