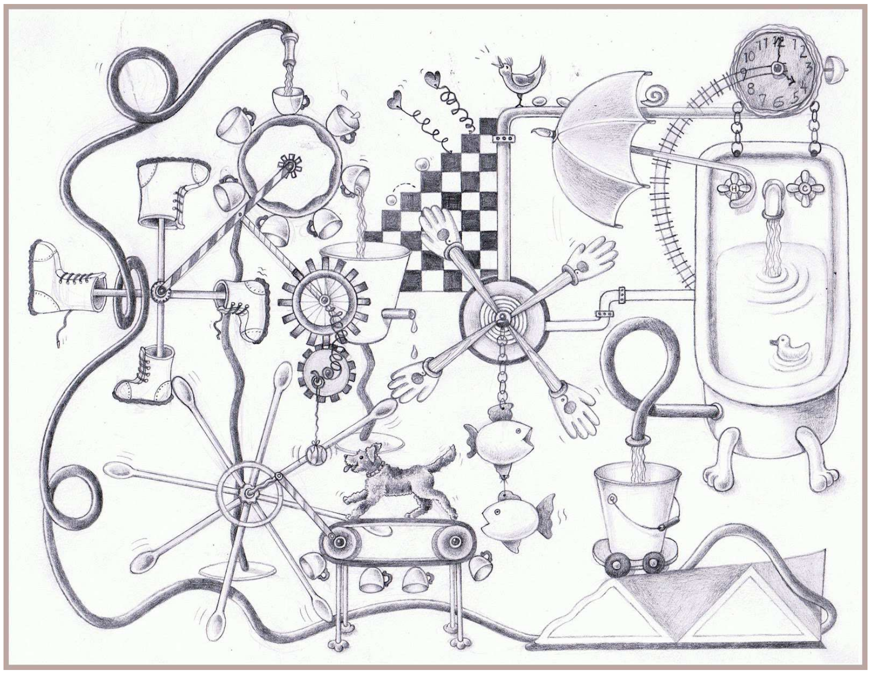


On Evaluating the Secondary Impact of Leadership Programs: A White Paper

Justus Baird and Aaron Dorfman

Processes and methods to evaluate the impact of leadership programs run by nonprofit agencies, particularly of programs designed to leverage leaders for impact on their communities and society.

Produced with the support of a Wexner Graduate Fellowship Alumni Collaboration Grant

Summary

This report outlines an approach and some specific methods for evaluating a particular, and often neglected, aspect of nonprofit leadership training and development programs. While the primary (or direct) impacts of these programs on the participating leaders are often evaluated and understood, it's less clear how leadership programs ought to think about and evaluate their secondary impact on the communities in which their program alumni work. This report offers a conceptual framework and tools for staff of leadership training and development programs to use in evaluating that secondary impact.

INTRODUCTION

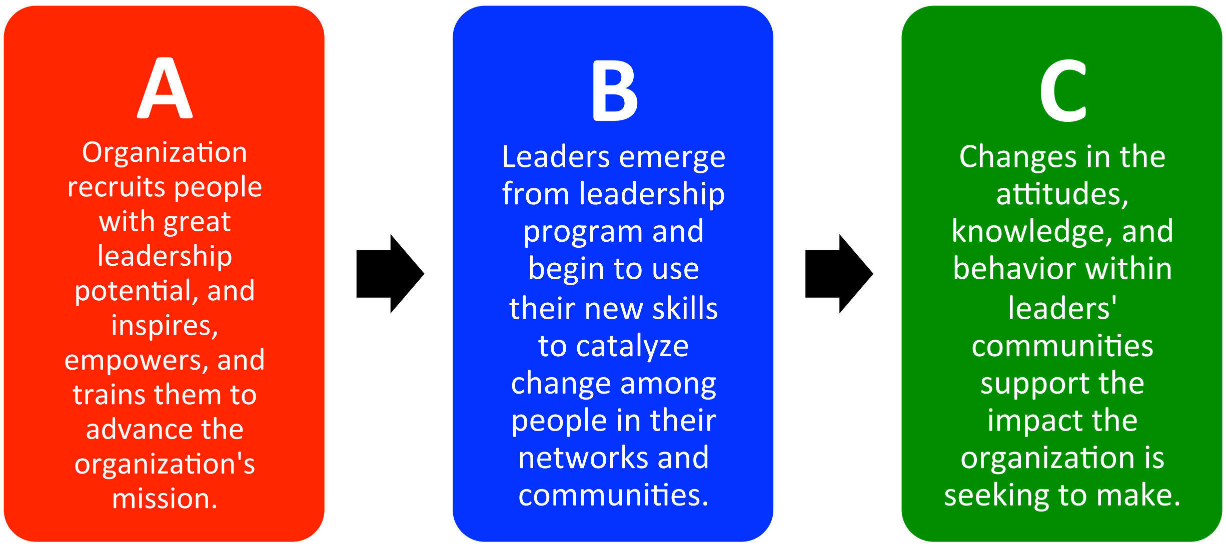
There's an apocryphal story in which a Harvard Business School student visits the course materials office at HBS and is disgruntled by the lackadaisical service. Fed up with waiting, he finally raps his knuckles on the counter and barks, "Hey! Can you hurry it up? I'm the customer!" In response, the clerk casually lowers the crossword puzzle in which she's been engrossed and remarks, "You're not the customer. You're the product." This joke encapsulates a perception that is often held among those who manage programs focused on leadership training and development: that the participants are the "product" of the learning process.

