Outcome Evidencing: A Method for Enabling and Evaluating Program Intervention in Complex Systems

Rodrigo Paz-Ybarnegaray[1] and Boru Douthwaite[2]

Abstract

This article describes the development and use of a rapid evaluation approach to meet program accountability and learning requirements in a research for development program operating in five developing countries. The method identifies clusters of outcomes, both expected and unexpected, happening within areas of change. In a workshop, change agents describe the causal connections within outcome clusters to identify outcome trajectories for subsequent verification. Comparing verified outcome trajectories with existing program theory allows program staff to question underlying causal premises and adapt accordingly. The method can be used for one-off evaluations that seek to understand whether, how, and why program interventions are working. Repeated cycles of outcome evidencing can build a case for program contribution over time that can be evaluated as part of any future impact assessment of the program or parts of it.

Many programs that set out to make a difference to people's lives are increasingly being understood as complex interventions in complex systems (Pawson, 2013). Complexity science provides a range of insights into how change happens, including the idea of emergence and positive feedback loops. Taking these insights onboard means program staff must necessarily grapple with uncertainty as to the nature, size, and pathways to intended program impact (Douthwaite, Kuby, van de Fliert, & Schulz, 2003; Patton, 2011), which can be problematic when donors require reassurance about the likely returns on their investments. This is particularly so for agricultural research for development programs, which are adaptive research programs in which researchers are expected to specify and implement according to program theory linking their research to uptake and impact (Mayne & Stern,

[1] WorldFish, Cochabamba, Bolivia
[2] WorldFish, Cushalogurt, Ireland

Corresponding Author:
Boru Douthwaite, WorldFish, Cushalogurt, Westport F28 W654, Ireland.
Email: bdouthwaite@gmail.com

2013). Starting with research questions, whose answers are uncertain, adds a step in the causal chain linking program intervention and its eventual impact compared to straight development programs that start with already-proven treatments or technology.

In complex systems, there are rarely ever any magic bullets: No intervention will ever work the same way, everywhere for everyone. In some contexts, some program offerings will work, and in others, they will not (Pawson, 2013). Evaluation methods therefore need to help understand how different aspects of programs work, for whom in different contexts. In other words, they need to help unpack the causal black box between program intervention and program outcomes (Astbury & Leeuw, 2010; Stame, 2004). Most traditional impact evaluation methods focus on establishing whether a program made a difference and less on understanding how it worked, or didn't, in different contexts (Mayne & Stern, 2013; Stern, 2015). Given long time lags between research and impact, for example, 20 years or more for agricultural research (Collinson & Tollens, 1994), many impact evaluations are carried years after the program has finished. Hence, traditional impact evaluation methods generally do not help staff in ongoing programs to identify and learn from the parts of the program that are working and have the potential, if supported and scaled, to make a real difference.

The literature that calls for complexity-aware monitoring and evaluation to fill this gap is large and growing (e.g., Britt & Patsalides, 2013; Douthwaite et al., 2003; Mayne & Stern, 2013; Patton, 2011; Rogers, 2008; Stame, 2004; van Mierlo, Arkesteijn, & Leeuwis, 2010). Advantages claimed for such approaches include helping program participants to identify system barriers to tackle and the actors to involve in doing so through reflexive learning (van Mierlo et al., 2010), supporting adaptive management and innovation (Douthwaite et al., 2003; Patton, 2011), and identifying unexpected and emerging opportunities and challenges (Snowden, 2010).

Theory-based evaluation, which began with Chen and Rossi in the 1980s, is increasingly seen as a way to deal with complexity, both of programs and the issues they address (Stame, 2004). A complex program is one which works in an integrated fashion on several scales or issues, hoping for synergies to bring about system change that working on a single issue would not bring about (Stame, 2004). Complex issues can neither be defined nor solved by a single actor or organization. Tackling them requires the involvement of various actors in a social learning process that reduces uncertainty, increases agreement on the way forward while at the same time surfacing and challenging the rules and relations that maintain a problematic status quo (Arkesteijn, van Mierlo, & Leeuwis, 2015). Developing and collectively revisiting theories of change (also sometimes called program theory) is a way of supporting social learning in practice that puts theory-based evaluation at the center of programmatic effort to support and trigger learning, innovation, and change in complex systems.

