APPLYING EVALUATION THINKING AND PRACTICE TO FORESIGHT

December 2024

ASSOCIATION OF
PROFESSIONAL
FUTURISTS

AUTHORS

Lead of APF Foresight Evaluation Task Force and Initiative

ANNETTE L GARDNER, PHD, MPH

Dr. Annette L Gardner is the Principal of ALGardner Consulting. She is currently developing resources to support foresight evaluation practice and evaluator foresight capacity.

APF Foresight Evaluation Task Force & Initiative Member

LAURENT BONTOUX, MS, PHD

Dr Laurent Bontoux is a Senior Foresight for Policy Expert at the EU Policy Lab of the European Commission's Joint Research Centre where he works to provide the best possible foresight evidence to EU policymaking.

APF Foresight Evaluation Task Force Member

ERIC BARELA, PHD

Dr. Eric Barela is a Senior Consultant with Raya Cooper Impact Consulting. He is currently combining effective evaluation and foresight approaches to provide a more comprehensive toolkit for evaluators and foresight practitioners.

For more information about this guide, please contact Annette L Gardner at: algardner.consulting@gmail.com.

ACKNOWLEDGEMENTS

ACKNOWLEDGEMENTS

APF Foresight Evaluation Task Force Members (2021-2023)

ACKNOWLEDGEMENTS

APF Foresight Evaluation Initiative Steering Committee (2024)

Annette
Gardner

Dr Laurent
Bontoux

Zan
Chandler

Rose
Thompson Coon

Tanja
Hichert

Steven
Lichty

Jean
Paul Pinto

Bruce
Tonn

TABLE OF CONTENTS

Abstract

A perennial concern of futurists and foresight practitioners is whether or not foresight work is having the desired impact and is viewed as credible. Recognizing the need for strengthened foresight practitioner evaluation capacity as well as increased assessment of foresight projects and activities, the APF Foresight Evaluation Initiative is developing resources to guide evaluation of foresight initiatives. In this practical guide, we curate and describe useful evaluation constructs and methods at each stage of the evaluation process that are applicable in diverse foresight settings. Second, we describe the contours of foresight, including how foresight is different from other types of initiatives and the challenges to evaluation, such as a long time-horizon and uncertainty. Where possible, we draw on the foresight evaluation literature and describe tailored measures, lessons learned from cases, and suggestions for outcomes and impacts. This guide is primarily intended for foresight practitioners with evaluation expertise and evaluators who are commissioned to evaluate foresight initiatives. However, it is also useful for orienting practitioners new to evaluation, deepening their understanding of the usefulness of engaging in evaluative thinking.

Background

Foresight evaluation is having a moment, characterized by increased demand for assessment of foresight initiatives and the growing realization by foresight practitioners that evaluation can greatly benefit their practice. In many situations, evaluation of foresight programs and activities will resemble traditional program evaluation.[1] Foresight evaluation involves assessing the quality and impact of foresight activities, which can include developing and using possible images of the future, identifying and monitoring signals and forces for change in the environment, and educating individuals, communities, and public and private sector organizations on how to use foresight. Foresight[2] evaluation designs will have similar parameters as designs of other types of programs—a time horizon, evaluation questions and methods, a budget, and scope of work. Many of the same methods will be applicable and will generate useful findings, including surveys, case studies, document analysis, and focus groups. The general rules for systematically assessing a program's merit will also apply (such as adhering to ethical research standards, use of social science methods, achieving a high degree of rigor, addressing stakeholder information needs, and communicating findings that inform decision-making).

That said, foresight evaluation needs to address specific challenges such as complexity, long-time horizons, the qualitative nature of the work, and documenting largely intangible and multiple foresight outcomes. However, these challenges are not unique to foresight work, and are shared by other types of programs that don't stabilize, such as advocacy and policy change initiatives that influence public and policymaker thinking and support. With respect to the latter, appropriate designs and methods have been developed by a dedicated group of evaluators and funders over the last 20 years. In all likelihood, foresight evaluation will follow a similar trajectory, or evaluators will identify the unique aspects of a foresight initiative and develop (or adapt) designs and methods accordingly. There will also be discussion about the purpose of evaluation at the field level, including research questions that are important to building the knowledge base and/or important to improving foresight quality and professional excellence, balancing of attribution and contribution, and long-term goals for building field evaluation capacity.

---

1. "Evaluation" is defined broadly in this paper or: "...to provide (credible) answers to questions about a program (or activity) that will be useful…to inform action." Source: Rossi, PH, Lipsey, MW. And Freeman, HE. (2004). Evaluation: A Systematic Approach. 7th Edition. Thousand Oaks, CA: Sage Publications. "Program" could refer to foresight trainings, process such as alternative scenarios, products such as movies, etc. "Research" is different from evaluation in that it engages in testing a hypothesis and generates knowledge to inform a research base.
2. By 'foresight,' we mean " …the capacity to thinking strategically about the future." Shallowe et al. (2020). *A stitch in time? Realizing the value of futures and foresight.* RSA. October 2020. p. 76.

More broadly, the evaluation field has been evolving and adapting to the changing times and has made significant progress towards being able to systematically assess programs, policies, and activities of all types. Evaluators know that there is no "one-size-fits-all" approach to evaluating programs and policies. They are comfortable with different worldviews, unique contextual factors, and mixing and matching constructs and methods. For example, they can determine whether an experimental or non-experimental design is appropriate in a particular setting, such as a randomized control trial (RCT) to test the effectiveness of new medical treatments or a rapid assessment process (RAP) approach for programs in developing countries where resources are limited. Additionally, they know that it is not solely the initiative which drives evaluation design, but also stakeholder information needs, which can vary greatly. Last, evaluators are very methods-friendly and can readily develop and/or adapt methods (such as big data analysis, process tracing, outcome harvesting and GIS mapping).

In this paper, we highlight evaluation theories and practices that are important to keep in mind when designing a foresight evaluation, calling out specific constructs and methods that are useful in diverse foresight settings. For example, in public sector foresight it is important to be mindful of not just government foresight, but of the political context, structure of government, and government bureaucracy. For foresight trainings and workshops, there are useful educational evaluation frameworks and metrics that can be adapted to a short-term training setting, such as Kirkpatrick's 4-stage model, Ahvenharju's *Futures Consciousness Scale*,[3] and the European Commission's *foresight competence progression model*[4] as part of the competence framework for innovative policymaking.

This paper is intended for evaluators and foresight practitioners with evaluation expertise. It will be particularly helpful for evaluators new to foresight and to help them navigate the foresight space. For foresight practitioners with limited or no evaluation expertise, this paper will not fully address their learning needs, but we believe describing the contours of foresight evaluation can provide a foundation for getting started.

3. Ahvenharju, S., Minkkinen, M., & Lalot, F. (2018). The five dimensions of Futures Consciousness. *Futures* 104: 1-13.
4. European Commission. (2023). Competence framework for 'innovative policymaking'. https://knowledge4policy.ec.europa.eu/visualisation/competence-framework-innovative-policymaking_en

Designing an Appropriate Evaluation

This section describes the mainstream and emerging methods germane to foresight evaluation and provides recommendations from the authors' experiences with evaluation capacity building (ECB) and with advancing evaluation of new topical areas, such as advocacy and policy change evaluation.

Getting Oriented

First, evaluators of foresight should take some time to reflect on their own evaluation practice and their strengths as well as where they might need to bolster their expertise. They will want to examine their design and methods "toolkit" and make sure they're up to date on the latest advances in evaluation thinking and practice, particularly methods to address program complexity, models and frameworks to advance equity, and new ways to increase the usefulness of the evaluation to program participants and funders. Also, evaluators will want to bring to bear their "evaluative thinking" mindset, or: " ...identifying assumptions, posting thoughtful questions, pursuing deeper understanding through reflection and perspective taking, and making informed decisions and preparation for action."[5] This is particularly useful for undertaking and advancing foresight evaluation, which many times will have a participatory component that may also be cross-cultural and multi-disciplinary.

Before launching a foresight evaluation, evaluators would be wise to undertake an evaluability assessment (EA) to determine how ready a foresight activity is for evaluation. EA is an assessment of the extent to which an intervention can be evaluated in a reliable and credible fashion. The results of an EA should indicate what preparatory steps are needed to make an intervention ready for an evaluation and what remaining challenges will need to be addressed by an evaluation team. An EA will examine: (a) the adequacy of the intervention's Theory of Change or objectives, (b) the availability of relevant data and supporting systems, (c) stakeholders' interests in the evaluation, especially their evaluation questions, and (d) constraints arising from the surrounding institutional context.[6]

Where programs can be evaluated, the evaluators of foresight are encouraged to consult the literature on existing foresight evaluation frameworks and cases that can provide useful lessons learned and technical guidance. For example, Makarova and Sokolova (2014) and Ko and Yang (2024) conducted a detailed literature review of evaluations of foresight projects.[7] Makarova and Sokolova went a step further and developed an evaluation model that builds on project evaluation, but takes a tailored approach to evaluating foresight, such as using primarily qualitative methods.[8] Poteralska and Sacio-Szymanska (2014) examined national and transnational foresight projects and provide a review of systemic foresight evaluation frameworks and their application.[9] Last, Georghiou and Keenan's (2006) seminal article describes the use of evaluation strategies in evaluating national foresight activities and strengthening foresight as a policy tool.[10] The article provides a thoughtful overview of the reasons for evaluating foresight or accountability, justification, and learning; a key first step in designing an evaluation. The challenges to evaluating foresight are carefully laid out, providing evaluators a 'map' for matching the evaluation

5. Vo, A. T., Schreiber, J. S., & Martin, A. (2015). Citing Buckley et al. Toward a Conceptual Understanding of Evaluative Thinking. *New Directions For Evaluation*. No. 158. Summer 2018. p. 31.

6. Davies, R. (2013). *Planning Evaluability Assessments: A Synthesis of the Literature with Recommendations.* Department for International Development. Working Paper 40.