Many nonprofits run programs to invest in future skilled leaders and professionals. While we are deeply committed to the intrinsic value that these programs offer to those who participate in them, we should be equally committed to understanding the impact that the leaders in whom we've invested have on their communities and networks, what are referred to in this report as "secondary impacts." In fact, the real measure of our success is not the leader's experience alone, but also the change they catalyze in their communities, their networks, and the larger world. In this context, our participants are also a means to support our larger visions for change. Our implied theory of change is that investing in leaders translates to better-equipped leaders who can more effectively pursue the change we all need to see in the world. The measure of our success is how much more effectively they realize that change. This report was born out of a shared need of its authors, staff leaders at two different nonprofits, American Jewish World Service and Auburn Seminary. We were both overseeing major programs

Case study from the Talmud: Training leaders to achieve secondary impacts
(This story about leadership in a moment of crisis comes from the Babylonian Talmud, Bava Metzia 85b.)
One day Rabbi Hanina and Rabbi Hiyya were arguing, and Rabbi Hanina said to Rabbi Hiyya: 'How can you argue with me? If, Heaven forbid, the Torah were forgotten in Israel, I would [be able to] restore it by my argumentative powers.' To which Rabbi Hiyya rejoined: 'How can you argue with me, who actually prevented the Torah from being forgotten in Israel? How did I do it? I went and sowed flax, made nets [from the flax], trapped deer and gave the meat to orphans, and prepared scrolls [from their skins], upon which I wrote the Five Books [of Torah]. Then I went to a town [where there were no teachers] and taught the Five Books to five children, and the six orders [of the Mishnah] to six children. And I told them: "Until I return, teach each other the Five Books and the Mishnah;" and thus I preserved the Torah from being forgotten in Israel.' This is what Rabbi [Yehuda haNasi meant when he] said, 'How great are the works of Hiyya!'

that aspire to strengthen the leadership of a cohort of leaders. For both of us, the investment in leaders was clearly a means to an end: the real desired outcome was making an impact on society, and the leaders we were training served as vehicles for achieving that impact. We were operating under a hypothesis about social change that presumes that great leaders make great things happen, so investing in people to help them become great leaders—inspiring, empowering, and training them to catalyze communal transformation—would lead to great change.

Our bare-bones theory of change looked something like this:



While we had many tools for evaluating our program's impact on the leaders themselves (A->B or "primary impact"), when it came to understanding or measuring how each leader's experience in our program contributed to an impact on the communities they led (A->C or "secondary impact"), we were at a loss.

Stirred by this frustration, we produced a two-day seminar for nonprofit staff to study approaches to evaluating the secondary impacts of leadership programs. We are grateful to the Wexner Foundation for underwriting the costs associated with the seminar (through an alumni collaboration grant), held in April 2015 at the New York offices of American Jewish World Service.

Twenty-two colleagues who lead similar leadership programs (mostly in the Jewish nonprofit sector) participated in the seminar. We hired two experts in evaluation to lead us in our learning and reflection: Dr. Jewlya Lynn from the Spark Policy Institute in Denver and Tanya Beer from the Center for Evaluation Innovation in Washington, D.C. The report shares the tools and insights we gained through the seminar, in the hopes that other nonprofit staff and funders designing and supporting leadership efforts can better assess the impact their leaders have on the communities they serve, in order to learn and improve those programs to effect yet more positive change for our world.

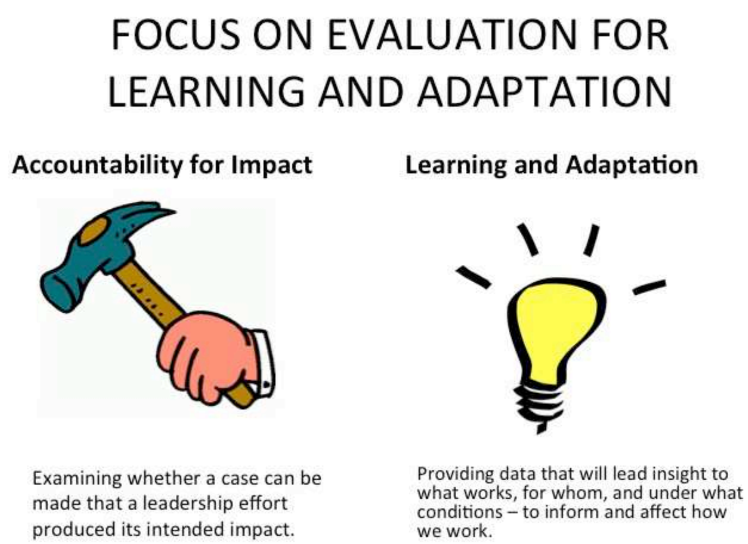
EVALUATING IMPACT IN A COMPLEX SYSTEM

The leadership programs at issue address what are often called complex problems or systems. As opposed to simple problems (baking a cake) and complicated problems (sending a rocket into space), complex problems or systems—such as raising a child, changing immigration policy, or solving climate change—are characterized by a multiplicity of interconnected factors, some of which are unknown or invisible; an inability to predict consequences; and unstable contexts.