Theory-based evaluation approaches deal with internal validity by focusing on the building and testing of theory. The theories can model complex processes, including the presence of feedback loops and tipping points, which can then be tested. They deal with generalizability by identifying contextual factors that influence whether programs trigger causal mechanisms or not, while at the same time building mid-level theories that allow abstraction and the accumulation of learning about what works and what doesn't across sectors (Pawson, 2013). Theory-based evaluation has been seen as particularly useful in formative programs and small "*n*" studies, where experimental approaches are not appropriate because of small sample size (White & Phillips, 2012). They can strengthen experimental approaches by providing causal explanation alongside statistical proof of treatment effect. Theory-based approaches, in particular realist evaluation, are based on a generative view of causality that holds that causality can be established by identifying the causal mechanisms that link cause to effect (Pawson, 2013; Westhorp, 2014). In contrast, experimental approaches to evaluation hold a "secessionist" view that causality cannot be seen but only inferred "from repeated succession of one such event by another" (Pawson & Tilley, 1997, p. 5). Secessionist methods attempt to

exclude every conceivable rival cause from the experiment to be left with just one secure causal link (Pawson & Tilley, 1997).

The literature has less to say about the experience of developing and using complexity-aware impact evaluation. This article describes the development of a rapid evaluation approach called *outcome evidencing* that is based on the development and revisiting of theories of change. Outcome evidencing was developed within a systems-focused research for development program of the CGIAR, formerly known as the Consultative Group on International Research. The CGIAR is a worldwide partnership addressing agricultural research for development carried out by 15 research centers through 15 CGIAR Programs. CGIAR work contributes to the global effort to tackle poverty, hunger, and environmental degradation. As of 2014, the CGIAR employed more than 8,500 researchers and support staff worldwide, with an annual budget of US$800 million (Agropolis International, 2015).

Our objectives are 2-fold: to describe and critically reflect on an evaluation approach that may be of interest to other programs and to share the practical considerations involved in starting to use complexity-aware evaluation methods.

The Aquatic Agricultural Systems (AAS) Program

The goal of the CGIAR Program on AAS is to improve the well-being of poor people dependent on AAS (2011) by putting in place the capacity for communities to pull themselves out of poverty. AAS began in 2011 by establishing programs of work in five locations bounded by an important AAS. The five locations, known as "hubs" were in Zambia (the Barotse Floodplain), Bangladesh (the Southern Polder Zone), Cambodia (Tonle Sap Floodplain), the Philippines (Visayas–Mindanao coastal areas), and Solomon Islands (Malaita and Western Province coastal areas). The program had an aspirational goal to make positive difference on the livelihoods of 6 million poor and marginalized by 2023 (AAS, 2014). By the end of 2013, AAS staff were engaging in between 8 and 16 focal communities and with hub- and national-level stakeholders in each hub.

In the same period, the program developed the research in development (RinD) approach as its main vehicle for achieving impact. The RinD approach involves research teams engaging at both community- and hub-scales to tackle a commonly agreed development challenge relating to the dominant hub AAS. For example, in both Zambia and Cambodia, the challenge related to seasonal flooding. The RinD approach creates new and safe dialog and action spaces for stakeholders and communities to engage with one another long enough to build trust, motivation, capacity, and insight to tackle deep seated issues and constraints that maintain the status quo.

AAS staff worked with theory of change/program theory from the outset (Douthwaite, Apgar, & Crissman, 2014). The overarching program theory is based on the premise that agricultural research processes (e.g., multipartner collaborations) and outputs (i.e., new technologies) work to catalyze and foster processes of rural innovation. It is these innovation processes, that maybe technical, institutional, or both, that lead to development outcomes. The RinD approach is a way of building collaborations across institutional and scale boundaries (e.g., between farmers and researchers or between different government ministries). This program theory was unorthodox: Most CGIAR Programs build their program theory around the adoption and use of new technology rather than building a more enabling environment for local innovation (see Dugan, Apgar, & Douthwaite, 2013, for more detail on the AAS Program and the RinD approach).

The authors, both with responsibility for program evaluation, were aware that the investment being made in AAS was contingent on demonstrating that AAS' program theory was working in practice in the first phase of the program. We expressed the evaluation challenge in terms of two evaluation questions:

- What outcomes is AAS contributing to?
- Do these provide evidence that the AAS' program theory of change is credible, and how do they help us understand why (or why not)?

The Outcome Evidencing Method

We developed outcome evidencing to meet AAS' evaluation challenge. Outcome evidencing is an adaptation of outcome harvesting (Wilson-Grau & Britt, 2012), enriched by concepts from Scriven's (1976) modus operandi method and systems thinking (in particular from agricultural innovation systems thinking, e.g., Brooks & Leovinsohn, 2011). Outcome evidencing is a participatory, theory-based evaluation approach that seeks to identify outcomes, resulting from program implementation fast enough to influence ongoing program implementation and rigorously enough to make plausible causal claims to substantiate or challenge the overarching program theory.

We developed outcome evidencing through what began as a pilot of outcome harvesting in the five AAS Program hubs. The result is a method with 10 steps, which we describe in Figure 1. In subsequent sections, we describe the theoretical and practical reasons for the changes made to outcome harvesting and give examples of how the outcome evidencing steps worked.