7. Kwon Ko, B., & Yang, J. (2024). Developments and challenges of foresight evaluation: Review of the past 30 years of research. *Futures,* 155 103291.

8. Makarova, E. A., & Sokolova, A. (2014). Foresight evaluation: lessons from project management. *Foresight.* Vol. 16, No.1 2014. 75-91.

9. Poteralska, B., & Sacio-Szmanska, A. (2014). Evaluation of technology foresight projects. Eur J Futures Research, 15:26.

10. Georghiou, L., & Keenan, M.(2006). Evaluation of national foresight activities: Assessing rationale, process, and impacts. Technological Forecasting and Social Change.73 761-777.

approach to the rationale. Three cases of evaluation of national foresight activities are described, allowing for a comparison of evaluation designs and challenges. Evaluators outside the foresight arena will find the article helpful in understanding foresight as an evaluand (the thing that is being evaluated) and the likely dimensions of an evaluation design, such as the primacy of process evaluation versus impact or summative evaluation.

When consulting the literature on existing foresight evaluation frameworks, impacts, indicators, and methods, it's important to tap into the literature on theory (or theories) that can inform the development of a Theory of Change (ToC) and ground the evaluation in the scholarship. Areas that are germane to foresight evaluation include:

- Organizational development (OD) theories on leadership, innovation, and adaptation;
- Behavior theories from the psychology, sociology, and public health arenas.[11]
- Political science and public policy theories on power, influence, democracy, and representation; and
- Education and learning theories, such as cognitivism and social learning.

Clarify the Initiative Goals, Objectives, and Activities

Specifying the evaluand and what is being evaluated is critical to the evaluation design. The complexity and maturity of an initiative's strategy, specific foresight tactics, and levels of engagement (international, national, state, and local) are linked to specific methods and tools of the evaluation design. There are evaluation frameworks and tools to help with this. For example, the Institute for Development Research (IDR) framework uses five dimensions to help bound the universe of change initiatives and their impacts and help with method selection: 1) policy (national, provincial, local, international, other); 2) private sector (national, local, international, multinational, other); 3) civil society (NGOs, popular organizations, community-based organizations, ally organizations, others); 4) democracy or political systems and culture (democratic space, participation of civil society, political legitimacy of civil society, accountability of public institutions, transparency of public institutions, other); and 5) individual wellbeing (material, attitudinal, other).[12] This is a useful framework for public sector foresight even if it is not directly involved with a specific policy arena. The foresight sponsors and project participants are, however, intrinsically linked to a complicated and fluid policy landscape.

As is the case in the evaluation of most programs and services, the foresight evaluation design stems directly from an initiative's scope and purpose. The stated goals, objectives, and plan of action will provide much of this information, but one cannot assume that these goals and objectives are realistic or will not evolve over the course of the initiative. Understanding the different actors and their interests and information needs over time, such as what constitutes a "success," will help determine the outcomes and inform the model and level of stakeholder engagement in the evaluation design. Project participants and their funders (hereafter referred to as stakeholders) may have unachievable, aspirational goals and overlook the incremental gains that are significant achievements in their own right.

---

11. *Theory at a Glance: A Guide for Health Promotion Practice.* (2005). US Department of Health and Human Services. National Institutes for Health. National Cancer Institute.
12. Gardner & Brindis citing Chapman and Waymeyo in Gardner, A. G., & Brindis, C. B. (2017). *Advocacy and Policy Change Evaluation: Theory and Practice. Stanford University Press.*

Development of a program theory of change (ToC), a characterization of how change happens in a particular program or context, and/or logic model that describes the links between resources, activities, and outcomes is very useful in characterizing the evaluand. It is very likely that program planners and funders will have different understandings of how program elements relate to one another. Arguably, foresight can be emergent, making cause and effect difficult to discern. However, understanding the limitations of a theory of change and the constraints of a linear change model can be balanced with a nuanced understanding of the context and stakeholder perspectives. This is an area where there is increasing alignment between foresight and evaluation, with efforts underway to infuse foresight into developing a theory of change, particularly a deeper understanding of the anticipatory assumptions that went into developing a theory of change.[13]

Useful Frameworks and Approaches at the Evaluation Design Stage

The foresight field consists of different kinds of approaches (and paradigms) and foresight can have many kinds of goals. How to choose an appropriate evaluation approach and methods for your foresight? What are the relevant questions that guide designing evaluation, meaning what kind of evaluation questions could be set and how to produce evaluation knowledge of them? The aim of this section is to help to identify a purposeful evaluation design and suitable methods based on the nature of foresight and help practitioners to navigate existing evaluation repositories. **Figure 1** describes a simplified framework developed by evaluators and futurists that provides evaluators with an easy-to-use starting point for designing an evaluation. The first task is to identify what is the purpose and focus of the foresight that will be evaluated? Second, what are the interests and purpose of the evaluation? What to evaluate and what is relevant? What are the criteria for success? Based on the answers to the first two questions and on the constraints linked to the evaluation exercise, choose the methods that are best suited to generate the desired answers.

---

13. Carden, F. (2023). Back to the future: are we trapped in the past? *Evaluation and Program Planning*. Vol. 97.

Figure 1: Framework for selecting an evaluation approach

What is the goal and focus of the foresight activity?	Support to (contingency) planning and strategy development	Visioning, Development and adoption of a shared vision	Transformation, Futures literacy, foresight capacity, connecting futures thinking to making change
What are the endpoints to evaluate?	Scenarios Drivers of change Milestones towards the future Applied analytics 'Systemicity' Contribution of foresight to a theory of change or other decision-making	Improvement of the coherence of action between the various 'silos' of an organization Long-term 'uptake' of the vision	Capacity
What kind of evaluation questions and criteria are typical/useful?	Methodological adequacy (fitness for purpose) How "future proof" is the strategy? How alternative developments have been considered? Preparedness/agility/adaptiveness of decision making and strategy Quality of process, (robustness, inclusiveness, etc.)	What was the value of vision: as a process and a product? Quality of process (robustness, inclusiveness, etc.) Commitment Coherence of action	How futures thinking has been linked to action? Changes in individual's/communities' futures thinking and capacity? Novelty in action? Decolonializing of futures? Diversity Inclusiveness
What kind of evaluation approach is suitable?		Qualitative case study Process evaluation Before and after assessment of coherence of action between 'silos'	Qualitative case study, stakeholder survey/interviews, Self-capacity measurement, ethnography, action research

Source: APF Foresight Evaluation Task Force: Work Group 3 Members Stephen Aguilar-Milan, Laurent Bontoux, Jay, Gary, Laurie Smith, and Katri Vataja

An appropriate evaluation approach for working with the emergent and complex nature of foresight work is *Developmental Evaluation (DE).* A type of participatory evaluation, Developmental Evaluation can assist with developing foresight initiatives in complex or uncertain environments. It takes stakeholder involvement to the next level with its emphasis on ongoing learning and adaptation, such as sense-making and working with participants to identify patterns, integrate new information, and consider the implications of what is happening. Evaluation is considered an essential, partnership-based function that is responsive to stakeholder learning and information needs. Ongoing discussions with participants, providing feedback, and course corrections are embedded in the evaluation design.[14]

Foresight work, such as developing alternative scenarios, can be uncertain both in outputs and the process itself; it can change as the context changes, such as abrupt changes like Covid-19 and climate events. An *adaptive management* approach may be helpful here. Going beyond the usual adaptation involved in good management and modifying plans in response to changes in circumstances, it includes understanding and using information to inform these decisions. It is particularly relevant in a monitoring and evaluation context and is a nimble approach to managing under conditions of ongoing uncertainty. It represents a paradigm shift from classic, linear approaches to planning, implementation, and evaluation. It is well-suited for the emergent aspects of foresight activities, such as alternative scenario projects where the outputs are unknown.[15]

Since much foresight work focuses on preferred images of the future, it is helpful to work with evaluation tools that are compatible with the aspirational aspects of foresight. Appreciative Inquiry and the proactive examination of positive images of organization-level activities put stakeholders more in the driver's seat. *Appreciative Inquiry* focuses on strengths rather than on weaknesses. It is particularly well suited for assessing organizational capacity and can be used to build relationships among stakeholders, strengthen stakeholder evaluation capacity, and guide the evaluation design and its implementation. Appreciative Inquiry is often presented in terms of a 4-step process around an affirmative topic choice: 1) Discover: What gives life? What is the best? Appreciating and identifying processes that work well; 2) Dream: What might be? What is the world calling for? Envisioning results, and how things might work well in the future; 3) Design: What should be the ideal? Co-constructing - planning and prioritizing processes that would work well; and 4) Destiny (or Deliver): How to empower, learn and adjust/improvise.[16]

Much foresight work is complex in design and/or focuses on complexity. *Complex Adaptive Systems* (CAS) constructs can be helpful for characterizing and working with this complexity in an evaluation design. A complex adaptive system is a system that is complex in that it is a dynamic network of interactions, but the behavior of the ensemble may not be predictable according to the behavior of the components. It is adaptive in that the individual and collective behavior mutate and self-organize corresponding to the change-initiating micro-event or collection of events.[17]

---

14. Patton, M. Q.(2011). *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. The Guilford Press.

15. Rogers, P., & Macfarlan, A. (2020). *An overview of monitoring and evaluation for adaptive management. Monitoring and Evaluation for Adaptive Management* Working Paper Series, Number 1.

16. Coughlan, A. T., Preskill, H., & Catsambas, T. T. (2003). An Overview of Appreciative Inquiry in Evaluation. *New Directions for Evaluation*, 100:5-22.

17. Sibthorpe, B., Glasgow, N., & Longstaff, D. (2004). Complex adaptive systems: A different way of thinking about Health care systems. The Australian National University.