When evaluating impact in the context of a complex system, experts encourage evaluators to develop a plausible case for how a program contributed to the outcomes, as opposed to overstating the case. The Spark Policy Institute encourages evaluators to:

- Aim for contribution rather than attribution (i.e. don't assume that any particular outcome is entirely attributable to a single intervention);
- Look for emergent outcomes, especially those you didn't anticipate or expect;
- Keep outcomes realistic;
- Differentiate between outcomes that you "expect to see," "would like to see," and "would love to see"; and
- Use a "golden spike"[1] to link to other evidence: connect your impact to already proven causal relationships, focusing your evaluation on the gaps or new learning.

Perhaps the most important approach to evaluating complex systems is to focus more on "learning and adaptation" than on traditional "accountability for impact."



Evaluations that focus on learning and adaptation encourage developmental learning and risk taking. The complex problems our programs are addressing are never entirely solved; what worked last time may not work in the future; and what works in one context may not work in another. Developmental evaluation methods track effects as they unfold, not expecting methods or even intended outcomes to remain stable. In this context, evaluation encourages staff to adapt to what is learned, acknowledge what isn't working, and use the new knowledge to inform future decisions. In their book Getting to Maybe: How the World is Changed, Westley, Zimmerman, and Patton call this a "safe-fail" approach, as opposed to a "fail-safe" approach[2]. The point is that in the context of complex systems, spending lots of time and resources on measuring the original, predicted impact may not be an effective way to get to the ultimate outcome.

THREE-STEP CYCLE FOR BETTER EVALUATION

Moving from the *idea* of more effectively evaluating leadership programs in a complex system to actually *doing it* can be daunting. We have found in our own work that it takes significant time and energy to get institutional buy-in and involve colleagues in new behaviors. Whether you try a DIY ("do it yourself") approach or involve a consultant, these three steps form the core elements of an evaluation and learning cycle.

The first step is to **reflect on what you are trying to achieve (outcomes) and how you are trying to achieve it (change pathways)**. It is impossible to effectively measure or learn about impact until you have first articulated as clearly as possible what the desired outcomes are and how you think they will come about. The second step is to **identify and utilize methods to collect and analyze data that can measure the desired outcomes.** Evaluation experts have developed a wide range of creative methods and tools for this step, well beyond the surveys, interviews, case studies, and focus groups that we are most familiar with. And even for some of these tried-and-true methods, experts propose refinements that make them more effective in the context of complex systems. The third step is to **feed the learning back into the ongoing program design and implementation.**

Step 1: Articulate Meaningful Outcomes and Change Pathways

Most programs treat evaluation as an exercise in counting heads (e.g., How many people applied to / participated in / graduated from my program?) and/or customer satisfaction (e.g., How would you rate your experience in my program on a scale of 1-5?). We design leadership programs to leverage leaders to effect meaningful change in the world. The first step in evaluating these programs is to articulate the change we seek to produce and the way we think it will materialize: What do we want to see happen in the communities our leaders serve? Will more people join those communities? Will community members engage in more—or better—political activism? Will they engage more productively in civil discourse around contentious issues? Will they manage their institutions more sustainably?

One way to begin articulating the outcomes and change pathways of a leadership program is to develop a **persona**, an iconic/archetypal story of a program participant who exemplifies what the program is trying to accomplish. Using personas can help concretize program design in the observed interests, experiences, and perspectives of real people. Developing a persona consists of the following key steps:

1. Identify target users
2. Develop a research process (interview and observation protocol)
3. Learn from target users (through interviews, focus groups, observation, etc.)
4. Analyze and synthesize learning into one or more personas
5. Produce a written profile for each persona