Step 1: Agree on the Evaluation Questions and the Use of the Evaluation Results

Program leadership agree on the evaluation questions and how the evaluation results will be used. The AAS questions are the ones identified above.

Step 2: Identify Areas of Change

Knowledgeable program staff, in particular "change agents," identify areas of change to which the program is contributing. Change agents are the people implementing the program in the field and is a term borrowed from outcome harvesting. The areas of change are where program intervention has taken hold and participants are starting to work together. They may fall outside the initial program theory of change. They are spaces where program staff think things are starting to change, whether they are expected or not.

Step 3: Identify and Describe Outcomes

Step 3 is to identify and describe outcomes occurring within the identified areas of change. This is done through asking change agents and looking for outcomes recorded in process documentation, particular records kept by field staff. Either way, the outcomes should be described in terms of a single phrase that can be written on card to allow for subsequent clustering in the next step. Other basic information should also be recorded for each outcome on a simple template.

Step 4: Identify Outcome Trajectories

A number of outcomes will be identified for each area of change. The next step is to make sense of these. This happens in a workshop attended by staff and stakeholders involved in implementing the program. Participants first cluster outcomes that they think are causally related. They then build a multi-cause diagram (Burge, 2013) as a way of collectively agreeing on what those relationships are and in so doing add in or reject some outcomes. The result is a stakeholder-generated theory of change called an outcome trajectory. Outcome trajectories identify and explain the causal links connecting program intervention to outcomes within areas of change.
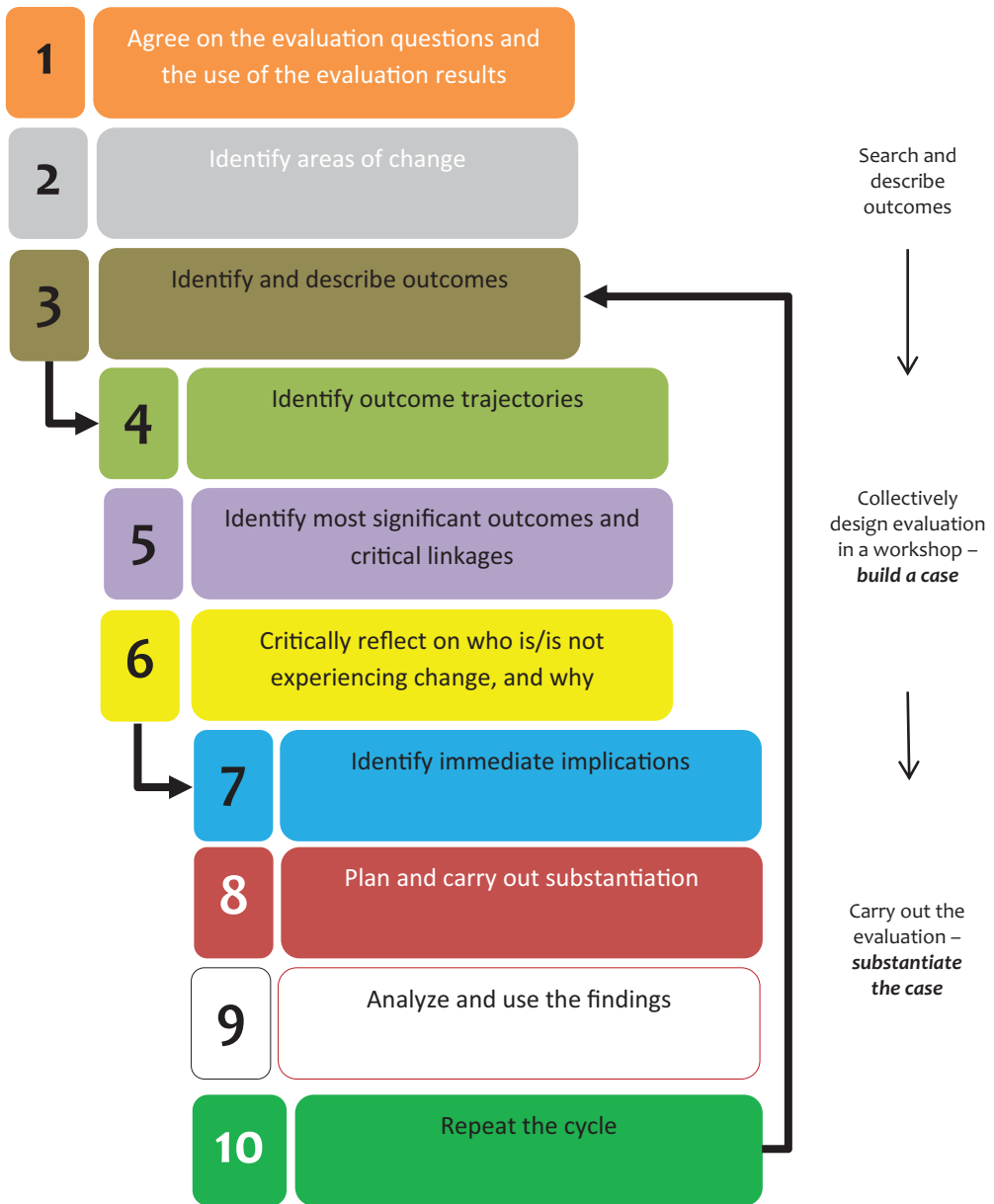
Figure 1. Ten steps of an outcome evidencing process.

