

# Outcome Trajectory Evaluation (OTE): An Approach to Tackle Research-for-Development's Long-Causal-Chain Problem

American Journal of Evaluation  
2023, Vol. 44(3) 335-352

© The Author(s) 2023



Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/10982140221122771

[journals.sagepub.com/home/aje](https://journals.sagepub.com/home/aje)



Boru Douthwaite<sup>1</sup> , Claudio Proietti<sup>2</sup>,  
Vivian Polar<sup>3</sup>, and Graham Thiele<sup>3</sup>

## Abstract

This paper develops a novel approach called Outcome Trajectory Evaluation (OTE) in response to the long-causal-chain problem confronting the evaluation of research for development (R4D) projects. OTE strives to tackle four issues resulting from the common practice of evaluating R4D projects based on theory of change developed at the start. The approach was developed iteratively while conducting four evaluations of policy-related outcomes claimed by the CGIAR, a global R4D organization. The first step is to use a middle-range theory (MRT), based on “grand” social science theory, to help delineate and understand the trajectory that generated the set of outcomes being evaluated. The second step is to then identify project contribution to that trajectory. Other types of theory-driven evaluation are single step: they model how projects achieve outcomes without first considering the overarching causal mechanism—the outcome trajectory—from which the outcomes emerged. The use of an MRT allowed us to accrue learning from one evaluation to the next.

## Keywords

theory of change, theory-driven evaluation, realist evaluation, middle-range theory, outcome trajectory, complex adaptive systems, CGIAR

## Introduction

Policies create enabling or constraining institutional environments for innovation processes. Informing policy is therefore central to international research for development work because policy change is likely to be an integral part of any pathway linking research to impact at scale (Jones, 2011; Renkow, 2018). This is particularly true in recent years in which research for development (R4D) organizations are increasingly under pressure to demonstrate their contribution in

<sup>1</sup> Selkie Consulting Ltd., Cushalogurt, Kilmeena, Westport, Co. Mayo, Ireland

<sup>2</sup> Center for International Cooperation in Agricultural Research for Development (CIRAD), Montpellier, France

<sup>3</sup> CGIAR Research Program on Roots, Tubers and Bananas (RTB), led by CIP, Lima, Perú

## Corresponding Author:

Boru Douthwaite, Selkie Consulting Ltd, Ireland.

Email: [bdouthwaite@gmail.com](mailto:bdouthwaite@gmail.com)

addressing responding to societal challenges. This has led to the broad adoption and use of theories of change (ToC) to make research-initiated causal chains explicit. It has also led to diversified and mixed evaluation methods, and to an increase in the use of theory-driven evaluation approaches that use ToCs (Belcher & Hughes, 2021; Faure et al., 2020; Joly & Matt, 2017; Reed et al., 2021).

Compared to development projects,<sup>1</sup> the evaluation of research for development (R4D) projects is harder because of longer causal chains that link research activity and outputs to its eventual outcomes and impact. With development projects, any research, necessary to develop solutions has typically already taken place. The long-causal-chain problem is particularly acute in evaluating R4D contribution to policy change, because of the particular complexity of public policy making processes (Mueller, 2020).

Common evaluative practice in both development and R4D is to develop a ToC at the start of a project and then evaluate the project against it at the end. Research's long causal links make this problematic in four ways:

1. At the start of a R4D project, uncertainty and non-linearity, both characteristics of the complex systems in which the project will intervene, make it practically impossible to predict the long chain of cause-and-effect relationships that will link research activity to future outcomes and impact. Hence, evaluating a project at the end of its lifespan, against its ToC developed at the start, is unlikely to be very useful.
2. Project ToC by definition places the project at the centre of the causal model, risking too much causal power being attributed to the project.
3. Significant outcomes and impact to which research has contributed nearly always require at least one project cycle, usually two to four project cycles, to emerge. Project ToCs, with their focus on a single project, do a poor job of describing and learning from the longer dynamic of keeping an outcome trajectory going from one project to the next.
4. Project ToCs are based largely on ad hoc stakeholder theory—the usually implicit theories of change held by those close to the program (Breuer, et al. 2016; Donaldson, 2007). They are rarely built on published conceptual frameworks or theories as to how and why behaviour change happens. The lack of use of common conceptual frameworks makes it harder to compare, contrast and learn across evaluations of similar types of projects, and harder for evaluation to contribute to theory testing and building.

This paper describes a novel approach to theory driven evaluation that seeks to provide a partial answer to the four issues. Our approach is called outcome trajectory evaluation (OTE). It was developed in the context of the CGIAR, the World's largest agricultural innovation network, which engages entirely in R4D. CGIAR has an overall strategy and results framework covering the period 2016 to 2030 (CGIAR Consortium Office, 2015) that set ambitious developmental targets for itself and its partners, such as 100 million people, of which 50% are women, assisted to exit poverty by 2030. While aspirational, the targets set high expectations for what constitutes significant CGIAR outcomes and impact.

## **The Cases for Which the TDE Approach “Outcome Trajectory Evaluation” was Developed**

We (the evaluation team) developed Outcome Trajectory Evaluation (OTE) to carry out evaluations of if and how the CGIAR Research Program (CRP) on Roots, Tubers and Bananas (RTB) had contributed to four policy-related outcomes. RTB was one of fifteen CRPs that made up the CGIAR's R4D portfolio prior to 2022.

**Table 1.** Four Policy Change Cases for Which OTE was Developed.

Case Title	Main Outcomes Achieved
1. Development of a cassava seed certification system in Tanzania	Cassava seed standards, from pre-basic to quality declared seed (QDS), have been passed into law to provide the regulatory framework for a functioning and sustainable cassava seed system. The system is being progressively put into place and current advances include strengthening of national and local technical and organizational capacities, and development of an online app to manage the certification process.
2. Development of a cassava seed certification system in Rwanda	The same as Tanzania, but in less time, having benefited from prior experience.
3. Control of potato purple top (PMP—Spanish acronym) in Ecuador	Establishment of a national-level technical committee who have drafted a coordinated national control strategy.
4. Mainstreaming of biofortification in the African Union (AU)	A continental declaration has been drafted by the AU Commission (AUC) that endorses regional- and country-level operationalization of biofortification as a strategic step in accelerating the scale-up and adoption of biofortified crops and products.

The four cases (see Table 1) were selected by RTB staff based on the fact that policy change had taken place, that the changes were judged to be significant, sources of information were available and there was interest to document and analyze the cases.

The objectives of each of the four policy outcome evaluations, as stipulated by RTB who commissioned the work, were:

1. To determine and document how and in what ways CGIAR R4D contributed to the cases;
2. To identify other major contributing factors, actions and actors;
3. To contribute to a cross-case synthesis of findings and lessons learned.