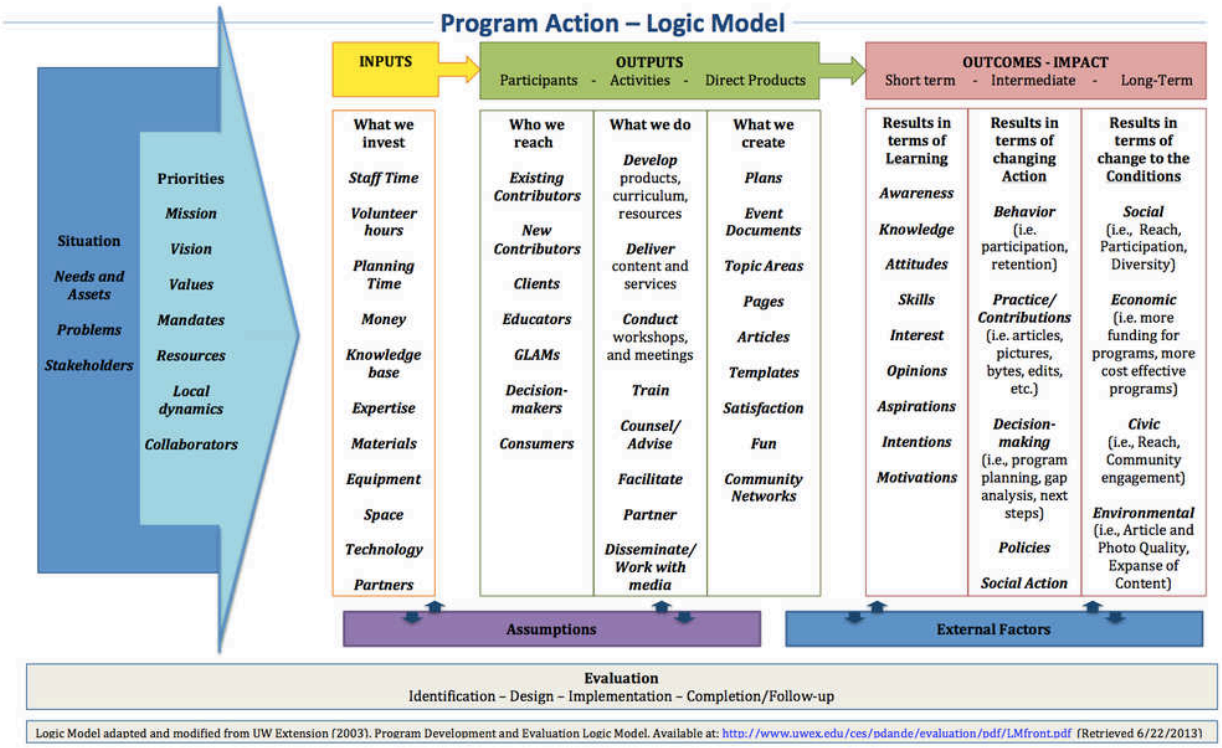
Shlomo Goltz has written a helpful guide for developing robust program personas. Below, you'll also find a case study from one of the program directors who participated in our evaluation seminar. While this is a description of an actual program participant and not a synthesized persona, it conveys the value of having an individual story to capture the kind of change your program is designed to produce.

Another valuable tool for articulating outcomes and change pathways is the **logic model**. (See W.K. Kellogg Foundation's logic model development guide.) A logic model is a flow-chart diagram depicting the causal relationships among the various elements of program design. While there are many logic-model formats, most include and describe relationships among the following elements.

- Planned work:
    - Inputs: resources invested in the program (e.g., funding, staff time, materials, political capital, etc.)
    - Activities: activities undertaken using inputs (e.g., training programs, political advocacy, direct service, etc.)
- Intended results:
    - Outputs: direct and usually quantifiable products resulting from activities (e.g., # of people served, # of lobbying visits or petition signatures, customer satisfaction targets, etc.)
    - Outcomes: specific changes in program participants' attitudes, behaviors, knowledge, skills, etc., that result from the program (e.g., reduction in child malnutrition, passage of specific legislation, increase in # of skilled volunteers, etc.)
    - Impact: systemic transformation supported by the program's outcomes (e.g., healthier families and children, reduction in crime rates/violence, increase in civic engagement, etc.)

Some logic models also include articulations of the problem the program is designed to solve, underlying assumptions that inform the logic model design, and critical factors in the external operating environment.

Logic models can be used for many purposes in both program design and evaluation, but their core utility is in making explicit and transparent the causal relationships among the various elements of a program.



[3]

An Example in Practice: Bronfman Youth Fellowships (BYFI)

Julie (not her real name) grew up in a part of the U.S. with a small Jewish community. She belonged to a Reform Jewish synagogue, attended Hebrew School, and participated in Reform Jewish summer programs growing up. Her limited exposure to Jewish diversity resulted in very few friendships with Jews outside of the Reform movement.

Julie traveled to Israel with BYFI the summer before her senior year of high school and continued with BYFI seminars during her senior year. The program exposed her to Jewish pluralism, text study, and Israel in a multifaceted way. She built close relationships with Jews different from her (e.g., she emerged from the program best friends with an observant Orthodox young woman).

Julie went on to attend an Ivy League school. Because of her experience in BYFI, she decided to get involved and joined a prayer community, but came to see that the Hillel—dominated by Orthodox students on her campus—seemed to alienate Jews from more secular backgrounds and those with left-leaning perspectives on Israel.

As a junior, Julie decided to run for Hillel president against a politically conservative and religiously Orthodox fellow student. She was the underdog, but spent a lot of time reaching out to students on the fringes and those in the center. Her BYFI experience gave her valuable tools for how to speak with Jews from diverse backgrounds and helped her honor each person's point of view as equally valid. She gained people's trust and won the election.

A month prior to the election she was feeling disheartened about polarized attitudes toward Israel on campus. She attended a BYFI seminar for college alumni designed to create a safe space to explore the Israel discourse on campus, provide discussion tools, and build community for students who care about Israel across different political perspectives. She broke down in tears at the seminar, but had several mentoring conversations with staff and peers that bolstered her. After the seminar she wrote to staff saying the retreat had rejuvenated her and given her increased confidence and strength to keep her campaign going. The sense of being part of a community that believed in her and the reminder that there are Jewish spaces that can enable heartfelt-yet-civil discussion about Israel gave her hope and energy.