Step 5: Identify Most Significant Outcomes and Critical Linkages in the Outcome Trajectories

The next step is to identify the critical outcomes and linkages within outcome trajectories upon which the program's claim to have made a contribution most depend. This step is carried out in the same workshop by the people most familiar with the outcome trajectory. Choices are presented in plenary to allow for challenge.

Step 6: Critically Reflect on Who Is Experiencing Change and Who Isn't

AAS uses research to trigger or support processes of innovation. Innovation processes benefit participants more than nonparticipants (Rogers, 2010). AAS' goal, shared with many other programs, is to benefit the poor and marginalized who are usually bypassed by mainstream development activity. Hence, we include a step that involves analyzing outcome trajectories in terms of social and gender equity, inclusion, and power.

Step 7: Identify Immediate Implications

The workshop produces learning and insight about which there is sufficient agreement to be acted upon immediately. To make sure, this happens a workshop report identifying these measures is written and circulated to relevant people as soon as possible. Another strategy is to hold the outcome evidencing workshop immediately before annual planning, so that the people involved in both can take the learning with them.

Step 8: Plan and Carry Out Substantiation

The workshop provides sufficient information to plan and carry out the substantiation of the outcome trajectories. Substantiation is carried out by an evaluator who may be internal or external. Internal or "self-evaluation" has been found to be more self-critical, and the results are more useful than when an external evaluator is used (Douthwaite et al., 2003), whereas external evaluation may carry more weight with an external audience. Developing and implementing the plan requires a number of decisions to be made as to which key informants to interview, which documentation to check, and the evaluation report length and structure. Where relevant, the evaluator looks for rival explanations for the program's claims and adjudicates between them.

Step 9: Analyze and Use the Findings

The evaluator who has carried out the substantiation and other staff leading the outcome evidencing process analyze the findings from the substantiation to complete the evaluation report. The evaluator should use the findings to challenge and support the existing program theory. Outcome evidencing was designed to be repeated annually or biannually within a program that needed the results to inform its adaptive management. Outcome evidencing can also be used for one-off evaluations. In either case, the authors of the evaluation report have a responsibility to promote the use of the findings.

Step 10: Repeat the Outcome Evidencing Cycle

Repeating the outcome evidencing cycle annually allows AAS to explore how the outcome trajectories first identified, have evolved, and grown. This is done in subsequent repetitions of Step 3 by collectively deciding if new outcomes map onto existing outcome trajectories, and if they do whether they add to or challenge the outcome trajectory theory of change. New outcome trajectories may emerge in this process if new outcomes do not map onto existing trajectories. New outcomes could also surface new outcome trajectories. In doing so, the overall program theory is built, challenged, and substantiated.

Design Rationale

In this section, we explain and justify the design choices made in developing the approach just described and explore what outcome evidencing might offer, at least in theory.

Table 1. Comparing the Steps in Outcome Evidencing With Outcome Harvesting.

Outcome Evidencing	Outcome Harvesting (Adapted From Wilson-Grau & Britt, 2012, p. 4)
Implemented by: one or more evaluators who facilitate the inquiry and evaluate parts of it and who may be internal to the program, external, or both	Implemented by: "harvesters," usually external to the program being evaluated
Step 1: Design the evaluation	Step 1: Design the evaluation
Step 2: Identify areas of change	Step 2: Gather data and draft outcome descriptions. Harvesters glean information about changes that have occurred and how the change agent contributed to these changes. The harvesters write preliminary outcome descriptions
Step 3: Identify and describe outcomes across areas of change	Step 3: Formulate outcome descriptions. Harvesters engage directly with change agents to review and classify the draft outcome descriptions and identify and formulate new ones
Step 4: Identify outcome trajectories	
Step 5: Identify most significant outcomes and critical linkages within the outcome trajectories	
Step 6: Critically reflect on who is/is not experiencing change and why	
Step 7: Identify immediate implications of findings and insight coming from the workshop	
Step 8: Plan and carry out substantiation of significant outcomes and critical linkages	Step 4: Substantiate. Harvesters obtain the views of independent individuals knowledgeable about the outcomes and how they were achieved
Step 9: Analyze and use the findings	Step 5: Analyze and interpret. Harvesters organize outcome descriptions through a database in order to make sense of them, analyze and interpret the data, and provide evidence-based answers to the evaluation questions
	Step 6: Support use of findings. Drawing on the evidence-based, actionable answers to the evaluation questions, harvesters propose points for discussion to harvest users, including how the users might make use of findings. The harvesters facilitate discussion among harvest users
Step 10: Repeat the outcome evidencing cycle	