The evaluation findings were also expected to help inform the CGIAR's reformulation under a new phase called One CGIAR.<sup>2</sup>

The objectives shaped the design of OTE. For example, the “how?” question in the first objective led to the choice to use TDE in the first place.

## The Approach

OTE is based on four starting assumptions. The first, which is embedded in the name, is the idea that project outcomes are not single, one-off events; rather, they are generated over time by an interacting and co-evolving system of actors, knowledge, technology and institutions (Axelrod & Cohen, 1999). This system is called an outcome trajectory (Paz-Ybarnegaray & Douthwaite, 2017). Outcome trajectories are different in their starting point from impact pathways which are what project ToCs, as defined by Mayne and Johnson (2015), are commonly called in the CGIAR. Outcome trajectories are constructed by working backwards from existing outcomes while project ToCs generally begin with project activities and outputs to describe how they are expected to produce outcomes and impact in the future. Project ToCs are by definition project-centric and look to the future while outcome trajectories are built by looking backwards. Outcome trajectories have the same starting point as outcome harvesting (Wilson-Grau, 2018)—an achieved and verifiable outcome (see Table 4 for OTE's similarities with other approaches to evaluating policy outcomes). OTE takes a retrospective view, called “back casting” in the outcome harvesting literature.

The second assumption is that many of the long-causal-chain outcomes to which research contributes, are policy oriented. OTE was designed to evaluate policies. According to Renkow (2018, p. 2), agricultural R4D contributes to five types of policy-oriented outcomes:

- Changes in laws and regulations governing economic incentives in agriculture or natural resource management;
- Creation of institutions;
- Changes in government investment priorities and budget allocations;
- Innovations to the operations and management for government agencies and programs;
- International treaties, declarations, or agreements among parties reached at major policy conferences.

The third concept is that a number of theories exist to explain how research and other types of activity bring about policy-oriented change. Sabatier (2007) edited a seminal book called “Theories of the Policy Process” that identified six that met the following criteria:

1. The theory/framework must do a reasonable job of meeting the criteria of a scientific theory; that is, its concepts and propositions must be relatively clear and internally consistent, it must identify clear causal drivers, it must give rise to falsifiable hypotheses and must apply to most of the policy process in a variety of political systems;
2. Each framework must be the subject of a fair amount of recent conceptual development and/or empirical testing;
3. Each framework must be a positive theory seeking to explain much of the policy process;
4. Each framework must address the broad sets of factors that political scientists have traditionally deemed important: conflicting values and interests; information flows; institutional arrangements and variation in socioeconomic environment.

The fourth key concept is Merton’s (1968) idea of middle-range theories (MRTs) that inhabit a middle ground between ad hoc explanations of singular cases (e.g., stakeholder-developed project ToC) and “grand,” universal systems theories that explain all features in a stylized way (e.g., Sabatier’s six theories) (Hedström & Ylikoski, 2010; Pawson, 2017). MRTs are useful in providing a “reusable conceptual platform” identified by realist evaluators as helpful in accruing learning from one evaluation to the next of a similar set of well-bounded phenomena (Meyfroidt, 2016; Pawson, 2013). Critical to our approach is the idea that Sabatier’s theories can be used to build MRTs which can then serve as reusable conceptual platforms to accumulate learning across evaluations of similar types of outcomes, in our case, policy-oriented outcomes in tropical agriculture.

Bringing the four concepts together, the approach we developed consists of eight steps:

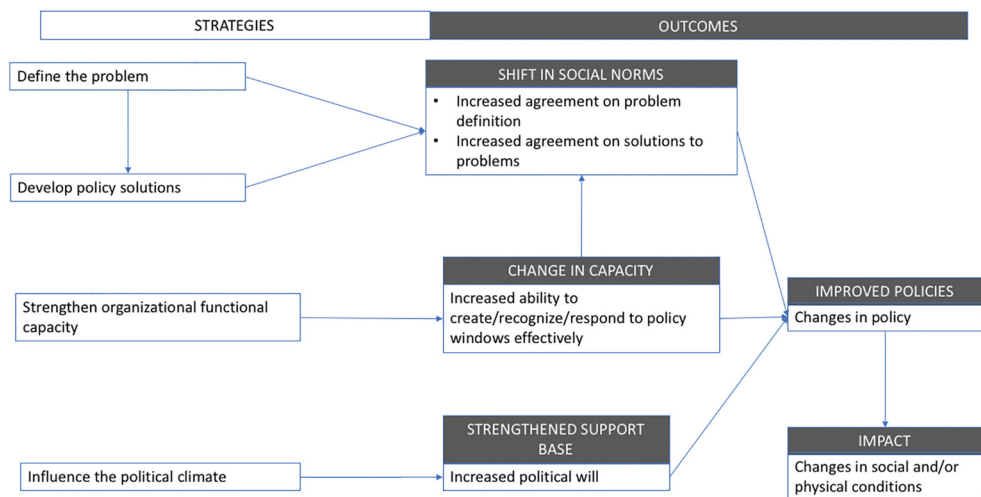
1. Select or develop an MRT to act as both a reusable conceptual framework and the “theory of the case” (Rule & John, 2015).
2. Use the MRT as the “theory of the case” to help identify and describe the outcome trajectory for the first case. Describe the outcome trajectory using a timeline of the main events and processes that led to and sustained the policy outcome.
3. At the same time, adapt and specify the MRT based on what is learned.
4. Validate this adapted MRT, and the outcome trajectory timeline, in a workshop with key stakeholders. The first iteration of the MRT can be understood as a ToC of the outcome trajectory.
5. Use the validated MRT and timeline to answer the evaluation questions in a draft report, along with in-depth interviews and the results of a document review.

6. Subject the draft report to fact and inference checking by key stakeholders.
7. Repeat steps 2 to 6, starting each time with the previous iteration of the MRT, updating it after each iteration.
8. Write a synthesis paper taking all cases / OTs into account to accumulate learning and generate insights.

We conducted the four case evaluations sequentially, beginning with the case study where participants had worked more explicitly and for longer on influencing policy (i.e., mainstreaming of bio-fortification in the African Union). The four evaluations were reviewed by participants and published (Douthwaite, 2020a, 2020b, 2020c & 2020d).