Foresight is considered a type of systems thinking since it examines the relationships between drivers, trends, events, and signals of change. *Systems thinking* approaches and looking at an initiative as a web of relationships enables evaluators to understand an initiative in a holistic, dynamic way, and not overlook aspects and changes that are not included in a linear model. Using a systems lens has the potential to capture unintended changes as well as surface recommendations for program improvements and identifying levers for change. While systems thinking and thinking about relationships and connectedness goes back many decades, evaluation tools are still being developed. For example, Kumu, an online systems mapping tool, is readily available,[18] as well as Netlogo, an open-source tool widely used by complexity researchers.[19] Evaluators are adapting systems concepts and principles to characterize their initiatives and inform development of theories of change.[20] Regardless, it's helpful to have some mastery of systems thinking concepts and tools and see how they apply to foresight work.

A key aim of many types of foresight is to increase participant foresight capacity. Therefore, it is helpful to look at other areas where capacity has been examined and there are models on which to draw. A specific focus of evaluation and related to evaluation use, *evaluation capacity building* (ECB) is increasingly a component of many evaluations is being embedded in evaluation designs.[21] There is the potential for ECB measures and resources to be adapted to assessing organizational foresight capacity, complementing foresight constructs such as the futures consciousness scale developed by Sanna Ahvenharju and her colleagues[22] and the future preparedness model developed by Rohrbeck and Kum.[23]

Participatory foresight and approaches that focus on co-creating a preferred future by a community may require approaches that support participant engagement and co-ownership of the evaluation process. The evaluator is less of an authority figure and more of a facilitator and stakeholders have greater control of the evaluation. Such an increased stakeholder involvement approach is designed to provide groups with the tools and knowledge they need to monitor and evaluate their own performance and accomplish their goals. It is particularly suited to the evaluation of comprehensive community–based initiatives or place-based initiatives.[24]

Last, the call for evaluation to support and strengthen transformational change puts new demands on evaluators to be flexible in the face of change and adopt new ways of thinking or what is referred to as *transformational evaluation*. There are also calls within the evaluation arena to transform evaluation so that evaluations consider all interventions in their broader context and how they interact with human and natural systems.[25] Evaluation also needs to move beyond a focus on individual projects and their stated objectives to

18. Kumu io: https://docs.kumu.io/getting-started/kumu-101

19. NetLogo: https://ccl.northwestern.edu/netlogo/

20. Gates, E. F., Walton, M., Vidueira, P., & McNall, M. (2021). Introducing systems-complexity-informed evaluation. *New Directions for Evaluation*, 1:13-25.

21. Lemire, S., Peck, L., & Porowski, A. (2020). The Growth of the Evaluation Tree in the Policy Analysis Forest: Recent Developments in Evaluation. *Policy Studies Journal.* Vol. 48. No. A1.

22. Ahvenharju, S., Minkkinen, M., & Lalot, F. (2018). The five dimensions of Futures Consciousness. *Futures* 104: 1-13.

23. Rohrbeck, R., & Kum M.E. (2018). Corporate foresight and its impact on firm performance: A longitudinal analysis." *Technological Forecasting & Social Change*. 129: 105 – 116.

24. Fetterman, D., Rodrigues-Campos, L., & Zukoski, A.P. (2018). Collaborative, Participatory, and Empowerment Evaluation: Stakeholder Involvement Approaches. New York: NY. Guilford Press.

25. Van den Berg, R. (2021). Transformational Evaluation for The Crises of our Times. International Development Evaluation Association.

consider their impact on wider systems. We may be witnessing the emergence of post-normal evaluation or evaluation thinking and practice that is better at navigating change and complexity, less tethered to linear thinking and more able to collaborate and take into consideration alternative mindsets and world views.[26] To support this shift from transactional to relational evaluation, the Equitable Evaluation Initiative has developed an evaluation framework that includes useful principles to advance equity and ways to shift mindsets and address barriers.[27]

Determine Evaluation Purpose, Design, and Questions

Determining the purpose of the evaluation follows closely on the heels of developing a detailed understanding of the initiative. The evaluation theory tree in **Figure 2** can help clarify the intent of the evaluation and it provides a set of decisions that will shape the evaluation design. At the Roots level, it is important to know to what extent the evaluation will focus on Social accountability and the need to demonstrate positive outcomes or impact, Social inquiry and the systematic use of social science methods, and/or Epistemology and the focusing on the nature and validity of knowledge. The next level is the three evolving and overlapping Branches—Use branch, Methods branch, and Valuing branch. The Use branch and using evaluation findings to inform decision-making also includes process use or using evaluation for program operations. For foresight practitioners that are new to evaluation and are primarily interested in learning about their activities, this is a good starting point. The Methods branch covers not just the research methodology and design (experimental/quasi-experimental/non-experimental) but also theory-based evaluation, data collection, and qualitative and quantitative constructs. It is the "toolkit" of evaluation. Last, the Valuing branch is how evaluative judgements are done and by whom to determine a program's merit or goodness. For example, it is the branch where discussions on cultural competence in evaluation practice take place.[28]

As argued by Georghiou and Keenan (2006), based on their examination of three foresight programs in the United Kingdom, Germany and Hungary, the evaluation design will vary by foresight initiative.[29] However, there are some guidelines that are applicable in many foresight settings. Conventional wisdom says that if an organization's capacity is the key focus of an initiative, then a formative (or process) evaluation design is the appropriate evaluation strategy. This can be a very fruitful area of inquiry. Many foresight activities and products lend themselves to counting and ongoing monitoring, such as contacts by policymakers, number of times foresight project participants express satisfaction, and attendance. The formalized tracking of tactics and changes that are characteristic of Monitoring, Evaluation, and Learning (MEL) practice can also be useful. Additionally, utilization-focused evaluation and asking and answering actionable questions, particularly as an initiative evolves can reconcile differences in stakeholder information needs.

While summative evaluation design and the assessment of whether a program worked or not is difficult in situations where there is a short time horizon and/or lack of measures, evaluators have multiple options for developing a sound summative evaluation design. Focusing on interim outcomes that are under the control of an initiative, such as changes in participant knowledge and support, can provide useful information about program effectiveness. Approaches well suited for complex environments, such as systems thinking, can reduce the uncertainty and make more transparent the linkages between initiative elements.

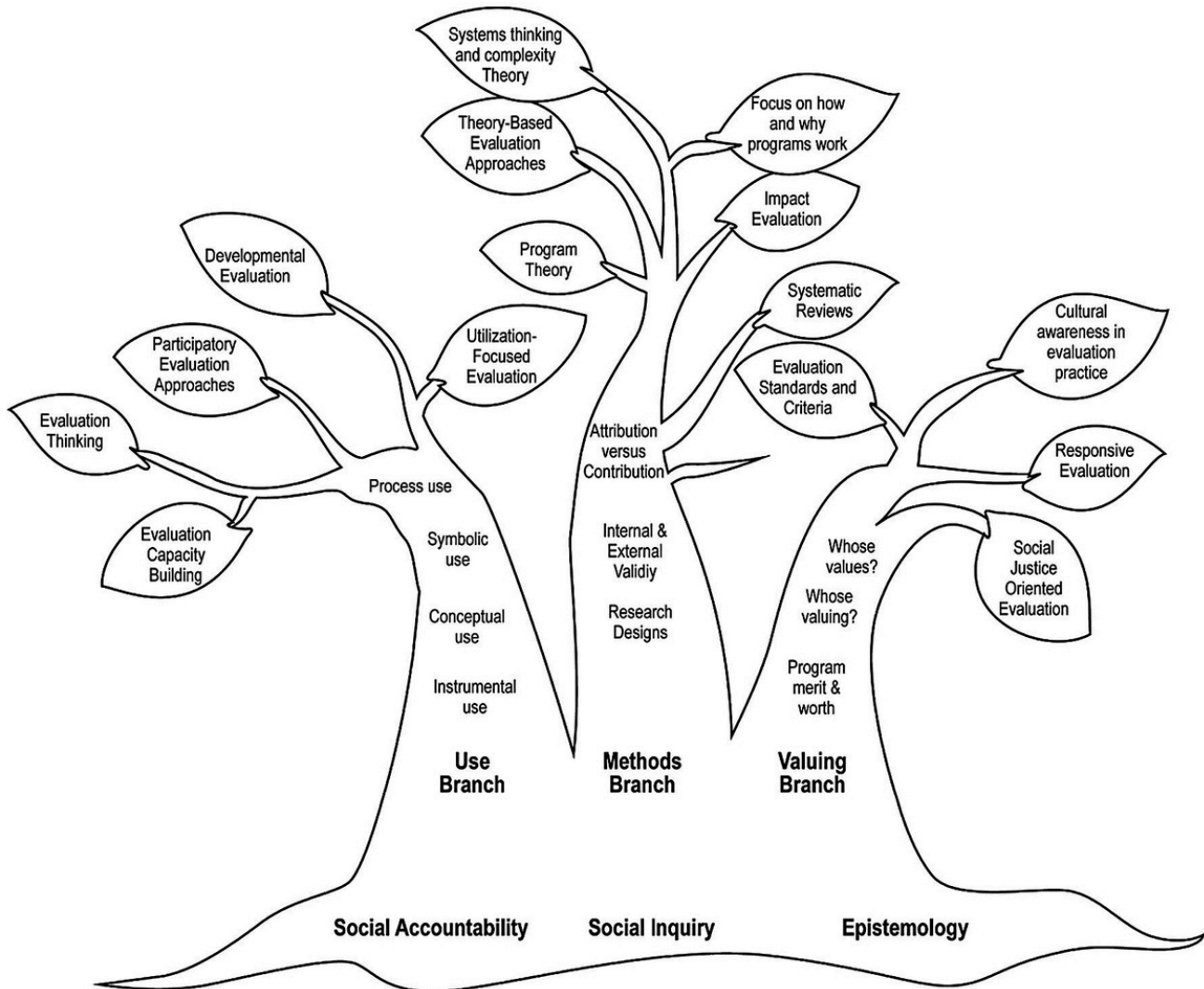26. Schwandt, T. A. (2019). Post-normal Evaluation? *Evaluation*. Vol. 25(3) 317-329.
27. *The Equitable Evaluation Framework.* Equitable Evaluation Initiative. 2023.
28. Lemire, Peck, & Porowski 2020, op. cit.
29. Georghiou & Keenan 2006, op. cit.