As Hillel president she focused on making those on the margins feel welcome. She hosted the first ever pre-Passover bagel brunch, with over 250 students attending (most of whom were not "the usual suspects"). She has also convinced the Hillel professional staff to develop programs that provide small-group settings for Israel conversation to bring people together.

One last point: Julie went to college intending to major in political science. Because of her exposure to a range of Jewish texts and ideas through BYFI, she decided to try Yiddish and is now a Yiddish major.

~Becky Voorwinde, Executive Director of The Bronfman Youth Fellowships

Step 2: Open the Evaluation Toolbox: Utilize a Range of Methods to Collect and Analyze Data

If the first step is articulating meaningful outcomes and the means by which we intend to achieve them, the next step is to open the evaluation toolbox to find methods and tools that capture and analyze data relevant to those outcomes.

The central task of this step is to match one or more evaluation tools to your desired outcome(s). To do that, you need to explore the evaluation toolbox, and then reflect on how you would use the data that a specific tool would generate. So first we'll introduce you to the toolbox, and then we'll offer a set of questions to reflect on as you consider potential evaluation methods.

One of the most exciting moments of the seminar was when our two evaluation experts walked us through a set of 15 evaluation tools they and other experts have developed for program providers. While it is difficult to recreate that experience in a written report, we believe that tasting the variety of evaluation methods that are available is an important experience in the process of improving our approach to evaluating/assessing our programs.

Below are brief descriptions of a handful of these evaluation tools. At the end of this report, you'll find a brief summary of all 15 different approaches, a chart you can use to identify additional approaches that you may wish to explore. Depending on the evaluation experience of your staff, you may decide to hire an evaluation consultant to identify—and support you in using—the most relevant tools to collect data related to your desired outcomes.

Identity Leadership Inventory

Use the Identity Leadership Inventory to measure a leader's ability to mobilize and direct followers' energies.

The Identity Leadership Inventory (ILI) is a simple, validated[3] survey instrument designed to measure a leader's influence on her or his community vis-a-vis four interconnected relationships:

- The leader's ability to "be one of us"
- The leader's ability to "do it for us"
- The leader's ability to "craft a sense of us"

- The leader's ability to "embed a sense of us"

By deploying the ILI in a leader's community before, during, and after a leadership-training intervention, you can infer some of the intervention's value in supporting the leader's influence in her/his community. You can find the ILI survey instrument and guidance for administering it in Niklas K. Steffen, et al's
"Identity Leadership Inventory (ILI) Instrument and Scoring Guide (ILI Version 1.0)," downloadable here: https://www.researchgate.net/publication/271071707_ILI_Guide

Outcome Harvesting

Use Outcome Harvesting to identify changes in a complex system and work backwards to determine whether and how an intervention contributed to those changes.

The **outcome harvesting** methodology is unusual in that it does not seek to measure progress toward a predetermined set of outcomes. Instead, it starts by identifying changes (e.g., shifts in communal behavior, adoption of new regulations/legislation, emergence of new memes/messages, etc.) within a system affected by an intervention (e.g., a community whose leader has experienced a leadership-training program) and then works backward to determine whether there's credible evidence that the intervention contributed to the changes. It seeks to answer the following questions:

- What happened?
- Who did it (or contributed to it)?
- How do we know this? Is there corroborating evidence?
- Why is this important? What do we do with what we found out?

It's particularly applicable in contexts in which the connections between cause and effect are ambiguous. You can find detailed guidance on the outcomes harvesting methodology in Ricardo Wilson-Grau and Heather Britt's "Outcome Harvesting," downloadable here: http://www.outcomemapping.ca/download/wilsongrau_en_Outome%20Harvesting%20Brief_revised%20Nov%202013.pdf

Most Significant Change

Use Most Significant Change to assess not only what outcomes a program has achieved, but also how various stakeholders value those outcomes in different ways.

The **Most Significant Change** (MSC) technique is a highly participatory methodology designed to engage a broad cross-section of stakeholders in the evaluation of a given intervention. In the context of a leadership-training program, members of the leader's community would be invited to share stories of what they consider to be significant changes resulting from leader's participation. This process of sharing not only provides useful evidence for the impacts of the intervention, but also surfaces valuable insights into what community members consider "significant." Once these stories have been collected, a group of select stakeholders (staff, advisors, etc.) discuss the stories and determine which they consider "most significant," along with describing a rationale for their determinations. These analyses are then fed back to the community to focus its attention on impact, shared with leaders, and used to refine the training program. You can find detailed background information and implementation guidance in Rick Davies and Jess Dart's The 'Most Significant Change' (MSC) Technique: A Guide to Its Use, downloadable here: http://www.mande.co.uk/docs/MSCGuide.pdf

Collective Efficacy Scale

Use a Collective Efficacy Scale to assess a group's shared belief in its joint capability to organize and execute a course of action in pursuit of its desired results.