We chose Policy Window theory from Sabatier's list of six on which to base our MRT. We did so together with the researchers that had been most involved in the first case, taking into account that Sabatier (2007) found Policy Window theory to be widely applicable of the six. Also, choosing one theory as a basis for the MRT did not exclude adding in elements from other policy process theories that might provide a better or more detailed model of the phenomena we subsequently observed in the four cases. Indeed, we found that the concept of a coalition, from the Advocacy Coalition Framework (Stachowiak, 2013), proved particularly useful in two cassava seed system OTs. From a systems point of view, all models are wrong in some respects, but some are more useful than others in generating insights (Stermann, 2002). Our intention was to make the MRT more specific and more applicable to each of the four cases, and then develop a synthesis MRT applicable to all four (see Figure 3 & Table 3). In this way, our intention was to accrue learning from one evaluation to the next to make it available for future evaluations of similar phenomena.

The Policy Window theory, shown in Figure 1, comes from political science and was developed by Kingdon and Stano (1984). It is also known as the multiple-streams theory. The model proposes that policy changes during *windows of opportunity*, which help champions successfully connect two or more components of the policy process. The components are: the way a *problem* is



**Figure 1.** The first iteration of the middle-range theory, based on policy window theory, used to evaluate how CGIAR interventions in four cases contributed to policy change (reproduced with permission from the author of pathways for change: 10 theories to inform advocacy and policy change efforts. Source: Washington: Center for Evaluation Innovation. (Stachowiak, 2013).

defined; the *policy solution* to the problem; and, the *politics* surrounding the issue (Sabatier & Weible, 2007; Stachowiak, 2013). Windows of opportunity are moments when progress can be made. They can be created by natural events such as pandemics, droughts or earthquakes. For example, the latter is an opportunity to change building regulations. They can be human-made events like spikes in air pollution that lead to changes in clean air regulations. They can also be changes in government, budget cycles or landmark meetings and summits held as part of ongoing national, regional and global processes. Policy windows are often short in duration and may or may not be predictable.

Policy Window theory has been criticized for overly focusing on problem identification at the start of the policy cycle and ignoring the complex steps that follow (Jann & Wegrich, 2007). Our view is that the criticism does not apply to our approach because we work with the premise that policy changes happen as a result of outcome trajectories, that is, through the prolonged actions and interactions of stakeholders together with institutions and technology.

Policy Window theory has also been criticized for not providing enough falsifiable hypotheses (Sabatier, 2007). We also discount this criticism in our work because the Policy Window MRT that we used in the evaluations became increasingly more specific and hence more testable.

## Results: How OTE Worked in Practice

In this section we describe the OTE steps and how they worked in practice in evaluating the four policy outcome cases. We use one of the cases—the development of a clean cassava seed system in Tanzania – to provide explanatory examples where needed.

### *Step 1: Select the “Grand” Theory from Which to Develop an MRT, or a Suitable MRT if one Already Exists*

The reasons we chose Policy Window theory as the “grand” theory from which to develop an MRT, are described above. We also chose Policy Window theory because Stachowiak’s (2013) depiction of it in a causal diagram, and qualifying statements, had begun the process of developing an MRT (Figure 1).

After completing the first case, we used the more specified MRT developed in the process (step 5) as the MRT to begin the second case. We made additions to the MRT after evaluating each case by identifying and mapping the strategies used by trajectory actors onto the immediate outcomes the strategies sought to influence. Key participants in each case had the opportunity to validate and improve the degree to which their version of the MRT modelled what had happened in their case (step 4).

### *Step 2: Identify and Describe the Outcome Trajectory That Produced the Policy Outcome*

We developed timelines for the four respective outcome trajectories, based on initial interviews and review of available reports and online publications. Our approach was essentially a case study one in which understanding of underlying mechanisms at work came from rich, thick picture descriptions of events and processes gleaned from interviews in particular. We used the most recent iteration of the MRT as the “theory of the case,” to focus our inquiries so as to avoid the risk of becoming bogged down in too much spurious detail, which is a known issue with case study methodology (Baxter & Jack, 2008).

In all four cases, it was useful to keep in mind that the outcome trajectory of interest was itself nested within broader outcome trajectories. For example, understanding the outcome trajectory, from which a draft declaration by the African Union supporting biofortification emerged, required

us to also understand the broader outcome trajectory that led to the development and introduction of biofortified crops into Africa in the first place (Douthwaite, 2020a).

### Step 3: Use the MRT to Help Identify and Explain how Strategies Used by Trajectory Actors Contributed to Policy Outcomes

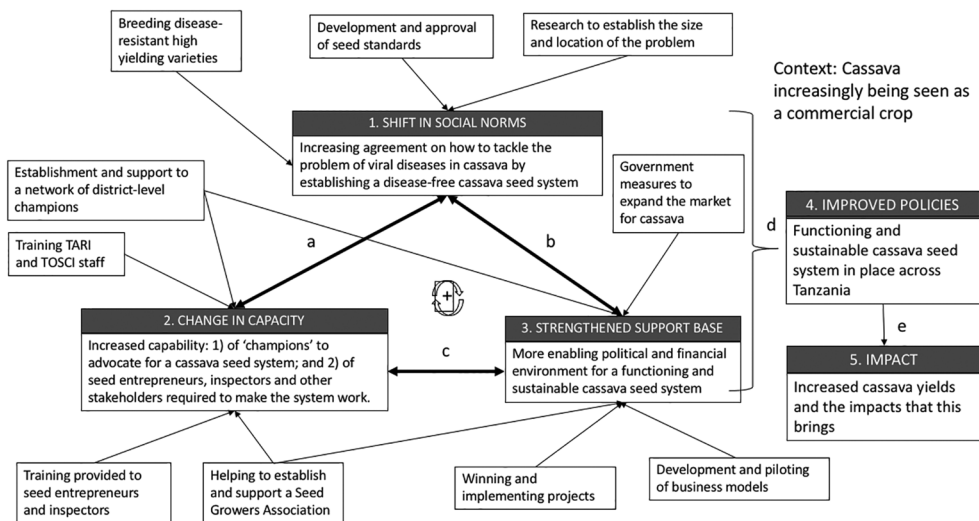
We started the evaluation of the first case using Stachowiak’s (2013) graphical depiction of Policy Windows theory (Figure 1) as the theory of the case. Specifically, we looked for the strategies OT actors had employed to contribute to the MRT’s three immediate outcomes, namely:

- Shift in social norms
- Change in capacity
- Strengthened support base

In each case, we carried out in-depth interviews, with the people who had been most involved in the respective outcome trajectory, to identify how the immediate outcomes were manifest in the respective outcome trajectories. For example, in the Tanzania case we found that “shift in social norms” outcome meant “increasing agreement on how to tackle the problem of viral diseases in cassava by establishing a disease-free cassava seed system” (see Figure 2).