Figure 2: Evaluation Theory Tree

Source: Lemire, S., Peck, L., & Porowski, A. (2020). The Growth of the Evaluation Tree in the Policy Analysis Forest: Recent Developments in Evaluation. *Policy Studies Journal*. Vol. 48. No. A1.

The design also depends on program stability and whether an initiative becomes a replicable model or if it continues to transform and adapt to changing circumstances. For those situations where the path forward is not so clear-cut, a developmental evaluation approach and real-time data collection provides evaluators with the means to assess progress and inform strategy. For example, this approach can accommodate the changing terrain that characterizes a public sector foresight initiative as well as strengthen the partnership between the evaluator and stakeholders.[30]

---

30. Patton 2011, op. cit.

Additionally, the varying information needs by stakeholders and sectors will shape the evaluation purpose. At the program level, the conventional focus on assessing program success and whether the original objectives were achieved may be less important to program managers who are more interested in the ongoing monitoring.[31] There are also differences in information needs by sector—government, education, corporations, nonprofits, and communities. In government, stakeholders may be interested in accountability and whether foresight as an intervention made a difference, and whether foresight activities have influenced short-term outcomes, such as public and policymaker awareness. In foresight education, the focus is on assessing student foresight learning outcomes and competency with new measurement tools that assess changes in student futures literacy, readiness, and consciousness, to name a few. Corporations and nonprofits are interested in the quality of the foresight process and/or whether it achieved its desired impacts, including informing strategy and future-proofing actions. Communities may engage in participatory foresight with the aim of co-creating a preferred future. Methods to assess strengthening of civil society, influencing decision-making, and development of creative and visionary images of the future are useful here. Last, at the field level, the purpose of the evaluation may focus more on generalizability and whether a type of foresight makes a significant difference to outcomes important to the field, such as social action, informing strategy and decision-making, and changing peoples' lives.

The evaluation questions, the pillar of an evaluation design, will flow from the evaluation purpose. Evaluator Eleanor Chelimsky describes four types of questions: 1) descriptive or "how" and "what" questions; 2) "normative" questions or demonstrating program outcomes as compared to a standard; 3) "cause-and-effect" questions that focus on attribution; and 4) "knowledge-based" questions, such as lessons learned. However, similar to the evaluation purpose, client information needs ultimately decide the evaluation questions.[32] A foresight practitioner will want to inquire on program quality and areas for improvement and what was achieved by their activities, as well as keep in mind the information needs of commissioners of foresight. Many funders of foresight want to know whether a foresight initiative has been implemented in the way it was intended, the quality of implementation, if an initiative achieved its outcomes, and what role their support played. Depending on the maturity of the initiative, they may also want to learn whether an initiative is applicable elsewhere or if it is adapted elsewhere, what the likely outcome will be. Additionally, an analysis of the economic gains may be required by a funder to justify the initiative to its board of directors.

General foresight formative (or process) evaluation questions could include:

- How has the initiative been implemented, and how have its objectives been achieved?
- What have been the outputs (e.g. meetings with other sectors, a foresight process, such as scanning)?
- How well is the foresight initiative being conducted? How competent is the project team? What could be improved?
- How satisfied are initiative participants?
- What evidence is there to demonstrate activities/participant/funder contribution?
- What are the facilitating and limiting factors (internal and external) that have been encountered in the implementation of the initiative?
- What is being learned from the initiative that can inform strategy, tactics and decision-making?
- What is the role of the funder/client in supporting foresight?

31. Van der Steen, M., & van Twist, M. (2012). Beyond use; Evaluating foresight that fits. *Futures*. 44 (475-486.
32. Chelimsky, E. (2001). What Evaluation Could Do to Support Foundations: A Framework with Nine Component Parts. *American Journal of Evaluation.* 22 (1): 13-28.

Similarly, typical summative foresight evaluation questions include:

- To what extent were the short-, intermediate, and long-term outcomes of the initiative achieved? Are these outcomes sustainable?
- What strategies and/or activities were most effective in achieving the desired outcome?
- What role did the activities, participants, and funder play in achieving the outcomes of the initiative?
- What was the impact of the initiative strategy/activities on the target audiences, policymakers?
- To what extent did the initiative result in individual, organizational, population, and/or system change?
- What contributed to (or impeded) initiative success?
- What can be learned from these successes and failures that can inform strategy, such as adoption of the initiative in other settings?

While answers to the formative questions can be obtained in the short-term, many answers to the summative questions, especially those related to impact, only emerge over time.

Figure 3 below is one framework for organizing evaluation questions by foresight purpose adapted from a framework by Ramos et al., (2019) for evaluating participatory futures.[33] The first dimension of the framework concerns the level of impact that might be evaluated from individual to community to institution. The second dimension of the framework is based on the five stages of decision-making or purposes of foresight outlined by Ramos et al which are:

- Mapping horizons: deepening awareness of changes on medium- and long-term time horizons
- Creating purpose: developing a sense of meaning and direction
- Charting pathways: creating high level strategies and socially acceptable pathways for desired change
- Acting together: mobilizing collaborative action and distributed innovation across a community to realize a desired future
- Testing ideas: generating feedback and learning about a specific idea of the future, a scenario, or prototype

33. Ramos J. et al. (2019). *Our futures by the people for the people.* Available from: https://www.nesta.org.uk/report/our-futures-people-people/ Accessed: August 2022

Figure 3: Evaluation Questions by Purpose of Foresight and Level of Impact

Level of impact		Purpose of foresight				
		Mapping horizons	Creating purpose	Charting pathways	Acting together	Testing ideas
Individual		How did foresight deepen individuals' understanding of emerging issues within the context? How did foresight engender a greater sense of responsibility and agency for the future among individuals?	How did foresight shift awareness of the future as a tool to drive action (temporal, optimism, agency) among individuals? How did foresight make stakeholders feel more involved in decision-making?	How did foresight foster individuals' agency in strategizing about the future? How did foresight promote greater ownership among individuals over change processes (e.g. strategic planning)?	How did foresight foster individuals' agency in creating the future? How did foresight generate value and knowledge for individuals? What does this mean in this context? How did foresight help to change individual behaviors?	How did foresight support individuals to explore the impacts of this prototype on their own lives, their jobs or that of their families? How did the foresight experiment help the exploration of individual values and aspirations?
Community		How did foresight support creative exploration of the challenges facing the community? How did foresight help the community identify opportunities for change?	How did foresight create a more positive vision for the community? How did foresight create a more inclusive vision for the community? To what extent did it increase a shared understanding of issues, strengthen shared values, improve social cohesion?	How did foresight help the community to identify the need for change and strategies for change? How did foresight help align community values and aspirations with stated priorities?	How effective were the strategies for initiating change? How did foresight help the community to mobilize the energy and resources of its members coherently for social change and sustainability?	How did foresight help consider the long-term impacts of possible decisions on the community? How did foresight help make the prototype more relatable or desirable for the community? How did the foresight process generate feedback from the community that led to changes or different decisions?

Purpose of foresight

	Mapping horizons	Creating purpose	Charting pathways	Acting together	Testing ideas
Level of impact — Institutional	How did foresight reduce institutional blind spots? What assumptions were challenged? How did foresight help create coherence between institutional silos? How did foresight help the institution identify new opportunities?	How did foresight enhance the institution's purpose? How did foresight generate more diversity/equity/inclusion with respect to the institution's purpose?	How did foresight expand the number of options for creating change that were considered by the institution? In what ways did foresight help align institutional strategies with community needs and desires?	In what ways did foresight make the institution's actions more effective? How did foresight improve trust in the institution?	How did foresight help the institution to identify blind spots and opportunities with respect to planned action? How did participatory foresight help make adoption or rejection of an idea by the institution easier? How did foresight help the institution improve the quality of its engagement with stakeholders?

Also, there is a foundation of foresight-specific evaluation schemas and questions in the peer review literature on foresight evaluation, with some types of foresight being better represented than others, particularly public sector foresight where there has been significant thought given to the role of foresight in how policy is made.[34] For example, Piirainen, Gonzales and Bragge developed a systemic evaluation framework with evaluation questions that are organized by level of analysis (Utility and delivery, Technical, Ethical) and by program maturity (pre-activity, post-activity, and long-term).[35]

Foresight Inputs, Activities, Outputs, Outcomes, and Impacts

While the evaluation questions are the bedrock of an evaluation design, it is highly likely that there will be an assessment of the achievement of impacts and outcomes, which themselves may be evaluation questions. First, evaluators must clarify the meaning of "impact" in their specific situation. In the international development arena, impact evaluation is: "… principally concerned with results of interventions (programs, projects, policy measures, reforms) on the welfare of communities, households, and individuals, including taxpayers and voters. In this context, impact evaluation is one tool within the larger toolkit of monitoring and evaluation (including broad program evaluations, process evaluations, ex ante studies, etc.). In other settings, an impact evaluation must establish the cause of the observed changes, with the OECD definition providing some parameters or: "positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended."[36] We provide some sample evaluation questions using a conventional logic model template below in **Figure 4**.

---

34. van der Steen & van Twist 2012, op. cit.

35. Piirainen, K. A., Gonzales, R. A., & Bragge, J. (2012). A systemic evaluation framework for futures research. *Futures*. 44: 464-474.

36. OECD (2010), *Quality Standards for Development Evaluation*, DAC Guidelines and Reference Series, OECD Publishing, Paris, https://doi.org/10.1787/9789264083905-en.