The **Collective Efficacy Scale** emerged from criminal-justice research into how communities' shared values and willingness to cooperate could exert influence on community members to reduce violence and crime. It has been adapted for use in schools and many other contexts in which community members' sense of themselves as a mutually supportive community capable of collective action is important. By deploying the Collective Efficacy Scale in a leader's community before, during, and after a leadership-training program, you can infer some of the program's value in building the community's sense of efficacy. You can find more information on the Collective Efficacy Scale in John M. Carroll, Mary Beth Rosson, and Jingying Zhou's "Collective Efficacy as a Measure of Community," downloadable here: https://www.researchgate.net/publication/221519069_Collective_efficacy_as_a_measure_of_community.

Ripple Effects Mapping

Use Ripple Effects Mapping to reflect upon and visually map the intended and unintended changes produced by a complex program or collaboration.

Ripple Effects Mapping (REM) is a participatory qualitative evaluation method in which stakeholders collectively and visually map the interconnected outcomes resulting from a program or intervention. In an REM process, participants come together to reflect on the ways in which a program has affected their lives and the lives of others. Individual impacts are captured using mind-mapping software or on a large white board, and the group works together to identify and map connections among these impacts to generate a collective representation of the ripple effects of the program or intervention. For more information and a practical guide for using REM, see Washington State University-Extension's "Ripple Effects Mapping for Evaluation," downloadable at: http://extension.wsu.edu/stevens/wp-content/uploads/sites/19/2013/12/REM.Complete.pdf.

More details and more tools are included in the section titled "15 Evaluation Tools to Consider" near the end of the report.

TEST THE TOOLS ON YOUR OUTCOMES

Once you have identified one or two evaluation tools of interest, you can test them out on your outcomes. Do this by reflecting on the following questions:

1. What outcome(s) do you want to measure using this evaluation method?
2. How are you planning to use the approach? Who will be the informants? How will you collect the data?
3. Imagine a set of data you might receive from this method. What would you do with the data? How would it affect the decisions you make about the program?

Note that each of these methods also introduces a frame you can apply to your evaluation even if you aren't using the full method. For example, you can learn about the types of questions you might ask to capture emergent outcomes by becoming familiar with outcome harvesting, even if you don't plan to do the full verification process that is part of that technique.

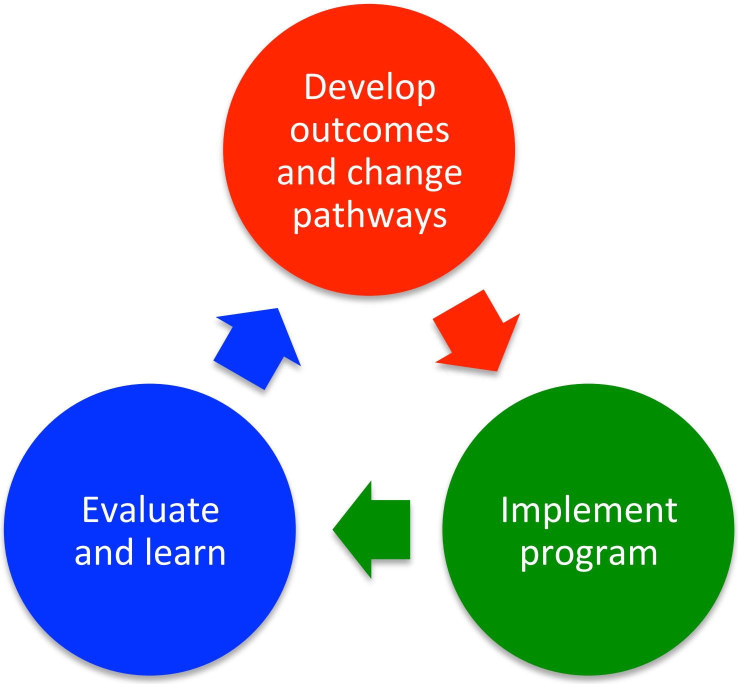
Evaluation experts offer various guidelines for selecting appropriate evaluation tools. No tool is perfect, comprehensive, or appropriate for all contexts, so deciding upfront what is most important about the tool can help. Some of the key criteria to consider in choosing evaluation tools include:

- **Accuracy**: The accuracy of your evaluation data is highly dependent on the constraints under which you are operating (e.g., having access to relevant respondents at the right time).
- **Timeliness**: Data that won't be available in time to meaningfully inform decisions won't be that helpful.
- **Actionability**: We often collect data that are interesting instead of data that are actionable. Consider a dress rehearsal to test out the usefulness of your expected data. Imagine results from a method, and then imagine what you would do differently based on having those data. If you can't figure out what you would change, those data may not be useful.
- **Impartiality**: Avoid asking leading questions or questions that assume positive or planned outcomes.
- **Representative of multiple perspectives**: Identify key perspectives you need to further develop the program (e.g., end users, funders, community stakeholders, etc.), and find a tool that will deliver those perspectives.

Note that these are criteria to balance—your tools need to meet all of them to a certain degree, but you will also have to balance how far to go (e.g. the most timely data collection may not be the most accurate or represent the most perspectives, but you don't want to go so far into being accurate and representative that the data doesn't show up until it's too late to use it).