We double checked what we were told about events and processes in interviews by cross referencing against relevant documents we had been provided with as well as those we could find on the Web. The latter was surprisingly useful, often throwing up other connected actors, events and processes to add to the timelines and the narratives explaining them. As we attempted to explain how trajectory actors had contributed to the immediate outcomes, the strategies they had used became clearer.

We used the findings from the analysis in each case to add specificity to our working MRT, to be used in the next case. For example, Figure 2 shows the version of the MRT we used in the Rwanda



**Figure 2.** The third iteration of the MRT describing how seed standards were developed and were being used in Tanzania, which formed the basis for the evaluation of the fourth case on how seed standards were developed in Rwanda.

case, incorporating learning from the Tanzania case. The full MRT includes a written description of the strategies and dynamics involved (see Douthwaite, 2020c, p. 23-24).

The main structural modification to the proto-MRT that we started with in the first case (Figure 1) was to show the three immediate outcomes interconnected at the center of the model, each linked to the others by self-reinforcing feedback loops (see Figure 2).

#### ***Step 4: Validate the Outcome Trajectory Timelines and the Specified ToC with key Stakeholders***

The evaluation team organized a workshop attended by key stakeholders to validate the timelines and the adapted MRT for their case. The workshops were held virtually because of Covid-19 travel restrictions in effect at that time. Nine participants took part in the Tanzania case from five of the organizations most involved in the cassava seed systems outcome trajectory. They were asked to review and add to the timeline and MRT using a virtual whiteboard platform. Their feedback was used to revise both.

#### ***Step 5: Use the Validated Timelines and ToC to Answer the Evaluation Questions***

We then used the validated timeline and specified MRT, together with notes from the in-depth interviews and information from document review, to answer the evaluation questions, which were also adapted to each case. For the Tanzania case, the questions were:

- What are the main outcomes resulting from the Tanzania cassava seed certification trajectory and how did the CGIAR contribute to them?
- Has the CGIAR contributed to integration/consideration of gender in the Tanzania trajectory and how?
- Is the seed certification trajectory likely to be sustained and scaled over the long term?

For the Tanzania case, the timeline allowed us to identify 13 cassava seed system-related outcomes from when the trajectory began in 2012 (Douthwaite, 2020b) such as the gazetting by the Government of Tanzania of Seeds (Amendment) Regulations for cassava, sweetpotato and potato in 2017. The time-ordering of events, the in-depth interviews and document review allowed us to identify if and how the CGIAR had contributed to them through various strategies, see Table 2.

We used an analogy with criminal justice to explain our approach in the validation workshops, which participants found clarifying. We said that we had used the timeline, interviews and document review to clarify the nature and extent of the outcomes achieved (the “crime”) and to identify who were the major “perpetrators” and what they had done. In each case we were able to demonstrate:

- CGIAR presence, prior intent and conspiracy to act together with other trajectory actors to achieve the outcomes;
- That CGIAR entities had the means (i.e., funding), the capacity (e.g., a trained network of champions) and motivation to champion the outcome trajectory.

The MRT was useful in answering the evaluation question about the future by identifying the strategies and interplay driving the outcome trajectory forward. In the Tanzania case, this analysis allowed us to conclude that future success will depend on continuing to simultaneously build the market for cassava, the availability of clean planting material and farmers’ willingness to pay for it (Douthwaite, 2020b p. x).

With respect to the evaluation question on relating to gender, the timeline helped us clarify that little explicit attention had been paid to gender in the first part of the Tanzania trajectory. For



**Table 2.** Strategies Used to Bring About Immediate Outcomes in the Tanzania Case, with Examples from the Respective Timeline (Douthwaite, 2020b p. 24, reproduced with permission from CIP as copyright holder).

Strategy	Examples of Resulting Outcomes
<b>1. SHIFT IN SOCIAL NORMS</b> —manifest as increasing agreement on how to tackle the problem of viral diseases in cassava by establishing a disease-free cassava seed system	
Research to establish the size and location of the problem	Increase in research papers on cassava viral diseases from 75 in 1990 to a cumulative total of 250 in 2010 including research on the causes, location and nature of the two diseases
Development and approval of seed standards	Development and gazettement of a cassava seed certification protocol and regulations to support it
Breeding disease-resistant high yielding varieties	IITA has a long history of breeding cassava in Tanzania, beginning in 1970s and including work under the Cassava Varieties and Clean Seed to Combat CBSD and CMD (5CP) Project
<b>2. CHANGE IN CAPACITY</b> —manifest as increased knowledge: 1) of “champions” to advocate for a cassava seed system; and 2) of seed entrepreneurs, inspectors and other stakeholders required to make the system work.	
Establishment and support to a network of district-level champions	BEST Cassava project trained district-level champions to advocate for districts to use funds allocated to them to support the production, processing and commercialization of cassava
Training TARI and TOSCI staff	SCP project trained Tanzania Agricultural Research Institute (TARI) staff in clean basic seed production and Tanzania Official Seed Certification Institute (TOCSI) staff in seed certification scheme implementation including upgrading lab skills
Training provided to seed entrepreneurs and inspectors	MMB and BEST Cassava provided training to more than 400 seed entrepreneurs
<b>3. STRENGTHENED SUPPORT BASE</b> —manifest as a more enabling political and financial environment for a functioning and sustainable cassava seed system	
Winning and implementing projects	Proposing and winning 5CP, MMB and Best Cassava projects involving IITA/RTB, MEDA, TARI and TOSCI to develop and implement a sustainable cassava system
Helping to establish and support a Seed Growers Association	MMB and BEST Cassava projects have set up and supported the Cassava Seed Growers’ Association to help seed entrepreneurs to coordinate the testing of their fields by TOSCI and to help access credit
Development and piloting of business models	MMB and BEST Cassava projects developed and piloted seed entrepreneur business model with 400 individuals
Government measures to expand the market for cassava, e.g., deal with China	Government strategy to expand the market for cassava independent of the seed certification trajectory. This includes establishment of cassava flour processing plant and striking a deal with China to buy large amounts of cassava

example, gender was not mentioned in the gazetted cassava seed standards in Tanzania. It may be that considering gender would have made little difference at the level of cassava seed standard policies and regulations, but this question does not seem to have been asked. This appears to be changing however, with MEDA, one of the partners in the Tanzania trajectory, starting a new project that will carry out a gender assessment of cassava seed systems.

### *Step 6: Subject the Draft Report to Review for Fact and Inference Checking*

A draft of the final case report was sent to the workshop participants and the key informants to check facts and inferences drawn by the evaluation team. The comments were collated and were

considered by the evaluation team, who made changes where they deemed them to be credible based on evidence available. The changes made were recorded, explained and sent back. For those who had also attended the stakeholder verification workshop, this was the second time they had to query the approach and the findings. There were no major disagreements by the time the report was published.