Designing an Appropriate Evaluation

Figure 4: Examples of Evaluation Questions by Inputs, Outputs, Short, Intermediate, Long-term Outcomes

General	Alternative Scenarios	Scanning	Foresight Trainings and Workshops	Participatory Foresight
Activities: Was the activity/process implemented as planned? Were there refinements? What were the challenges and facilitating factors?				
Outputs: What was produced?	Were scenarios produced or did participants work with scenarios to develop policy recommendations and/or a vision?	How many people were engaged in a scanning system? How was the scanning process embedded in the organization? How many weak signals were generated?	How many 'students' were trained in foresight?	How many media messages on foresight were produced? How many people were engaged in the process? Were relevant sectors represented?
Process: What was the program participants experience? What was the quality of the process?	Were participants satisfied with the scenarios and process? Did the scenarios reach a sufficient level of quality to serve the needs?	Were participants satisfied with the scanning system and its outputs? What was the quality of the signals and their analysis?	Were participants satisfied with the workshop/course? How did they rate the quality of the course process and content?	Were participants satisfied with the process and its outputs? How participatory was the process, e.g., diversity, level of engagement?
Short-term Outcomes: Were there positive changes in awareness, knowledge, and behavior?	Did the scenarios change participant and/or organization: 1) foresight competency; 2) agency to shape the future; 3) identification of a preferred future	Were there positive changes in organization: 1) foresight understanding and competency; 2) resilience; 3) Openness to new sources of information	Did the trainings change participant and/or organization: 1) foresight understanding and competency;[37] 2) agency to shape the future. Were the participants able to implement their learnings in their work?	Did the workshops generate new knowledge, change participants and/or organizations: 1) foresight understanding competency; 2) agency to shape the future;

37. Hines, A. Gary, J., Daheim, C., & van der Lan, Luke. (2017). Building Foresight Capacity: Toward a Foresight Competency Model. *World Futures Review*, 9(3) 123–41.

General	Alternative Scenarios	Scanning	Foresight Trainings and Workshops	Participatory Foresight
Activities: Was the activity/process implemented as planned? Were there refinements? What were the challenges and facilitating factors?				
Intermediate Outcomes: Did the activity/process contribute or result in more use of foresight?	Did the alternative scenarios contribute: 1) informing strategy; 2) informing decision-making; 3) increase systemic understanding	Were the scanning results: 1) used in strategy; 2) used in decision-making;	1) Did the trainings contribute to increased 'futures literacy'[38]2) Did the number of foresight trainings increase?	Did the process contribute to: 1) informing decision-making; 2) increased commitment to long-term goals, future generations
Long-term Outcomes: Did the activity/process achieve a system-level change?	Did decisions informed by alternative scenarios result in better outcomes for organizational or community well-being?	Did the decisions informed by scanning result in prevention or avoidance of a crisis, seizing new opportunities?	Did the trainings increase the number of leaders who engage in long-term thinking?	Did the process result in passage of policies that takes into consideration future generations?

38. Definition of 'futures literacy:' "FL is a capability. It is the skill that allows people to better understand the role of the future in what they see and do. Being futures literate empowers the imagination, enhances our ability to prepare, recover and invent as changes occur." Source: UNESCO. https://www.unesco.org/en/futures-literacy/about

However, a major challenge to evaluating foresight is the difficulty in identifying the changes due to a foresight activity because of the long time it takes for foresight work (such as foresight targeting policy) to manifest its impacts and because foresight work is only one of many inputs leading to political decision making. There has been considerable work done to develop foresight impact schemas and typologies, as well as balancing the knowledge gained from assessing impacts with that gained from evaluation questions on program quality and the stakeholder experience, or what Ian Miles (2012) calls "dynamic foresight evaluation."[39] For example, Van der Steen and Van Twist (2012) developed an impact evaluation framework consisting of five elements and providing suggestions for evaluating the impact of futures studies in the public sphere to help bridge the gap between foresight and policy.[40] Johnston (2012) provides a comprehensive review of eight frameworks to evaluate policy outcomes and the impacts of foresight. It is a useful repository for identifying impacts in diverse policy settings, such as matching impacts to six functions of foresight for policymaking and sorting impacts by type (accountability, justification, and learning). Additionally, he describes three practical guidelines (heuristics) to guide foresight practice and ensure a well-designed evaluation, such as clear and timely communications with the client, creating strong public/private partnerships, and linking foresight outputs into decision-making. Johnston's contribution is the Foresight Impact Schema, which focuses on four types of impact: Awareness raising, Informing, Enabling, and Influencing. For each impact type, he provides two or more outcomes and possible metrics for each outcome, noting that impacts need to be tailored to the type of foresight work, outputs, and anticipated time horizon, e.g., short, intermediate, and long-term outcomes.

Indicators or qualitative and quantitative evidence that address evaluation questions and/or demonstrate achievement of foresight outcomes is a work in progress. Johnston's Foresight Impact Schema includes qualitative and quantitative metrics for the four impact types. For example, for the impact Informing Policy, possible metrics include:

- Reported use of foresight concepts and data;
- Foresight findings regularly used as evidence basis for decision-making;
- New foresight issues being examined by government departments, companies, NGOs;
- Level of investment in and use of horizon scanning; and
- Confidence expressed in foresight concepts and data.[41]

Makarova and Sokolova (2014) provide indicators for financial analysis methods, such as rate of return methods and payback methods.[42] The foresight cases in Georghiou and Keenan's 2006 article provide some suggestions for indicators, such as the engagement of the voluntary sector in foresight activities in the UK Foresight Program.[43] However, the foresight activity, stakeholder information needs, and availability resources will determine in large part which indicators are meaningful and measurable.

---

39. Miles, I. (2012). Dynamic foresight evaluation. *Foresight*, Vol 14. No. 1 pp. 69-81.
40. Van der Steen & Van Twist 2012, op. cit.
41. Johnston, R., (2012). Developing the capacity to assess the impact of foresight. *Foresight*, Vol. 14, No. 1. p.. 56-68.
42. Makarova & Sokolava 2014, op. cit.
43. Georghiou & Keenan 2006, op. cit.

In many evaluations, evaluators work with initiative stakeholders to identify appropriate impacts and can use these frameworks as helpful guides. However, identifying meaningful measures and feasible methods to demonstrate achievement of these impacts and outcomes may confound the best thought-out impacts. The evaluation arena has developed and tested new approaches for situations like this, such as outcome harvesting or a six-step process of working backwards from changes achieved by a program to determine whether the intervention made a difference.[44]

Methods and Sources of Data

Selecting appropriate methods and data collection instruments that address the evaluation questions and assess achievement of outcomes will entail consideration of program maturity and timing of the evaluation. Unfortunately, many evaluations are done after a program is up and running or at the end, a lost opportunity for gathering pre/post data to determine impact. Ideally, an evaluation starts with program planning but barring that, it is most helpful to the program and foresight practitioner if the evaluation commences at the program start. Whether the design is an experimental, semi-experimental or non-experimental approach will also contribute to the selection of methods. For experimental designs, determining attribution will most likely require quantitative tests of significance while non-experimental designs might not. There will also be a determination of whether qualitative and/or quantitative methods will be used, with an understanding that both have their strengths and weaknesses. However, both types of data may be part of one instrument, such as a survey with closed and open-ended questions, and there are guidelines for how they can complement each other. However, a key factor in determining methods will not be so much technical but will be determined by the availability of resources. In many programs, the budget for an evaluation is 10 percent of the overall budget or less, precluding inclusion of many methods. A very likely scenario is a combination of 3 or 4 methods to answer a subset of the evaluation questions.

After these decisions are made, there will still be myriad methods from which to choose, though some data collection approaches are more standard practice than others. As we found in a survey and case studies of advocacy and policy change evaluations, these methods usually include surveys, interviews, observations, document analysis, and one tailored or unique method.[45] That said, evaluation of foresight is not as rich in measurements as other areas, such as educational evaluation and health care services evaluation. While this may compromise evaluation of some dimensions of foresight, this is a temporary situation and social scientists and evaluators are developing new ways to measure impacts. Also, because of differences in foresight processes and programs, having one metric is not always possible and the evaluators will need to adapt or develop a metric for their specific situation. This requires some expertise in instrument design, testing, and analysis, which could be a challenge when resources are scarce.

Useful Evaluation Methods

A few existing evaluation methods and tools can be adapted to foresight. It should be noted that the evaluation arena is very supportive of innovation and out-of-the-box thinking when it comes to constructs and methods. While there is a clear demand for assessing foresight project outcomes and impacts, such as influencing decision-making, foresight practitioners will want to determine how well their project or workshop is doing early on when there is time to refine the program and areas for improvement. Monitoring is a process to periodically collect, analyze, and use information to actively manage performance, maximize positive impacts and minimize the risk of adverse impacts. It is an important part of effective management because it can provide early and ongoing information to help shape implementation in advance of

44. Wilson-Grau, R., & Britt, H. (2013). *Outcome Harvesting.* Developed for the Ford Foundation.
45. Gardner & Brindis 2017, op. cit.

evaluations. This is a resource-rich area, with many monitoring and evaluation how-to guides. It plays a dual role as an evaluation capacity building approach while informing program quality.[46]

Foresight work produces outputs (such as alternative scenarios) that are imaginary and set in the future to inform decision-making and strategic reflection in the present. This makes the determination of attribution challenging as successful foresight outcomes do not relate directly to many outputs. Contribution Analysis is an approach for assessing causal questions and inferring causality in real-life program evaluations. Developed by John Mayne, it offers a step-by-step approach designed to help managers, researchers, and policymakers arrive at conclusions about the contribution their program has made (or is currently making) to particular outcomes.[47] The six-step process of testing a program theory of change and assessing a contribution claim is particularly useful in the case of foresight as it was designed to help in situations where the program is complex and doesn't lend itself to an experimental approach.