We originally chose outcome harvesting as the method best able to answer the AAS evaluation questions because it starts with emerging outcomes and works backward to establish if and how program interventions had contributed by reconstructing and validating causal pathways (Wilson-Grau & Britt, 2012). Outcome harvesting has 6 steps as shown in Table 1 compared to 10 in outcome evidencing, as shown in Table 1.

When we started using outcome harvesting (Step 2), we realized that the outcomes key staff identified occurred in areas of change that staff in the country hubs could identify, for example, small-scale fisheries management or mango production in different barangays in the Philippines. The innovation systems' literature helped us understand these as potential sociotechnical niches. Niches are protected spaces that allow people to experiment with novelty in technology and/or institutions (Klerkx, Van Mierlo, & Leeuwis, 2012) and are a core concept of strategic niche management (Kemp, Schot, & Hoogma, 1998). According to this theory, when niches are properly

constructed and linked, they can act as building blocks for broader societal changes toward sustainable development (Schot & Geels, 2008). Hence, strategic niche management provides some detail to the AAS' program theory described above, specifically that the program creates, supports, and guides sociotechnical niches to be building blocks that come together to help achieving the program's goal. We realized that focusing rapid evaluation on if and how program intervention is contributing to niches was a way of answering the second evaluation question relating to the credibility and workings of the AAS Program theory. We also realized that evaluation findings could guide how the program intervenes in the future to link the niches to bring about broader change.

Our next insight was that the outcomes staff thought the program was causing were not independent: Rather, they appeared to clump together in causal clusters. Hence, it made sense to evidence these clusters as a whole rather than the individual outcomes, as outcome harvesting does; hence, we introduced a step to do so (outcome evidencing Step 4). We thought the best way was to have key staff and partners involved in implementation to agree the clusters in a workshop because collectively they would have the best sense of what might be causing what. We called the clusters "outcome trajectories" to signify that the clusters appeared to have some momentum and coherence to them. We defined outcome trajectories as stakeholder-generated theory of change that identify and explain the causal links connecting program intervention to outcomes within areas of change. Outcome trajectories, like the niches themselves, can be protected, nurtured, or dampened down and killed off. The idea of nurturing or dampening outcome trajectories borrows from Snowden (2010) who recommends working with "beneficial coherence" (i.e., patterns of activity that start to resonate with people) as an effective strategy for fostering change in complex adaptive systems. To do this in practice requires rapid evaluation (McNall & Foster-Fishman, 2007) that can provide actionable findings within annual planning cycles.

We also realized that the clusters of outcomes that participants were identifying were similar to Scriven's modus operandi. Scriven (1976) argues that interventions create distinctive patterns of effects, which are its modus operandi, or signature. Programs contribute to change through triggering characteristic causal chains of events. Finding evidence of this signature helps the program claim it made a causal contribution. Identifying the causal pattern also helps the program understand how its interventions are working, a design criterion for complexity-aware evaluation.

Step 5 in outcome evidencing involves identifying within an outcome trajectory the most significant outcomes and critical linkages for substantiating causal claims. The rationale for this step comes from Popper (1992) who argued that in any theory, some parts of it can be taken for granted while others require greater scrutiny. Given this, we assume that some outcomes and causal links are more crucial for understanding and substantiating program impact claims than others: These require greater scrutiny. We assume the scrutiny will also help clarify the program's unique modus operandi.

Both the modus operandi method and outcome harvesting use a detective analogy, that is, they both claim that they establish causality in the same way a detective might by building a case. In doing so, they are signaling the assumption of a generative view of causality described above that outcome evidencing inherits. A generative approach to causality is commonly used in everyday life by doctors, mechanics, and so forth as well as detectives. A doctor can diagnose a disease through observing a pattern of critical symptoms (its modus operandi) just as a car mechanic can diagnose a car is overheating due to a stuck thermostat by feeling the temperature of certain hoses (Mohr, 1999). In contrast, most impact evaluation carried out of agricultural research has assumed a secessionist view of causality and with it the requirement for a "counterfactual" which is usually a control group. However, in addition to the small n and black box issues already discussed, establishing "without" control groups when intervening in complex systems can be ethically and practically problematic (Scriven, 2008) because researchers take the time of control group members without giving anything

back and may work to stop the spread of technology to them to preserve their trial. Hence, a method that can work without control groups is a better match to AAS' ethical position and its evaluation and learning needs. In place of a control group, Outcome evidencing borrows from the modus operandi method and realist evaluation to develop rival causal explanations and adjudicate between them (Pawson & Tilley, 1997; Scriven, 1976).