### ***Step 7: Repeat Steps 2 to 6, Starting Each Time with the Previous Iteration of the MRT, Broadening its Applicability After Each Iteration***

Step 7 is required when working on more than outcome trajectory as part of an overall evaluation, or on a set of evaluations into a particular phenomenon. If OTE is used for a single OT then the last step is to identify lessons learned and/or conclusions and/or recommendations.

### ***Step 8: Accumulate Learning and Generate Insights as Part of Writing up a Synthesis Paper***

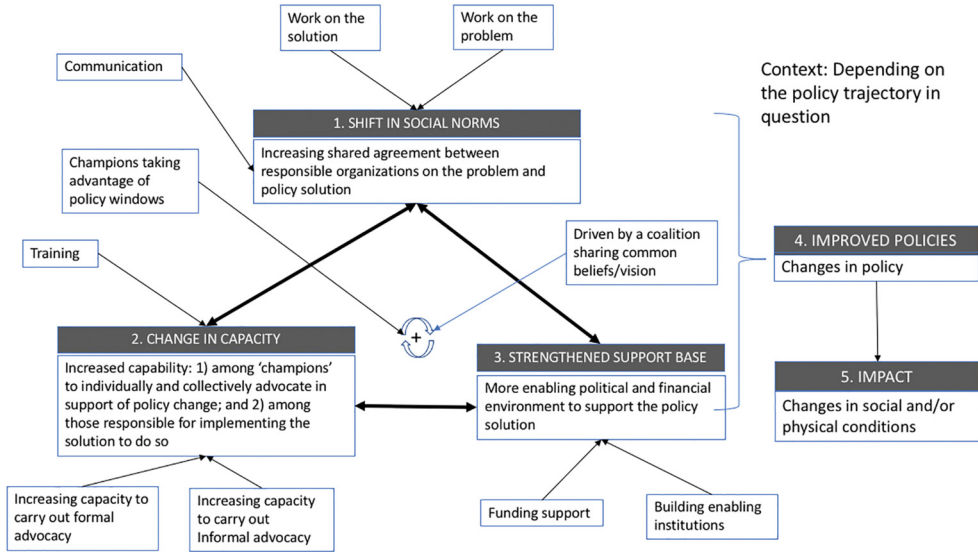
After we carried out the evaluations of the four cases, we wrote a synthesis paper (Douthwaite, 2020e) to accumulate learning and develop more generalizable findings framed by five questions:

1. How can the Policy Window theory be adapted to model four policy outcome trajectories?
2. What are the characteristics of the four policy outcome trajectories, what has been achieved so far, and what is the potential for impact?
3. What are the main policy outcomes resulting from the four cases and how did CGIAR contribute to them?
4. How has CGIAR contributed to the integration/consideration of gender in the four cases and how?
5. Are the four policy change trajectories likely to be sustained and scale over the long term?

The first synthesis question on the theory was answered by listing out the strategies used in each case. This allowed us to generate a more generic MRT (Figure 3) that applies to the types of policy change covered by all four cases, specifically agricultural producer- and consumer-oriented policy outcomes (FAO, 2015).

We indicated which cases used which strategies and briefly described them, as shown in Table 3 for the “shift in social norms” immediate outcome. We developed similar tables for the other two outcomes.

A complete set of findings from the four cases can be found in the respective evaluation reports (Douthwaite, 2020a, 2020b, 2020c & 2020d) and the synthesis paper (Douthwaite, 2020e). For example, the synthesis report found that the way that the CGIAR contributed to the four policy-related outcomes was by being a major participant and contributor to the four outcome trajectories from which the respective case outcomes emerged. CGIAR Research Programs engaged in a number of strategies (Figure 3 & Table 3) that contributed to three immediate outcomes—shift in social norms, change in capacity and strengthened support base. These outcomes interacted and reinforced each other to establish the dynamic that drove the outcome trajectories, helped by champions taking advantage of policy windows and coalitions with shared intent (Figure 3).



**Figure 3.** The final iteration of the MRT developed during the synthesis process, that has the potential to inform planning and evaluations of similar policy interventions.

## Discussion: Issues Addressed by OTE, its Relationship to Other Evaluation Approaches and its Limitations

### Issues Addressed

Our purpose in writing this paper is to share an approach that seeks to provide a partial answer to four issues facing evaluation of R4D projects, resulting from generally long causal chains linking research activity to developmental outcomes, especially true for the influence of research on policy-related outcomes.

*1. Using outdated ToC developed at the start of an R4D project in summative evaluations.* OTE provides the opportunity to evaluate projects against significant outcomes they claim to have achieved rather than what they thought they would achieve at the start. It is standard practice to evaluate projects at the end of their lifespan against a ToC developed at the beginning, when potential causal links are very many and largely unknown, making it likely that any forward-looking ToC will prove to be either too general to be useful, or based on too many unverified hypotheses to be credible, or a bit of both.

This was identified as an issue for the CGIAR in 2017 when an evaluation of results-based management in the CGIAR highlighted the critical importance of working with sound ToC that is updated over the years, rather than working with it as a fixed point of reference (CGIAR-IEA, 2017, p. 26). The evaluation found some ToCs were overly-simplistic and unchanged.

In the four cases, OTE was used to evaluate CGIAR contribution to significant, achieved policy-related outcomes stakeholders chose themselves. By focusing on successes, we took a portfolio approach to evaluation in which the overall rate of return on R4D investment is assumed to depend largely upon a few known project successes. A portfolio approach reduces the number of outcome pathways that a summative evaluation needs to look at, making it potentially a more parsimonious and simpler way of reaching a sense of a program's value. By focusing on significant,

**Table 3.** Strategies Identified in the Four Cases That Brought About a “Shift in Social Norms.” Together with Strategies Identified to Bring About “Capacity Development” and “Strengthened Support Base,” it Makes up Part of the Synthesis MRT Developed for the Four Cases.

