



## Resource - Matching Evaluation Approaches to Expectations 2.0

Sometimes, confusion about evaluation arises because there is a mismatch between the approach and purpose or expectation. Consider the following five basic evaluation approaches:

**Facilitated critical reflection** or what we colloquially refer to as the pizza party approach is about acknowledging that sometimes the best way to learn and evaluate a program or service is to get all relevant stakeholders in a room (perhaps over pizza) and have an open and honest dialogue. This approach is less complex than some of the others in terms of data collection requirements for instance, but it nonetheless requires a skilled facilitator to keep the discussion focused. It also requires dedicated and uninterrupted time to work through the issues and a willingness by participants to share and reflect. This approach may be most appropriate in situations where there is a high level of trust among stakeholders and where there is an opportunity to learn from others who are coming at the issue from a different perspective.

**Performance measurement** or program monitoring is the ongoing, day-to-day data gathering that program staff and volunteers do as part of their job. It tends to use low-cost, less intrusive data gathering techniques. It often focuses on tracking program processes and outputs (e.g., attendance rates, demographic information about participants, or basic feedback on satisfaction). Simple satisfaction surveys and attendance sheets are classic examples of performance measurement methods. Performance measurement is good at generating data that managers can use quickly, efficiently, and frequently. It isn't as good for addressing larger, more complex questions such as those that deal with program impact or ways in which the program could be redesigned.

**Program evaluation**, when used in its narrower and more technically correct sense, refers to data gathering work that is more intensive, more formal, and more time-limited than ongoing performance measurement. While program monitoring work typically produces a simple summary of key statistics or a dashboard, a program evaluation project typically begins with a critical analysis of the theoretical assumptions underlying a program (using a theory of change) and produces an analytical report with conclusions and recommendations. Program evaluation work often involves a deeper investigation into the outcomes or impacts as well as questions related to the process. It often uses a greater variety of data gathering methods and attempts to determine whether a program led to change and why. Program evaluation work is good for generating evidence of impact as well as practical, actionable ideas for how programming can be improved or buy-in from various stakeholders can be increased. However, it is more technical and more resource intensive than performance measurement. It can be more intrusive in the lives of participants and it often takes a bit longer to generate insights.

**Systems evaluation.** Governments or other large funders may commission many related evaluation projects and then seek to combine their findings into a report on the impact of a

complex and diverse set of community investments. This process is quite different from any of the three approaches discussed so far. We refer to this approach as systems evaluation. This work involves data collection by many people in different locations. It often requires pulling together various kinds of information, originally gathered for dissimilar reasons. However, performing good systems evaluation is more complicated than simply “rolling up” the findings of many local program evaluation reports. Systems evaluation is designed to answer questions that are different from those in program evaluation or applied research. It often seeks to determine whether services have been implemented consistently across sites. It can also seek to understand how different kinds of interventions have helped or hindered one another in a local community and whether they have worked together to produce collective impact.

**Applied research** is more time consuming, more theory-driven, and more expensive than program evaluation or performance measurement. It often looks at data from multiple programs or program sites and focuses on a small number of focused research questions. It is typically designed and carried out by academic researchers who are content experts. Its primary purpose is to create generalizable new knowledge and, therefore, it may not always generate practical recommendations for immediate local action. Program evaluation is different from applied research in that it considers local context, the values of the people involved, and the program’s side effects.

In the updated table below, we consider six common purposes or motivations for conducting evaluation work and we consider which of our five evaluation approaches is the best fit for each:

- Where there is a good fit between approach and expectations, there is a **green dot**.
- Where there is a bad fit, there is a **red dot**.
- Where it is best to proceed with caution, there is a **yellow dot**.

The blue arrow from left to right indicates the increasing complexity of the approaches.






























#### For further reading

Alaimo, Salvatore P. 2008. “Nonprofits and Evaluation: Managing Expectations from the Leader’s Perspective.” Edited by J.G. Carman & K.A. Fredericks. *New Directions for Evaluation* (Nonprofits and evaluation) 73-92.