Because foresight focuses on change in its many forms, methods that characterize and prioritize change in a particular setting are very relevant. The Most Significant Change (MSC) technique is a form of participatory impact monitoring and participatory evaluation that involves soliciting and analyzing personal accounts of change and deciding which of these accounts is the most significant and why. Learning opportunities in this evaluation approach occur at two levels: (a) in the choices of which types of change are more versus less preferred (e.g., the direction of change), and (b) in the choice of what criteria of value are most important in each context (e.g., the nature of the objectives being pursued by those changes). In addition, the way in which the MSC story selection processes are structured highlights the differences and similarities in views of different stakeholder groups in a foresight process in relation to both levels of learning.[48]

Last but not least, Advocacy and Policy Change (APC) Evaluation is an arena that has been developing its own unique methods since the early 2000s. A topical focus of a growing number of evaluators, advocacy and policy change evaluation examines a broad range of activities that seek to expand advocacy capacity, influence policy-maker support, shape a policy, strengthen democracy and/or bolster civil society. The field has produced many outcomes, methods, and metrics to assess these activities, as well as tailored evaluation toolkits that combine constructs and methods and are intended for advocates. Advocacy and policy change evaluators argue that policy change efforts by their very nature do not lend themselves to more traditional scientific investigation and demonstrating a cause-and-effect relationship between tactics and outcomes is nearly impossible. Additionally, evaluators have had a dialogue on growing this area of practice to incorporate the advocate perspective as well as proposed parameters to add in evaluation design, such as emphasizing contribution over attribution. This area is particularly germane to foresight activities that target decision-makers, civic engagement activities, such as participatory foresight and futures generations foresight, and adds to an existing foundation of foresight impact schemas developed by Georghiou, Keenan, Johnston, and others for government foresight programs.[50]

Define the Role of the Evaluator

It is important to be mindful of how an evaluator relates to the initiative and the multiple roles they can play to facilitate the successful planning, execution, and application of evaluation findings. The ideal approach is

46. Markiewicz, A. (2014). *Core Concepts in Developing Monitoring and Evaluation Frameworks*. Anne Markiewicz and Associates.
47. Mayne, J. (2012). Contribution Analysis: Coming of Age? Evaluation, 18(3), 270-280.
48. Davis, R., & Dart, J. (2005). *Most Significant Change (MSC) Technique: A Guide to its Use.*
49. Gardner & Brindis, 2017 op. cit.
50. Johnston 2012, op. cit.

one whereby the neutral perspective, resources, and strengths of an external evaluation are combined with the insider, ground-level perspective of the participant. One way is to adopt a "critical friend" perspective and establish the external evaluator as an objective insider. Defined as "a trusted person who asks provocative questions, provides data to be examined through another lens, and offers critiques of a person's work as a friend," a critical friend takes the time to fully understand the context of the work presented and the outcomes that the person or group is working towards. The critical friend perspective blends the objectivity that is expected of an evaluator with the trustworthiness and sensitivity that are the attributes of a friend.[51]

The evaluation is at risk if the evaluator does not have the cultural competence to work with diverse stakeholders who may not share the same values and mindset, particularly in how they regard evaluation. Considered a critical skill for the evaluation profession, one of the American Evaluation Association's Guiding Principles for Evaluators is that evaluators should be able to "demonstrate cultural competence and use appropriate evaluation strategies and skills to work with culturally different groups." For example, in a public sector foresight context, evaluators need to factor into the design the power dynamics and identify whose interests are served by the evaluation. They need to be sensitive to the inequalities and injustices in everyday social relationships and arrangements. The evaluation community has identified seven requirements to maximize the cultural competency of an evaluation:

1) considering the community for whom the evaluation plan is created;
2) pretesting survey instruments with different ethnic groups;
3) obtaining information about other attributes related to ethnicity beyond self-identification of ethnic group;
4) building a process check into the evaluation by holding ongoing discourse with the evaluation team concerning their experiences with participants;
5) using triangulation of multiple information sources;
6) including expert cultural or ethnic consultants on the evaluation team; and
7) creating research reports that contain full discussions of the sample and sampling methodology used.[52]

Last, adopting participatory evaluation approaches that include clients and participants will help balance the evaluation.

Ethical Considerations
Regardless of what is evaluated, evaluators are expected to identify and address the ethical issues that might arise during the evaluation. At one level, there are the ethical implications of different designs (such as the use of a Randomized Control Trial (RCT) approach and withholding of the intervention from the control group) and the use of emerging technologies in evaluation practice (such as big data). In the U.S., evaluations in the university settings and public sector typically must undergo a human subjects review and demonstrate adherence to guidelines that respect informant privacy, such as securing participant permission or what is called "informed consent." However, all evaluations have an obligation to prevent harm pertaining to those associated with the evaluation, such as ensuring confidentiality. In addition to disclosing possible risks to evaluation informants and how they will be mitigated, it is expected that an evaluator will behave with honesty and be transparent in missing or weak skills or expertise and disclose any conflicts of interest.

---

51. Costa, A. L., & Kallick, B. (1993). Through the Lens of a Critical Friend. *Educational Leadership,* 51(2): 49-51.
52. Dunaway, K. E., Morrow, J. A., & Porter, B. E. (2012). Development and Validation of the Cultural Competence of Program Evaluators (CCPE) Self-Report Scale. *American Journal of Evaluation*, 33 (4):496-514.

Another ethical consideration is described above in the 2011 AEA Public Statement on Cultural Competence in Evaluation, or: "Effective and ethical use of evaluation requires respecting different worldviews." Evaluators must strive for an ethical commitment to fairness and equitable treatment and respect different worldviews and perspectives.[53]

Effective Communications

Putting some thought to the communications plan during the evaluation planning and implementation stages and creating an information feedback loop will ensure that findings contribute to foresight quality and achievement of impacts. Developing and disseminating evaluation findings early on can create a vehicle for change that continues well after the end of the initiative. With stakeholder buy-in, evaluation publications and other products can be used to fine-tune a foresight initiative, as well as build organizational capacity to build on foresight work. For example, descriptive cases of successful foresight work can educate others, as well as strengthen the case for future support.

---

53. American Evaluation Association. *Evaluators' Ethical Guiding Principles.* 2018.

Tackling Foresight Evaluation Challenges

Considering the future or sets of futures that have not happened yet distinguishes foresight from other types of activities with respect to evaluation. Indeed, a program or policy that has been implemented has usually delivered concrete outcomes that can be identified and evaluated. That said, outputs or images of the future are considered methodological steps toward desired near-term outcomes, such as enhanced capacity for strategic thinking. Arguably, there are programs and policies where the effects won't be perceived for many years to come (such as a policy to slow global warming), and the real aims are changes in decision-maker support for a particular policy. Nonetheless, evaluating foresight requires being creative in design options as well as knowledgeable about compatible areas in evaluation such as developmental evaluation, systems thinking, and complex adaptive systems. Building on the work of the APF Foresight Evaluation Task Force, we detail the challenges evaluators are most likely to encounter and possible strategies for addressing them. (See **Figure 7**)

Evaluation Capacity

Tonn, Gardner, and Lesieur (2024) characterize foresight evaluation capacity based on the findings from a 2021 survey of the APF membership. At the individual level, having evaluation expertise is a challenge for most foresight practitioners.[54] Only about 30 percent of APF members that responded to a survey demonstrated some expertise and were applying it. The good news, as described in **Figure 5**, is that those who evaluate their foresight know how to do process and impact evaluation. For practitioners with limited or no evaluation expertise there are many "How-To" evaluation guides that, while not targeted specifically to foresight activities, cover the core activities that comprise a solid evaluation. For example, evaluator Sheila Robinson's guide, Professional Development Program Evaluation For the Win!, not only provides step-by-step instructions to conducting a robust evaluation but the focus on professional development training is not so different from that of foresight workshops and trainings.

Figure 5: APF Members' Expertise to Conduct Process (N=27), Impact (N = 25) Evaluations
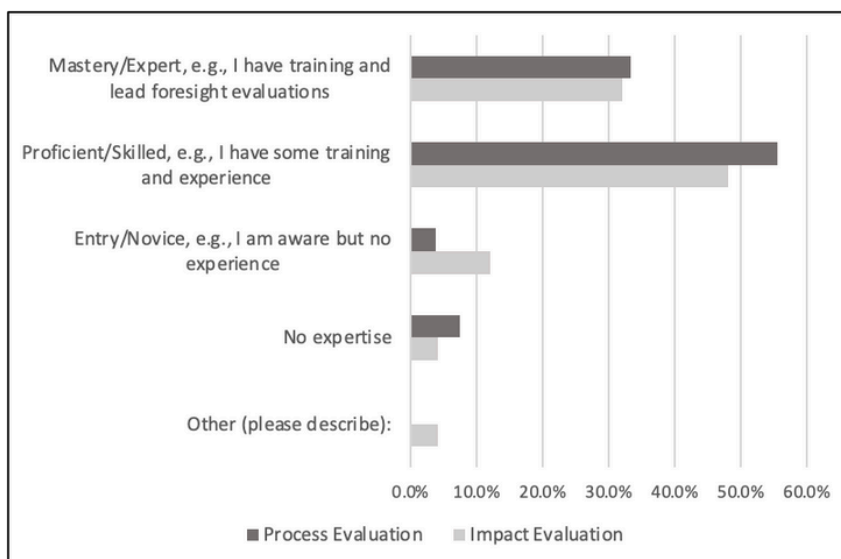
Source: Tonn, B., Gardner, A. L., & Lesieur, B. (2024). Expanding Foresight Practitioner Excellence: Assessment of Practitioner Evaluation Capacity. New Directions for Evaluation, (182).

---

54. Tonn, B., Gardner, A. L., & Lesieur, B. (2024). Expanding Foresight Practitioner Excellence: Assessment of Practitioner Evaluation Capacity. *New Directions for Evaluation*, (182).

However, limitations in resources available to undertake an evaluation are a reality in nearly all evaluations and range from financial to technical. The findings from the 2021 APF foresight evaluation capacity survey described in **Figure 6** speak to the many challenges besides the financial ones (35%), such as limited staff evaluation knowledge (12%), working with data, and communicating (19%), and using and effectively communicating findings (31%). This is not unique to foresight evaluation, requiring individual and field-level action to make the case to foresight commissioners and clients that evaluation is an important component of any foresight activity or program.