Taking the detective analogy and thinking of achieving impact as a "crime," then outcome evidencing seeks to at least prove that AAS has been an accessory to impact, in other words that program actions have in some way contributed to outcomes identified. This is different to "attribution" claims that seek to establish how much of the impact can be attributed to the program. Seeking to show contribution is more realistic within complex systems, where it is useful to think of outcomes following from the interaction of causal packages to which different stakeholders and different programs contribute parts (Mayne, 2011).

Steps 7 and 9 in outcome evidencing both involve use of evaluation findings. Step 7 allows for the rapid use of learning and insight which has clear and uncontroversial implications. In terms of behavior-change theory, this can be thought of as single-loop learning. Step 9 in contrast allows space for double-loop learning (Argyris, 1976; Argyris & Schön, 1978) that unlike single-loop learning involves questioning of underlying assumptions and mental models (i.e., theories of change). It is an opportunity for programs to challenge and build their program theory, usually developed before the program started, using the verified outcome trajectories that outcome evidencing reveals. If outcome trajectories contradict the program theory, then this may require program staff to change their underlying assumptions and potentially how the program is implemented. It is in this way that outcome evidencing can support adaptive management, defined as an actively adaptive, probing, and deliberative process of intervening in socio-ecological systems (Walters, 1986).

In sum, we adapted outcome harvesting to meet AAS Program needs informed by several bodies of literature and Scriven's modus operandi method. We call the result outcome evidencing to emphasize the importance of the case-building detective analogy for both outcome harvesting and the modus operandi method as well as the assumption of generative causality as opposed to a counterfactual one. Outcome evidencing is a method that theoretically can identify sociotechnical niches, understand how program intervention is contributing to their development as building blocks to achieving broader systemic goals, and identify and substantiate causal claims. In the next section, we present early experience from developing and using outcome evidencing before then critically reflecting on the degree to which the theoretical potential has been realized.

Using Outcome Evidencing in Practice

Piloting of outcome harvesting began in January 2014 in Bangladesh followed by Cambodia, Solomon Islands, Zambia, and finally the Philippines by the end of 2014. Bangladesh, as the first hub to start, is where the method went through the greatest change. The main adaptations were that we combined Steps 3 and 5 of outcome harvesting together—change agent description of the outcomes and their analysis and interpretation—in one workshop, in which the change agents collectively identified and agreed outcome trajectories. The substantiation step, Step 4, became verification of the theories of change developed to describe the outcome trajectories. We also dropped the use of "independent but knowledgeable people" to validate outcome claims. We realized that because we were validating emerging outcomes, the people knowledgeable about them were also likely to be involved with the program in one way or another and therefore not independent. We used evaluators—either external or internal—to carry out the verification step. It was also in Bangladesh when we first acknowledged the changes in the steps by calling the method "outcome evidencing." This was not necessarily
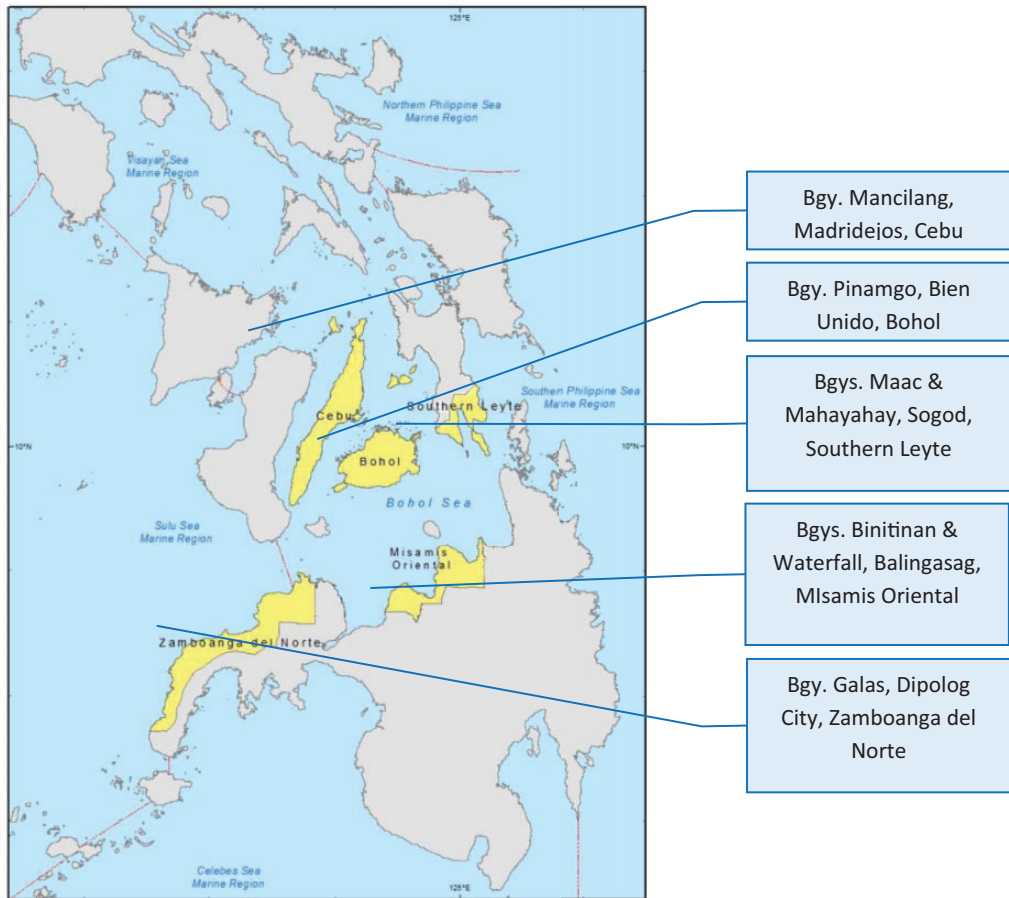
Figure 2. Visayas–Mindanao hub and the location of the Aquatic Agricultural Systems focal barangays (Brgy = barangay = smallest government administrative unit).