Type of Strategy	Specific Strategies Used	Cases That Used It
Framing the problem	Research to identify vectors and causal agents of Purple Top	Purple Top
	Research to document impact of Purple Top	Purple Top
	Bringing in outside experts to help understand Purple Top	Purple Top
Framing the solution	‘Gold standard’ research showing biofortification can significantly reduce the main types of micronutrient malnutrition	Biofortification
	Demonstrating that biofortified crops can be grown at scale in Africa	Biofortification
	Work to establish biofortification as a solution that can be defined and measured in National Agricultural Implementation Plans (NAIPs)	Biofortification
	Formation of a technical committee to develop seed standards	Tanzania and Rwanda
	Development and piloting of business models	Tanzania and Rwanda
	Formation of platforms, committees and groups to tackle Purple Top	Purple Top
	Development of effective Purple Top control measures	Purple Top
Communicating the problem or solution	Maintaining a clear and consistent message with regard to the relevance of biofortification in reducing nutritional deficiencies	Biofortification
	Framing of solution as complementary to other ways of reducing micronutrient malnutrition	Biofortification
	Issuing press releases	Biofortification, Tanzania and Rwanda
	Communication of the problem and/or solution in academic conferences	Purple Top

achieved outcomes, OTE also lends itself to formative evaluations that explore what is working, to inform design and implementation of the next phase. The latter applied to the purpose of this evaluation, which was to synthesize findings and lessons learned across the four cases to inform a change process in the CGIAR. The former also applied to this evaluation as the clients wanted to demonstrate to their respective donors that their support had led to some significant successes in line with the generic ToC developed at the beginning of the program; in other words, that their money had been well spent.

**2. Assigning too much causal power to projects.** OTE broadens the scope of project evaluation to avoid this risk. We did so in this evaluation by using an MRT to help delineate and understand the trajectories that generated the policy outcomes we evaluated, and then identifying project contribution to those trajectories. Other TDEs are single step: they model how the project achieved outcomes without first considering the overarching causal mechanism—the outcome trajectory—from which the outcomes emerged.

**3. Missing contributions from the past.** The cases showed that by starting with a significant outcome, and describing the OT that produced it, causal analysis automatically began at the OT’s origins,

however far back that was. For example, the Tanzania OT timeline (Douthwaite, 2020b) showed that it began in the late 1990s with a USAID-funded project that developed a quality management protocol for producing and distributing clean cassava seed. This idea was picked up in 2007 with the establishment of the Great Lakes Cassava Initiative (GLCI) that ran until 2012, funded by the Bill and Melinda Gates Foundation (BMGF), which distributed clean cassava seed to 1.15 million farmers in Central and East Africa, including Tanzania. FAO and AGRA funded similar initiatives that also worked in Tanzania during the same period. Work began in earnest in Tanzania in 2012 with the launch of three complimentary projects, building on the GLCI and also funded by BMGF. One of these projects—the so-called 5CP project—had a component on building a cassava seed system model for Tanzania. Another, the MMB project, piloted commercial approaches to providing farmers with clean seed. An evaluation of either of these projects individually, without the MRT, would likely have underplayed or ignored the changes that have already happened in terms of shaping social norms, building capacity and strengthening support base, or attributed these changes to the projects.

**4. Lack of learning from one evaluation to the next.** OTE is based on MRTs to which stakeholder theory can be added, rather than ToCs that are generally built entirely on *ad hoc* stakeholder theory. An MRT provides a reusable conceptual framework that allows learning and insight to build from one evaluation to the next, to potentially be used in program design, implementation and evaluation. This is evident in our evaluation in our development of the Policy Window MRT from case to case resulting at the end in an MRT that applied to all four cases.

Our iterative use of a Policy Window MRT helped us to generate a number of potentially generalizable findings. Firstly, the MRT's focus on advocacy generated the insight that effective advocacy strategies depend upon the proximity in network terms between researchers on one hand, and the policy decision-makers on the other. When the degree of separation is large then it was necessary to engage in "formal" advocacy practices such as explicitly recruiting and training policy champions to bridge the gap. When the degree of separation is small, "informal" advocacy can take place in which a coalition of CGIAR and national-level researchers are able to directly engage with and influence key decision-makers. This is important, because informal advocacy is not widely recognized as a scaling mechanism in the CGIAR. Coalitions, such as those that were evident in the three plant disease cases, are generally not recognized or valued.

The MRT's focus on policy windows helped us determine that two types of policy window in particular helped drive the respective trajectories forward. Regional- and global-level conferences provided opportunities for biofortification champions to link biofortification to the broader and well-supported global nutrition trajectory. Disease outbreaks were the most important policy windows for the three disease-related trajectories.

Our focus on learning from one case to the next helped us find that by learning from colleagues in Tanzania, Rwanda trajectory actors were able to develop and approve cassava seed standards in one year rather than the five years it took in Tanzania (Douthwaite, 2020b).

Our final iteration of the Policy Window MRT, developed during the synthesis process, has the potential to inform planning and evaluations of similar policy interventions, specifically of any future evaluation of agricultural producer- or consumer-oriented<sup>3</sup> policy change initiative. Any future evaluation using this this MRT would add a further iteration, potentially making the MRT more broadly applicable.

We found the synthesis Policy Window MRT (see Figure 3 and Table 3) particularly helpful in providing a checklist of different strategies used to bring about the immediate Policy Window outcomes—shift in social norms, change in capacity and strengthened support base. We adapted the MRT to make it specific to the four cases, thus adding to learning as to what R4D strategies work

**Table 4.** Similarities Between OTE and Other Approaches Used to Evaluate Policy Outcomes.

Evaluation Approach	Similarities to OTE
Process tracing (Collier, 2011)	<ul style="list-style-type: none"> <li>- Focus on unfolding events or situations over time to make causal inferences;</li> <li>- The idea that causal inferences can be affirmed through building up a weight of evidence the robustness of which may be established through various tests (e.g., straw in the wind, smoking guns, etc.).</li> <li>- Use of criminal justice system analogies in explaining how the approach works</li> </ul>
Outcome harvesting (Wilson-Grau, 2018)	<ul style="list-style-type: none"> <li>- The practice of “back-casting” from an established outcome to understand what has contributed to it</li> <li>- Interviews with knowledgeable stakeholders to validate or repudiate causal claims</li> </ul>
Contribution analysis (Mayne, 2012)	<ul style="list-style-type: none"> <li>- Use of a contribution story, similar to the timeline used in OTE</li> <li>- The development and refinement of a ToC as a part of the analysis</li> </ul>
Episode study (Carden, 2009)	<ul style="list-style-type: none"> <li>- Back-casting from a well-defined policy change</li> <li>- Development of a historical narrative to explain the policy change along with important documents and events, and identifying key actors</li> </ul>

to produce immediate Policy Window outcomes in which context for which type of intermediate policy-related outcome.

We have subsequently used the synthesis Policy Window MRT in the evaluation of policy-related outcomes relating to the establishment of biofortified crop breeding systems in Bangladesh, India and Rwanda.

Overall, we have found it better to conduct an evaluation with the framing and accrued learning that an MRT provides, than without one.