Newcomer, K and C. Brass. 2015. “Forging a Strategic and Comprehensive Approach to Evaluation Within Public and Nonprofit Organizations: Integrating Measurement and Analytics Within Evaluation.” *American Journal of Evaluation*, 36 (1), 1-20.

Pawson, Ray, Trisha Greenhalgh, Gill Harvey, and Kieran Walshe. 2005. “Realist review – a new method of systematic review designed for complex policy interventions.” *Journal of health services research & policy* 21-34.

Taylor, Andrew and Ben Liadsky. 2016. “Evaluation Literature Review.” Ontario Nonprofit Network. Accessed May 10, 2016. [http://theonnc.ca/wp-content/uploads/2016/01/Report\\_ONN-Evaluation-LiteratureReview\\_2016-01-21.pdf](http://theonnc.ca/wp-content/uploads/2016/01/Report_ONN-Evaluation-LiteratureReview_2016-01-21.pdf).

Purposes or motivations for doing something called “evaluation”.	Approaches				
	 <b>Facilitated Critical Reflection</b>	 <b>Performance Measurement</b>	 <b>Program Evaluation</b>	 <b>Systems Evaluation</b>	 <b>Applied Research</b>
For an external agent, such as a government department, to determine whether local programs, sites or agencies have	 Not a good fit. It can take time to build trust and create a safe space for critical reflection with grantees.	 Good fit... if data is used!	 Not a good fit. Evaluation methods are more complex than needed and the turnaround time for analysis may be too slow.	 Can work well... when time and energy are invested in shared performance measurement systems.	 Not a good fit. Applied research is not designed for day-to-day monitoring.
For a nonprofit to determine whether individual programs or sites have delivered a service as planned.	 Good fit! Sometimes a facilitated conversation with appropriate stakeholders about what works, what doesn't, and why is all that is needed.	Good fit... if data is used!	Not a good fit. Evaluation methods are more complex than needed and the turnaround time for analysis may be too slow.	 Not a good fit. Shared measurement tools aren't typically flexible or sensitive enough to track nuances of local programming.	Not a good fit. Applied research is not designed for day-to-day monitoring.
For an external agent to determine whether local programs, sites, or agencies have achieved impact as planned.	 Not a good fit. It is unlikely that this would allow enough time for an external agent who has not been involved in the day-to-day to determine the level of impact.	 Can work well... if performance measurement systems are sophisticated, specialized, and carefully monitored.	 Can work well... but evaluations undertaken for this purpose may not be as good at generating local insights or actions (see below).	 Can work well... when time and energy are invested in shared measurement systems, ongoing communication, backbone infrastructure, and a shared sense of purpose.	 Not a good fit. Applied research is not designed to inform action in a direct way.
For nonprofits or networks of nonprofits to develop insights about their work and its impact leading them to improve practice.	 Can work well... when combined with other approaches to distill lessons learned.	 Rarely works well. Measures are focused on outputs, buy-in is minimal, and analysis is basic.	 Good fit! Especially when time and energy is invested in buy-in, communication, clarity of purpose, and plans for use.	 Can work well... if local sites are engaged as partners in the process.	Not a good fit. Applied research is not designed to inform action in a direct way.
For networks representing a sector or the community as a whole to develop new knowledge about best practices and long-term impacts.	 Not a good fit. This approach is unlikely to lead to the kinds of insight necessary to develop new knowledge.	 Rarely works on its own (although performance measurement methods are often useful when incorporated into more complex systems evaluation projects).	 Rarely works. Local evaluations do not typically measure long term change and are not designed to generate generalizable knowledge.	 Good fit! Especially when time and energy are invested in shared measurement systems, ongoing communication, backbone infrastructure, and a shared sense of purpose.	 Good fit! Especially when the research questions are highly focused and specialized.
For large systems to develop insights about their work and its impact leading them to improve practice.	Not a good fit. This approach is unlikely to lead to the kinds of insight necessary to develop new knowledge.	Rarely works on its own (although performance measurement methods are often useful when incorporated into more complex systems evaluation projects).	 Rarely works. It is challenging to aggregate findings from different local evaluations.	 Can work well... if research findings are presented in an accessible way and provided in a timely manner.	 Can work well... if research findings are presented in an accessible way and provided in a timely manner.