to suggest we had developed a new approach, but rather that what we employing was different to outcome harvesting in a number of ways.

As the last hub to finish, the Philippines is our main worked example for this article (see AAS, 2014). It is also the most successful pilot because the implementation team was able to learn from other hubs.

Implementing Outcome Evidencing in the Philippines

The AAS Program in the Philippines began in 2012 focused on AAS in the Visayas–Mindanao area (see Figure 2). The program engaged local-, regional-, and national-level stakeholders to agree a common hub development challenge (HDC)—''increasing pressure from a rapidly growing population on an already degraded resource base within the context of climate change'' (AAS, 2014, p. 3). The program goal was to enhance resilience, improved well-being, and inclusive growth for the poor and marginalized in the hub. The program's overarching premise was that since the causes of marginalization and poverty are multifaceted and manifest at multiple scales, the goal could be best achieved by better linking hub- and community-level actors together through collaborative research

activity. To this end, AAS selected seven focal communities and facilitated a community-led participatory action research process to achieve community goals related to the HDC. These included rehabilitating abaca (a member of the banana family grown for fiber) production devastated by disease and improving community fisheries management in the face of collapsing catches. At the same time, AAS engaged hub-level stakeholders by setting up three research initiatives on sustainability, value chain development, and governance and policies. The AAS team facilitated linkages between the community-level participatory action research and the research initiatives as a mechanism to support, guide, and bring niches together. One of the main program outputs was the approach it developed to do all this, called the RinD approach.

The Philippines AAS team met to agree areas of change in March 2014 after the program had only been in operation little more than 1 year. The team identified five:

1. Small-scale fisheries management in Barangay Mancilang, Madridejos, Cebu.
2. Emerging Community Based Small Scale Fisheries Governance in Balingasag, Misamis Oriental.
3. Mango production in Barangay Pinamgo, Bien Unido, Bohol.
4. Rehabilitation of Abaca Production in three communities in Sogod, Southern Leyte.
5. Vegetable home gardening in Barangay Galas, Dipolog City.

With these areas in mind, the team organized an outcome evidencing workshop (October 14–17, 2014) to which they invited change agents involved at community- and hub levels, themselves included. Twenty-eight people attended the workshop who identified 80 outcomes (Step 3) within the five areas of change. They then identified two clusters of outcomes: those achieved at community-level and those achieved at hub level, across the five areas of change. Participants collectively built a causal diagram for each cluster to identify and describe an outcome trajectory operating at each scale (Step 4). The outcome trajectories were given descriptive names:

1. Communities recognizing their strengths, resources, and gaining better linkage with institutions to undertake actions to improve their lives.
2. Hub-level partners are recognizing that the RinD approach is markedly different from their approaches and starting to adopt aspects of it.

The resulting causal diagrams were complicated involving over 40 outcomes linked together with explanatory text. Figure 3 provides a simplification showing the partner trajectory of change.

Participants then identified evidence to support AAS contribution to the outcomes identified in the outcome trajectories. Table 2 shows key evidence for two outcomes as well as sources of verification.