### *The Similarities and Differences of OTE to Other Approaches Used to Evaluate Policy Outcomes*

Table 4 shows that OTE uses elements employed by four outcome evaluation approaches. It shares the importance of developing a causal narrative with three, and that the narrative should be historical with two. OTE shares the idea of back-casting with two. What sets OTE apart from all four is the explicit use of an MRT, based on a “grand” social science theory, as a framework to guide the identification and description of an outcome trajectory, manifest in an annotated historical timeline. Contribution analysis is the only other approach to use theory, but is different in developing a ToC of the program being evaluated, not the specification of an existing MRT. OTE is different in assuming, *a priori*, that the significant outcomes of interest emerged from outcome trajectories. It is also distinct in first establishing an outcome trajectory and then evaluating project contribution to it. Other TDEs are single step: they model how the project achieved outcomes without first considering the overarching causal mechanism—the outcome trajectory—from which the outcomes emerged.

### *Limitations of OTE*

OTE only works only when a suitable MRT exists and is agreed upon, or where the evaluator is able to develop and agree one from an existing “grand” theory. An example of an MRT that covers agricultural R4D in general, including policy change, is the three-pathway model published in

Douthwaite et al. (2017) that was recently the basis of a lesson-learning review of how the CGIAR has achieved development outcomes, to inform its next phase (Douthwaite & Child, 2021).

OTE also requires that the evaluand has contributed to a limited number of significant outcomes, for which outcome trajectories exist. OTE's employment of back-casting from existing outcomes is well matched to a portfolio approach that requires the client to accept that the bulk of a project's return on investment will come from these, and/or the client wishes to learn from evaluating significant outcomes. With respect to learning, the significant outcomes need not be all positive.

OTE was designed to work with long causal chains that have evolved over more than one project lifecycle. Projects that have achieved outcomes within their lifespan may be better served by other evaluation approaches.

## Conclusions

Long-causal-chains linking activity to eventual outcomes and impact characterize research for development work. This makes assessing R4D contributions to societal outcomes and impacts particularly complex. Researchers and evaluators have been developing theory-driven evaluation approaches in response, but issues remain. In this paper, we show how a new approach—outcome trajectory evaluation (OTE)—can partially address four issues.

- Missing or understating significant project outcomes by focusing the evaluation on a ToC developed at the start of a project when the causal links the project could activate are numerous and hard to predict. The OTE solution is to focus on the most significant outcomes achieved by the project, when the activated causal links are discoverable.
- Assigning too much causal power to projects by making them the unit of analysis, as is the case in with most project ToCs. The OTE solution is to put the OT at the centre of the analysis, and show project contribution to the OT.
- Linked to the latter, missing causal contributions from the past to outcomes claimed by projects. The OTE solution is to construct a history of the work that led to significant outcomes (i.e., the OT), including what happened before the project started.
- Researchers not learning from one evaluation to the next, and so not becoming better at contributing to outcomes through long causal chains. The OTE solution is to use MRTs as reusable conceptual frameworks to help this learning happen.

OTE borrows concepts and tools from four other evaluation approaches—process tracing, outcome harvesting, contribution analysis and episode studies and the realist evaluation idea that evaluations should accrue knowledge over time.

OTE is novel in assuming, *a priori*, that significant project outcomes emerge over time from an interacting and co-evolving system of actors, knowledge, technology and institutions, called an outcome trajectory. OTE first establishes the outcome trajectory for a cluster of significant project outcomes and then evaluates the project contribution to the cluster. Other TDEs are single step: they model how the project achieved outcomes without first describing and understanding the overarching causal mechanism—the outcome trajectory—from which the outcomes emerged.

OTE is also novel in using MRT both: (1) as a “theory of the case” to help structure and shape outcome trajectories; and (2) as reusable conceptual platforms that can, for evaluations that use the same MRT, help structure and accumulate learning from one evaluation to the next. For this paper, our four cases are all about processes of policy change in tropical agriculture so could use the same MRT. Our MRT was based on “grand” Policy Window theory, which we specified and expanded upon after each of the four evaluations. In addition to generating case-specific conclusions

and recommendations, we were also able to write a synthesis paper and develop an MRT based on all four evaluations, available for use in future evaluations of research-informed policy change processes in developing countries. The more the MRT is used, the better tested, specified and useful it is likely to become. Other MRTs exist, including a three-pathway model of how agricultural R4D in general works to achieve impact.

## Acknowledgements

This paper is based on an evaluation contracted by the International Potato Center (CIP) to assess selected policy changes brought about by the CGIAR Research Programs on Roots, Tubers and Bananas (RTB) and Agriculture for Nutrition and Health (A4NH). The authors would particularly like to express their thanks for the support of Amanda Wyatt, the evaluation manager for the A4NH case on the biofortification policy change.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Centro Internacional de la Papa.

## ORCID iD

Boru Douthwaite  <https://orcid.org/0000-0003-0018-4340>

## Notes

1. We use the term project to include programs.
2. <https://www.cgiar.org/food-security-impact/one-cgiar/>
3. As per FAO (2015) classification.

## References

- Axelrod, R. M., & Cohen, M. D. (1999). *Harnessing complexity: Organizational implications of a scientific frontier*. New York: The Free Press.
- Baxter, P., & Jack, S. (2008). Qualitative case study methodology: Study design and implementation for novice researchers. *The Qualitative Report*, 13, 544–559. <https://doi.org/10.46743/2160-3715/2008.1573>
- Belcher, B., & Hughes, K. (2021). Understanding and evaluating the impact of integrated problem-oriented research programmes: Concepts and considerations. *Research Evaluation*, 30(2), 154–168. <https://doi.org/10.1093/reseval/rvaa024>
- Breuer, E., Lee, L., De Silva, M., & Lund, C. (2016). Using theory of change to design and evaluate public health interventions: A systematic review. *Implementation Science*, 11(1), 63. <https://doi.org/10.1186/s13012-016-0422-6>
- Carden, F. (2009). Understanding influence: The episode studies approach. In D. Tussie (Ed.), *The politics of trade: The role of research in trade policy and negotiation* (pp. 273–298). Canada: International Development Research Center.
- CGIAR Consortium Office (2015). *CGIAR Strategy and Results Framework 2016–2030*. Montpellier, France: CGIAR, <https://cgspace.cgiar.org/handle/10947/3865>
- CGIAR-IEA (2017). *Evaluation of Results-Based Management in CGIAR, Final Report*. Rome, Italy: Independent Evaluation Arrangement (IEA) of CGIAR.
- Collier, D. (2011). Understanding process tracing. *PS: Political Science and Politics*, 44(4), 823–830. Retrieved from: [http://polisci.berkeley.edu/sites/default/files/people/u3827/Understanding Process Tracing.pdf](http://polisci.berkeley.edu/sites/default/files/people/u3827/Understanding%20Process%20Tracing.pdf)