STEP 3: APPLY LEARNINGS FOR CONTINUOUS PROGRAM IMPROVEMENT

This is perhaps the simplest and yet most overlooked step in the program evaluation process. The learning generated through evaluation processes should be fed back into the program design and process to ensure a continuous cycle of program improvement. There's no "magic bullet" methodology for this—it's mostly a function of setting aside time for a planning team to reflect on the evaluation and identify ways in which the program can be refined to improve effectiveness. This should be an ongoing cycle in which the program design (outcomes and change pathways), and learning and evaluation methods are regularly interrogated and refined to ensure that the overall intervention is continually improving.



We hope this report has stimulated your thinking, and your practices, around evaluating your own leadership program.

As your evaluation practices evolve, you may wish to deepen your learning about evaluation methods. In the addenda you'll find a short list of resources to consider. You may also wish to partner with other organizations and pool resources or explore joint evaluation efforts. Could you design or use an instrument together? Work with a consultant together? Share data you have gathered and the learning it has offered?

Ultimately, evaluation is a core part of learning and improvement. Good evaluation gives us data to reflect on, helps us imagine better ways to make an impact, and guides our ongoing efforts at improvement.

NOTE TO FUNDERS

Donors and funders have a significant influence on the evaluation practices of nonprofits.

In most cases, this influence is very positive. The questions embedded in a grant application and the inquiries of a program officer about evaluation practices force program designers to reflect on the outcomes they hope to achieve and articulate how they hope to measure progress toward those outcomes. But sometimes a funder's requests for evaluation lead to a culture in which program directors feel pressure to show how a pre-determined target was achieved. When a program is situated in a complex system, such a linear approach to evaluation can shut down important learning and reflection that will lead to breakthroughs later on. We encourage funders and nonprofit staff to engage in open conversations about approaches to evaluation that invite curiosity and excitement and rather than compliance-driven fear. By committing to outcomes, independent of the success of any particular programmatic approach, and encouraging honesty, collaboration, and innovation, funders can shape a larger field of practice. Some funders we work with take a learning-focused approach and use language like: "We care more about making an impact than the particular success or failure of your program. We want to learn what you are learning and find out how to make a difference. If this program doesn't work well, we know we will all learn a lot from the experience, and then we can try something else together."

ACKNOWLEDGEMENTS

This report was produced with the support of a Wexner Graduate Fellowship Alumni Collaboration Grant.

The authors express gratitude to the Wexner Foundation for their investment in leadership development and for the financial support that made this initiative possible.

Authors of this report:
- Rabbi Justus Baird is Dean of Auburn Seminary.
- Aaron Dorfman is President of Lippman Kanfer Foundation for Living Torah. He was Vice President for National Programs at American Jewish World Service during phase 1 of this project.

We express appreciation to the almost two dozen staff who oversee leadership development programs who came to the April 2015 seminar that led to the development of this report:

1. Sarra Alpert, AVODAH: The Jewish Service Corps
2. Justus Baird, Auburn Seminary
3. Lauren Berkun, Shalom Hartman Institute
4. Shifra Broznick, Advancing Women Professionals
5. Aaron Dorfman, American Jewish World Service
6. Mirele Goldsmith, Hazon
7. Irit Houvras, American Jewish World Service
8. Leora Isaacs, SYNERGY
9. Steven Jacobson, Dorot
10. Jason Kimelman-Block, Bend the Arc
11. Kate Lauzar, SYNERGY
12. Sharon Miller, Auburn Seminary
13. Lev Nelson, T'ruah: The Rabbinic Call for Human Rights
14. Heidi Rosbe, Encounter
15. Nadia Roumani, American Muslim Civic Leadership Initiative
16. Shauna Ruda, JDC Entwine
17. Rachel Jacoby Rosenfield, American Jewish World Service
18. Judith Samuels, SYNERGY
19. Chris Scharen, Auburn Seminary
20. Rebecca Sirbu, CLAL: The National Jewish Center for Learning and Leadership
21. Emily Soloff, American Jewish Committee
22. Becky Voorwinde, The Bronfman Youth Fellowships

15 Evaluation Tools To Consider

Use this 2-page chart of Evaluation Tools to explore more methodologies for evaluating the secondary impact of your leadership development program.

Tool	Setting	# of Informants	Type of Information to Collect
Survey	On paper, by email, or by phone.	Typically a larger number of participants (20+).	Close-ended questions and open-ended questions tailored to capture data to learn about one or more outcomes.
Interview	By phone or in person.	Typically a small number of participants (around 5 – 15).	Open-ended questions that elicit more explanation and information than surveys usually provide, tailored to capture data to learn about one or more outcomes.
Focus Group	In person.	Typically a small number of participants (5 – 10).	Small number of open-ended questions designed to elicit dialogue among informants. Provides very in-depth information to learn about one or more outcomes.
Document Review	In your office!	No participants, but data is usually formal documents, including organization plans, earned media, policies, etc.	Can be analyzed by counting key pieces of information and/or summarizing the narrative, tailored to capture data to learn about one or more outcomes
Observation	In real time, as an action is being taken.	One or two observers, collecting data about everyone being observed (or listened to).	Counts and qualitative narrative regarding easily observable actions during a group event or activity of any type, tailored to capture data to learn about one or more outcomes.
Case Study	Mix of settings.	Typically includes a mix of document review, observation, interviews, and/or focus groups.	In depth learning from multiple perspectives about a specific example of change happening (e.g. a subset of leaders who worked together or a specific leader), tailored to capture data to learn about one or more outcomes.
Outcome Harvesting	Group dialogue.	Typically the team of program strategists with invitees who have a good perspective on the program	Participatory technique to identify, formulate, verify, and make sense of outcomes when relationships of cause and effect are unknown. Good for surfacing unplanned outcomes.