Figure 6: APF Member Evaluation Capacity

Resource (35%)

Limited staff time	13%
Insufficient financial resources	10%
Insufficient support from organizational leadership	8%
Insufficient support from organizational staff	5%

Knowledge (12%)

Limited staff evaluation knowledge, skills, and/or tools	7%
Not knowing where / how to get started with evaluation	5%

Process & Data (19%)

Difficulty in specifying variables to measure impacts	7%
Difficulty in collecting the requisite quantitative and qualitative data	5%
Difficulty in developing rigorous designs, e.g., valid, generalizable	5%
Difficulty in analyzing data	3%

Purpose, Use, & Stakeholder Expectations (31%)

Lack of follow-up and use of evaluation findings, e.g., to inform strategy.	10%
Clients are not interested in program achievements and longer-term impacts of activities	10%
Lack of appropriate outcomes, indicators, and methods that fit my organization's work	6%
Managing funder expectations and/or reporting requirements	4%
Difficulty in communicating results	2%

Source: Tonn, B., Gardner, A. L., & Lesieur, B. (2024). Expanding Foresight Practitioner Excellence: Assessment of Practitioner Evaluation Capacity. *New Directions for Evaluation*, (182).

Uncertainty

Commissioners and users of foresight work often want to know whether foresight will reduce uncertainty in decision-making and strategic planning. They also want to identify signals of change that are likely to happen. However, foresight cannot predict the future nor reduce uncertainty. What foresight can do is to explore a wide range of possible futures and to engage participants in "what if" exercises to help them deal with uncertainty and better understand possible consequences at the system level and inform their decision-making to be more effective at creating the future they want. Evaluators can surface assumptions about certainty and identify relevant questions for the evaluation as well as manage stakeholder expectations early in the evaluation.

Complexity

Foresight is ideally suited to help people deal with complexity, but foresight activities are themselves complex initiatives. They engage many stakeholders with diverse perspectives, working with complex information and knowledge from multiple arenas - social, technological, environmental, economic, and policy - often in rich, multi-step processes over many months. They also create detailed, complex outputs, such as second- or third-order consequences of an event through a futures wheel, a long multi-stakeholder course of actions in a scenario exploration, or a set of complex alternative scenario narratives. These outputs are designed to help people sense and manage complexity and understand the non-linearity and non-reproducibility of events. This helps participants gain new perspectives on complex issues and the multiplicity of possible 'solutions' (outcomes) and their trade-offs. The many components people have to deal with can interact in ways that often lead to unexpected outcomes. It is important to understand the features of complexity and get comfortable with nonlinearity. Evaluators of foresight must contend with both types of complexity, at content and at process levels. As described above, *Complex Adaptive Systems (CAS)* models are helpful here, providing frameworks which can inform evaluation design.

However, to be comprehensive, the internal focus of evaluation addressed in the previous paragraph must be complemented by a more contextual approach. This is where new analytical techniques, including *contribution analysis*, *contextual analyses* (such as realist evaluation), and *social network analysis (SNA)* are making it possible to navigate the complexity of the policymaking process as a whole and determine what contributes to achievement of interim and long-term outcomes. All these approaches can be used to connect the attribution dots and provide a more robust understanding of the role of funding, key foresight tactics, and how a policy change came about, or possibly important lessons learned if a change did not occur.

Last, the *wicked problems* framework, or characterizing intractable problems that do not lend themselves to easy resolution allows evaluators and stakeholders to view an initiative in a fundamentally different way. For example, in a public sector foresight context, it assists with understanding the underlying complexity of policy issues and the difficulty in achieving consensus on a particular solution, such as global warming. The framework forces stakeholders to take a hard look at the problem itself, the effectiveness of the intervention, and stakeholder involvement.[55]

Long time horizon and determining impacts

As mentioned several times already, the impact of foresight is often long-term, blurring the traces of its influence. Interestingly, it can also be something that does not happen (i.e., an outcome that is avoided), such as when foresight is used for preparedness and prevention. Indeed, considering an adverse possible, or even probable, scenario can inform the decision-making necessary to avoid this negative outcome.

Attribution vs. contribution

In most cases, the causal chain between a foresight intervention and its impact on decision-making or policymaking is unclear. Theories of change and methods such as the *Most Significant Change (MSC)* technique and *systems thinking* approaches described above can help better determine the attribution of

---

55. Sherman, J., & Peterson, G. (2009). Finding the Win in Wicked Problems: Lessons From Evaluating Public Policy Advocacy. *The Foundation Review*, 1(3): 87-99.

an outcome to an intervention. As this is often not possible, foresight evaluators may need to educate stakeholders on the value of focusing on how and to what extent foresight has contributed to an outcome and on the meaningful information that can be gained. Tools like *systems maps*, *contribution analysis*, and *process tracing* that look at contribution are very useful.

Emergent nature of foresight initiatives

Much of foresight work is by definition "emergence" since it considers possible endpoints that are hypothetical and far into the future. However, these future-oriented outputs are generally a means to a concrete end, for example to enrich understanding, reflection and decision-making in the present or to define shared long-term objectives (visions). The foresight processes can also lead to new dynamics between stakeholders and change the attitude of participants for the long-term. This requires a different evaluation approach that focuses on novel and unanticipated outputs and/or outcomes. Emergence refers to the process whereby novel collective behaviors, properties, patterns, or phenomena come into existence only when the parts of a system interact as a whole.[56] Additionally, the change mechanism adds another dimension of complexity and external factors add an element of unpredictability. At the field level, evaluators, such as the Emergent Learning (EL) Community Project are developing evaluation principles and processes based on a greater understanding of the role of emergence in social change.

In practice, evaluators must wrestle with this complexity without getting flummoxed by it. One strategy is to incorporate a *developmental evaluation* approach in which emergence is assumed. It characterizes complexity and its properties—nonlinearity, emergence, adaptation, coevolution, dynamic interactions, and uncertainty—and provides evaluators with a strategy and tools to anticipate and learn from a program or initiative that is constantly in a state of flux.[57] As part of this strategy, it's important to understand the role of emergence in foresight and identify and document emergent partners, strategies, and outcomes. Likewise, it is important to identify the aspects of a foresight initiative that are not emergent and can be approached using conventional evaluation methods, such as assessing participant satisfaction, changes in knowledge, and fidelity of implementation.

Credibility and Rigor

While determining attribution is a challenge in many foresight activities, evaluations can be designed to increase rigor, including:

- high participation and involving stakeholders with different perspectives to guard against "blind spots";
- collecting baseline data for a pre/post or longitudinal analysis;
- triangulation of data from diverse source--interviews, sites visits, surveys, financial analyses--to corroborate (or refute) findings;
- having a reasonable sample size and response rate in surveys;
- identifying a counterfactual (such as survey with placebo technique); and
- inquiring about unintended consequences.

Also, evaluators should inquire about attribution even if it is to ask what people think would have happened in the absence of the initiative.

---

56. Darling, M. J., Gubber, H. S., & Smith, J. S.(2018). *A Whole Greater than Its Parts: Exploring the Role of Emergence in Complex Social Change.* Fourth Quadrant Partners.
57. Patton, 2001 op. cit.

While an experimental design can limit certain biases, it depends in large part on the nature of the initiative and whether it can be isolated from other influences. In the situation where the intervention cannot be isolated, there are alternatives to random assignment and identifying a control group, such as constructing a comparison group or using a nonexistent foresight group which to compare real foresight groups to rule out alternative explanation or what is called "Survey with Placebo" (SwP).[58] There are several other ways to increase the rigor of non-experimental designs that lack a control group. For some types of foresight activities, such as an assessment of changes in participant "futures consciousness", it matters that a survey instrument has high construct validity and measures what it purports to measure and is completed by a large enough sample of informants to achieve adequate statistical power. It will have more robust findings about impact than a poorly designed and executed survey that includes a control group. For a Delphi Survey where experts are asked to assess the likelihood of an event or activity, statistical validity is less relevant since it is an exploratory foresight approach.

Evaluators are developing approaches for working with small samples when random sampling is not feasible. Non-random sampling, particularly purposive sampling and the selection of informants who can provide an in-depth understanding of the initiative will prove more informative than a potentially skewed random sample of a small population of informants. Identification of cases that speak to the range of scenarios under an initiative protects against overlooking outliers as well as providing a comprehensive understanding of the initiative.[59]

If an experimental or semi-experimental design is not feasible but your evaluation design includes a program theory of change and/or logic model, then you are well positioned to monitor progress and achievement of initiative outputs, outcomes, and impacts as well as test the strength of the links between the model components. It is also the means for explaining why program outcomes were or were not achieved and focusing the evaluation on key aspects of the initiative, strengthening the design and implementation of future initiatives. By the same token, in an uncertain and changing initiative, it is important to be flexible and have a nimble theory of change or logic model that can quickly adapt to emergent aspects of a foresight and policy change initiative, including revisions to the logic model.

Regardless of type of design and the strategy used to increase rigor, it is important to identify and consider rival explanations in any evaluation design. For example, are there other factors, such as leadership buy-in that are not supported by the initiative but that are instrumental in achieving the program outcomes? To what extent are contextual factors, such as a sudden economic downturn, responsible for the outcomes of a local participatory foresight vision workshop? This information can be captured as part of the evaluation contextual analysis which details the environment—events, changes, trends, political factors, etc.—in which the program is taking place. These questions are also excellent evaluation questions that can be asked of informants in interviews and surveys, such as asking about facilitating factors and barriers to program success.