- Donaldson, S. I. (2007). *Program theory-driven evaluation science: Strategies and applications*. New York: Taylor & Francis Group .
- Douthwaite, B. (2020a). *Mainstreaming of biofortification in the African Union: Evaluation of CGIAR contributions to a policy outcome trajectory*. Lima, Peru: CGIAR Research Programs on Roots, Tubers and Bananas & Agriculture for Nutrition and Health.
- Douthwaite, B. (2020b). *Development of a cassava seed certification system in Tanzania: Evaluation of CGIAR contributions to a policy outcome trajectory*. Lima, Peru: International Potato Center.
- Douthwaite, B. (2020c). *Development of a cassava seed certification system in Rwanda: Evaluation of CGIAR contributions to a policy outcome trajectory*. Lima, Peru: International Potato Center.
- Douthwaite, B. (2020d). *Control of potato purple top in Ecuador: Evaluation of CGIAR contributions to a policy outcome trajectory*. Lima, Peru: International Potato Center.
- Douthwaite, B. (2020e). *How the CGIAR contributes to policy change: Learning from four cases*. Lima, Peru: International Potato Center, <https://hdl.handle.net/10568/111287>
- Douthwaite, B., & Child, K. (2021). *How agricultural research for development achieves developmental outcomes: learning lessons to inform One CGIAR science and technology policy research*. Colombo, Sri Lanka: International Water Management Institute (IWMI). CGIAR Research Program on Water, Land and Ecosystems (WLE). 27p. (WLE Legacy Series 2).
- Douthwaite, B., Mur, R., Audouin, S., Wopereis, M., Hellin, J., Saley Moussa, A., & Bouyer, J. (2017). *Agricultural research for development to intervene effectively in complex systems and the implications for research organizations (No. 12)*. The Netherlands: Royal Tropical Institute.
- FAO (2015). Food and Agriculture Policy Classification. Available at: <http://www.fao.org/3/a-bc358e.pdf>
- Faure, G., Blundo-Canto, G., Devaux-Spatarakis, A., Le Guerroué, J. L., Mathé, S., Temple, L., Toillier, A., Triomphe, B., & Hainzelin, E. (2020). A participatory method to assess the contribution of agricultural research to societal changes in developing countries. *Research Evaluation*, 29(2), 158–170. <https://doi.org/10.1093/reseval/rvz036>
- Hedström, P., & Ylikoski, P. (2010). Causal mechanisms in the social sciences. *Annual Review of Sociology*, 36, 49–67. <https://doi.org/10.1146/annurev.soc.012809.102632>
- Jann, W., & Wegrich, K. (2007). Theories of the policy cycle. *Handbook of Public Policy Analysis: Theory, Politics, and Methods*, 125, 43–62.
- Joly, P. B., & Matt, M. (2017). Towards a new generation of research impact assessment approaches. *The Journal of Technology Transfer*, 47(3), 621–631. <https://doi.org/10.1007/s10961-017-9601-0>
- Jones, H. (2011). A guide to monitoring and evaluating policy influence. *ODI Background Note*. Overseas Development Institute. <https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/6453.pdf>
- Kingdon, J. W., & Stano, E. (1984). *Agendas, alternatives, and public policies (Vol. 45)*. xxx: Little, Brown & Co.
- Mayne, J. (2012). Contribution analysis: Coming of age? *Evaluation*, 18(3), 270–280. <https://doi.org/10.1177/1356389012451663>
- Mayne, J., & Johnson, N. (2015). Using theories of change in the CGIAR research program on agriculture for nutrition and health. *Evaluation*, 21(4), 407–428. <https://doi.org/10.1177/1356389015605198>
- Merton, R. A. (1968). *Social Theory and Social Structure*. Cambridge, UK: The Free Press.
- Meyfroidt, P. (2016). Approaches and terminology for causal analysis in land systems science. *Journal of Land Use Science*, 11(5), 501–522. <https://doi.org/10.1080/1747423X.2015.1117530>
- Mueller, B. (2020). Why public policies fail: Policymaking under complexity. *Economia (Pontificia Universidad Católica Del Perú. Departamento De Economía)*, 21(2), 311–323. <https://doi.org/10.1016/j.econ.2019.11.002>
- Pawson, R. (2013). *The science of evaluation: A realist manifesto*. London, UK: Sage.
- Pawson, R. (2017). Middle range theory and program theory evaluation: From provenance to practice 1, *Mind the gap* (pp. 171–202). Milton Park, UK: Routledge.

- Paz-Ybarnegaray, R., & Douthwaite, B. (2017). Outcome evidencing: A method for enabling and evaluating program intervention in Complex systems. *American Journal of Evaluation*, 38(2), 275–293. <https://doi.org/10.1177/1098214016676573>
- Reed, M. S., Ferre, M., Martin-Ortega, J., Blanche, R., Lawford-Rolfe, R., Dallimer, M., & Holden, J. (2021). Evaluating impact from research: A methodological framework. *Research Policy*, 50(4), 104147. <https://doi.org/10.1016/j.respol.2020.104147>
- Renkow, M. (2018). *A Reflection on Impact and Influence of CGIAR Policy-Oriented Research*. Rome, Italy: Standing Panel on Impact Assessment (SPIA), CGIAR Independent Science and Partnership Council (ISPC).
- Rule, P., & John, V. M. (2015). A necessary dialogue: Theory in case study research. *International Journal of Qualitative Methods*, 14, 4. <https://doi.org/10.1177/1609406915611575>
- Sabatier, P. A. (Ed.) (2007). *Theories of the Policy Process*, 3. Boulder, USA: Westview Press.
- Sabatier, P. A., & Weible, C. M. (2007). The advocacy coalition framework: Innovation and clarification. In P. A. Sabatier (Ed.), *Theories of the policy process*, 2 (pp. 189–220). Boulder, USA: Westview Press.
- Stachowiak, S. (2013). *Pathways for change: 10 theories to inform advocacy and policy change efforts*. Washington: Center for Evaluation Innovation, <https://www.evaluationinnovation.org/publication/pathways-for-change-10-theories-to-inform-advocacy-and-policy-change-efforts/>
- Sterman, J. D. (2002). All models are wrong: Reflections on becoming a systems scientist. *System Dynamics Review*, 18, 501–531. <https://doi.org/10.1002/sdr.261>
- Wilson-Grau, R. (2018). *Outcome harvesting: Principles, steps, and evaluation applications*. North Carolina, USA: IAP.