This material is adapted from data collection tools presented as part of Spark Policy Institute's advocacy evaluation toolkit.

Tool	Setting	# of Informants	Type of Information to Collect
Ripple Effect Mapping	Paired interviews and group dialogue	Typically a large group of program participants and members of the larger network.	Appreciative inquiry-style interviews to elicit stories of success followed by a group mapping of "ripples" of change. Captures a wide range of outcomes that may have happened in different locations, as well as group hypotheses about what likely secondary effects those outcomes will have.
Social Network Analysis	On paper, by email, or by phone.	Can be surveys or interviews, depending on use of information.	Helps to understand relationships between individuals or organizations, the resources leveraged in those relationships, and can also provide insight into the why and how behind the relationships when done qualitatively.
360s (e.g. Servant Leadership Questionnaire)	On paper, by email, or by phone.	A pre-defined interview instruments deployed to multiple audiences of the leader.	Gathers perspectives on leader character and performance from those in direct contact with the leader on a number of different dimensions, as well as from the leader her or himself.
Identity Leadership Inventory	On paper, by email, or by phone.	A pre-defined interview instruments deployed to multiple audiences of the leader.	A leadership inventory based on social identity theory that captures how a leader affects group cohesion and identity, embodies the group's sense of "we", and creates platforms or vehicles for the group to act on its principles and interests.
Collective Efficacy Scale	On paper, by email, or by phone.	A pre-defined survey instrument deployed to multiple audiences of the leader.	Demonstrates the extent to which a group has confidence that it is equipped to achieve its goals together. Collective efficacy is predictive of success because it increases a group's ability to overcome adversity and maintain commitment.
Appreciative Inquiry Interviews	By phone or in person.	Interviews conducted with audiences of the leader.	Focuses on surfacing past experiences, in this case with the leader, that are important repeat in the future (very focused on strengths).
Discourse/ Frame Analysis	On paper, by email, or by phone.	Interviews, surveys or focus groups	Explores the frames or narratives people use to understand an issue, a group, or themselves, and/or the extent to which new frames and narratives resonate with their existing values.

Recommended Resources

American Evaluation Association (http://www.eval.org/)
The American Evaluation Association's mission is to improve evaluation practices and methods, increase evaluation use, promote evaluation as a profession, and support the contribution of evaluation to the generation of theory and knowledge about effective human action.

Center for Evaluation Innovation (http://www.evaluationinnovation.org/)
Our aim is to push philanthropic and nonprofit evaluation practice in new directions and into new arenas. We specialize in areas that are challenging to assess, such as advocacy and systems change.

Free Resources for Program Evaluation and Social Research Methods
(http://gsociology.icaap.org/methods/)
This page lists free resources for Program Evaluation and Social Research Methods. The focus is on "how-to" do program evaluation and social research: surveys, focus groups, sampling, interviews, and other methods. Most of these links are to resources that can be read over the web.

Spark Policy Institute (http://sparkpolicy.com/)
Spark Policy Institute partners with stakeholders throughout the country to develop innovative, research-based solutions to complex societal problems. Spark combines community and stakeholder-driven research with practical, hands-on experience and best practices, allowing for solutions that bridge sectors, issues, beliefs, and values. By integrating diverse policy systems, the team at Spark identifies and develops the best solutions for all stakeholders.

Notes

[1] There's a tendency in program evaluation circles to think that, in order to be credible, a program needs to demonstrate and account for the full chain of causation from activity through outputs/outcomes to impact for each discrete program or intervention. The "golden spike" analogy (evoking the final connection between Central Pacific and Union Pacific Railroads to complete the First Transcontinental Railroad across the United States) short circuits this presumption by arguing that an evaluation can (and should) draw on existing research for analogous proof points to connect activities to ultimate impact without having to prove the full causal chain for each discrete intervention. Say, for example, that you're concerned about childhood obesity. You design an intervention that replaces processed snacks with fresh fruit and vegetable snacks in a local middle school cafeteria. The golden spike concept suggests that if you can demonstrate that students are consuming more fruits and vegetables and fewer processed snacks as a result of your intervention, and you can cite to research that demonstrates that children who increase their consumption of fruits and vegetables and decrease their consumption of processed snacks experience reductions in obesity, you don't need to prove that your intervention reduces obesity. You can simply rely on the "golden spike" connection between what you're doing and what the research indicates will happen as a result. The willingness to use the "golden spike" has important implications for evaluation budgets, timelines, participant resources, and ability to produce results that will be seen as credible and useful.

[2] Frances Westley, Brenda Zimmerman, and Michael Q. Patton, Getting to Maybe: How the World is Changed (Toronto: Vintage Canada, 2007).