---

58. Gardner & Brindis, 2017 op. cit.
59. Gardner & Brindis, 2017 op. cit.

Craft vs Science

The futures studies discipline and foresight processes and methods are many times characterized as "craft" and rely heavily on tacit knowledge and adaption of methods to different settings. Evaluator knowledge of foresight methods and their characteristics and aims is key. Also, there is no "one-size-fits-all" approach to evaluation design and evaluators should be comfortable with a certain amount of fluidity and messiness.

That said, taking a broader view and intentionally building the futures studies knowledge base through theory development has great value. There is a middle ground or having an evaluation design that balances the two perspectives, such as including grounded theory social science research questions in a process evaluation. One can assess quality and engage in rigorous knowledge creation.

Perceptions of Evaluation

Perhaps the biggest hurdle to evaluating foresight is the failure to see the value-added of foresight evaluation to:

- Foresight practitioners, such as improved foresight practice through critical self-reflection;
- The foresight profession, such as increased credibility through rigorous studies of the positive impacts of foresight on decision-making, and;
- The field of futures studies, such as knowledge building through generalizable evaluation findings on the benefits of positive futures.
- Participants of foresight work, such alternative scenario workshop participants who want to apply the process in their organization.

This could be due to the perception that if evaluation does not document a success, there will be negative consequences, such as terminating the program. This outdated and narrow understanding of evaluation is being replaced by an emphasis on optimizing learning by stakeholders throughout an evaluation process, from planning to communication of findings.[60] A second issue is that evaluation is often perceived to be primarily a backwards-looking activity but that too is changing as the evaluation community learns more about foresight and its usefulness in informing strategy.[61] For example, evaluators can wind tunnel their evaluation findings, increasing their robustness under different scenarios. Like foresight, evaluation is also seeking to inform strategy and foresight methods make this possible.

A commitment to building a culture of evaluation and building an accessible repository that includes tested designs and methods (qualitative and quantitative), rigorous cases, as well as networking opportunities and trainings will do much to dispel these concerns. Also, there is ample evidence that the knowledge base of foresight evaluation is significant and growing. These advances include:

- Three themed journal issues on foresight evaluation models and impacts, as well as published findings from evaluations of alternative scenario projects, national foresight programs, corporate foresight, environmental and horizon scanning systems, and foresight education, trainings, and workshops.

60. Twersky, F., & Lindblom, K. (2012). *Evaluation Principles and Practices: An Internal Working Paper*. The William and Flora Hewlett Foundation.
61. Carden et al. 2023, op. cit.

- A robust body of work on evaluating foresight work in the public sector, at the local and national levels, including development of impact schemas and typologies.
- Foresight evaluation, as standard practice, being incorporated in mainstream foresight guides, such as a section on Evaluating Impact in the 2020 RSA report, <u>A stitch in time? Realizing the value of futures and foresight.</u>[62]
- Foresight organizations developing internal evaluation capacity, such as The Finnish Innovation Fund's (Sitra) comprehensive process, outcomes, and impact evaluation approach.[63]
- Futurist and foresight associations such as the World Futures Studies Federation, the Association of Professional Futurists (APF), and UNESCO are committed to strengthening foresight evaluation.

Also, evaluators are beginning to learn about foresight and its role in supporting strategy, and are adopting a forward-thinking mindset. Increasingly, the evaluation arena is applying systems thinking and complex adaptive system approaches to evaluating transformative initiatives, as well as adding foresight methods, such as scenario planning and the futures wheel to evaluation practice. The bridge between the two disciplines is being built from both sides.

62. Shallowe, A. et al. 2020, op. cit.
63. Link to the Sitra approach: https://www.sitra.fi/app/uploads/2022/04/sitra-evaluation_framework_december_2021-006.pdf

Figure 7: The Challenges of Evaluating Foresight and Potential Solutions

Challenge	Description	Potential solutions
Certainty	Decision makers and "customers" who desire evaluation often want certainty. Did a foresight process work as intended? Will the foresight process reduce uncertainty? Foresight helps users understand and recognize uncertainty better, thereby helping people deal with it better.	Using evaluation to better frame foresight projects that intrinsically acknowledges the inherent uncertainty associated with foresight methods.

Educating decision makers / evaluators / customers that foresight explores uncertainty, so they do not expect too much certainty.

Educating foresight practitioners of the need/desire for certainty among decision makers, evaluations, and customers.

Using theory of change (ToC) as a tool to illustrate the purpose of foresight, its aims and assumptions of what they want to see happen as a result of the foresight. What are the relevant questions for evaluation in relation to the ToC? |
| Timing and salience of impact | The challenge of prediction. The impact of foresight is often long-term and can be something that didn't happen (e.g., the pandemics that were avoided).

Foresight can be used for preparedness and prevention. As a result, foresight success will be achieved by adverse outcomes not happening, something difficult to ascertain. | Understanding that predictiveness is not a desired impact of foresight. Nor is it a prerequisite for evaluation.

Interim / milestone measures that might show progress toward outcomes that might be further in the future.

Proxy measures that can be measured more immediately than final-outcome measures.

Theories of changes that show the potential pathway to impact, including interdependencies of effects and outcomes, and assumptions in relation to the change and context, even if that is not reached any time soon.

Process evaluation to understand whether methods were well implemented even if the outcomes are unknown. Note: process evaluation can be valuable in its own right rather than simply as something to use when outcome evaluation if not possible.

Process evaluation complements outcome and impact evaluation, helping to understand what happens and why and what works in what kind of context. Without process evaluation, evaluation is a so-called "black box" and its results are difficult to use in development or scaling. |

Tackling Foresight Evaluation Challenges

Challenge	Description	Potential solutions
		Active exploration and evaluation of counterfactuals to consider what didn't happen, which is often neglected. Foresight is not for predicting the future but informing present choices/decisions. Thus, evaluation is needed to inform if and how foresight generates its outcomes, e.g. how foresight has informed decision-making or help to increase people's preparedness for the future.
Goals	Different kinds of foresight approaches have different goals, which are not always upon agreed. They are also multiple and can be difficult to define.	Clearer problem / outcome definitions so goals are explicit and understood by all involved in a foresight engagement. Clearer taxonomy of futures methods and the sorts of problems that they can solve. The use of evaluation approaches that aim to identify goals, such as goal-free evaluation and facilitation around goals are a good first step. Developmental evaluation and pre-program planning is another approach that aims to identify intended/unintended, desirable/undesirable, anticipated/unanticipated outcomes
Attribution	Attribution of contribution of foresight work to outcomes and impacts is often difficult and the causal change between intervention, outcome, and impact can be unclear and long.	Theories of change that can illustrate the attribution or contribution of particular individuals or groups within a larger whole. Adapted Bradford Hill for causation criteria to better understand and attribute causality when understanding and attributing the role of foresight in outcomes and longer-term impact. The use of systems thinking methods, e.g. systems maps, to better understand attribution and contribution by multiple actors to an outcome and a longer-term impact. Using evaluation methods that focus on understanding complexity and analyzing contribution instead of attribution; e.g. systems change approaches, contribution analysis, process tracing.

Tackling Foresight Evaluation Challenges

Challenge	Description	Potential solutions
Craft	Foresight methods are a craft with lots of tacit knowledge and methods are mixed or adapted, most often tailored to the specific circumstances of a project.	Clearer taxonomy of futures approaches and methods, their characteristics, and what they can achieve.

Developing a framework to identify and understand archetypes of mixing foresight methods, their characteristics, and their value.

Understanding that evaluation is not something "designed" for only standardized or highly developed models/actions or simple/static contexts and best practices. The object and context of evaluation is rather messy and complex in all relevant/important evaluations. Evaluation could be used as a tool to make tacit explicit, to facilitate learning by analyzing how and why the path has evolved during implementation, what is relevant, what works, what has been learned and what is scalable or transferable into other contexts. |
| Cost | Evaluation can be expensive, particularly at scale, and foresight projects often do not have substantial resources. | Assess the applicability and feasibility of experimental, semi-experimental, and non-experimental or descriptive evaluation approaches. RCT is not ideal (rarely applicable) for complex, uncertain foresight activities and programs.

Evaluation does not always have to be expensive:
1. Prioritize what kind of evaluation knowledge is needed and utilized (and by whom). Is evaluation for internal needs and developmental purposes or more for accountability and external use?
2. Plan from the start, what kind of knowledge could be produced during the project/process for the use of monitoring, evaluation and learning.

Educating funders to require evaluation and build that into funding in an appropriate way that supports foresight in implementation and learning too. Developing evaluation as an investment for getting better/impactful projects in the future. |

Conclusion

While foresight evaluation is an emerging area of evaluation, there are good reasons to support and accelerate its development. Increasing interest in foresight for policymaking around the world requires robust ways to justify its usefulness and cost. This entails developing rigorous studies of the positive impacts of foresight on decision-making to increase its credibility and develop its practice. Although much work remains to increase use of foresight evaluation and develop tailored methods, foresight evaluation can provide useful findings NOW to practitioners, commissioners of foresight, foresight participants, and the field.

There is a solid foundation of foresight evaluation on which to build. However, there are systemic barriers that need to be addressed, such as the slow uptake of evaluation by foresight practitioners and uneven support for evaluation by commissioners of foresight. This paper intends to address these barriers by assisting skilled evaluators to do quality foresight evaluation. To the experienced evaluator who has expertise in designing evaluations for complex, evolving programs, working with qualitative and quantitative methods, and translating evaluation findings into meaningful information for stakeholders, evaluating foresight will be familiar territory. For skilled evaluation practitioners new to foresight, it will take some skill building in futures studies concepts and foresight methods, but the learning will be well worth it. Foresight also has much to offer evaluation, including methods that can support a more robust theory of change and test the resilience of evaluation recommendations under different scenarios.

In closing, foresight evaluation has many benefits to many stakeholders—the future studies discipline, foresight practitioners, commissioners of foresight, and evaluators themselves regardless of their level of evaluation expertise. The time is right to take this area to the next level or increased foresight professional excellence and high quality, effective foresight work.

ASSOCIATION OF
PROFESSIONAL
FUTURISTS