Step 6, to critically reflect on who is/is not experiencing change, and why was not carried out in the Philippines but was in Zambia. The Philippine team felt that the nature of their deep immersion in the communities meant that they did not require a separate step to consider who was and was not benefiting. The Zambian team pioneered the step by virtue of their strong gender research capacity. In Zambia, workshop participants discussed and analyzed outcome trajectories and the most significant changes along them from a social and gender equity perspective by answering the following questions:

- What vulnerable or marginalized groups are being, or could be, directly or indirectly affected by the change?
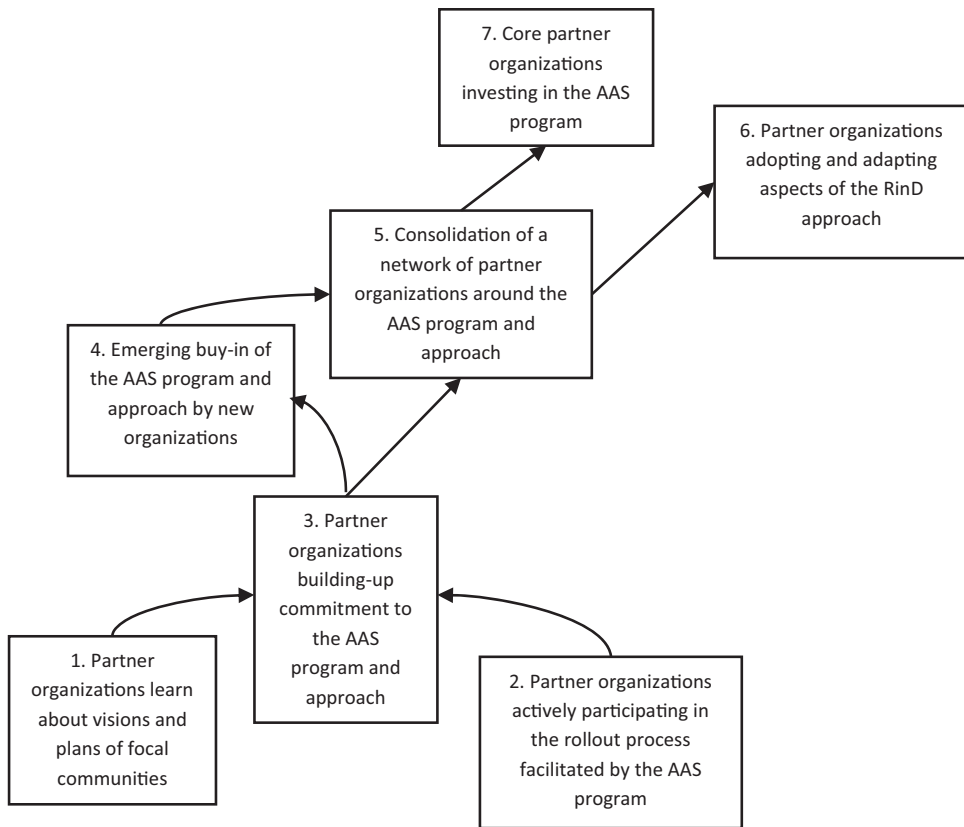- Does the outcome trajectory:

Figure 3. A simplified version of the multicause diagram developed to describe the hub-level trajectory of change resulting from Aquatic Agricultural Systems intervention.

- Promote equal opportunities for vulnerable and marginalized groups? *Yes/How is that happening? Or No/Why not?*
- Strengthen positive norms that support social and gender equality and an enabling environment? *Yes/How is that happening? Or No/Why not?*
- Challenge norms that perpetuate social and gender inequalities. *Yes/How is that happening? Or No/Why not?*

In Step 7, various immediate implications were identified during the workshop, two of which related to the hub-level partnership trajectory. The first was that while the RinD approach was working, the team now needed to follow up on partner commitments (e.g., funding, people, and processes). The second was the work closely and further engage with all five Regional Development Councils (RDCs), as this was a specific strategy that was proving particularly effective in providing impetus to the outcome trajectory.

Table 2 shows the key outcomes and sources of documentary evidence that guided substantiation of the partner outcome trajectory in the Philippines, as part of Step 8. The substantiation was carried out by the AAS team themselves to maximize their learning. They verified the outcome trajectories through a desk review. The quality of process documentation had been a program priority from the outset, and this helped the team build a case for the outcome trajectories. Triangulation between data sources was the main means of corroboration (see Kennedy, 2009, for description of triangulation).

Table 2. Evidence to Support the Aquatic Agricultural Systems (AAS) Contribution to the Hub-Level Outcome Trajectory in the Philippines.

Outcome in Outcome Trajectory	Specific Example Offered as Evidence	Source of Verification
Partners building-up commitment to the AAS Program and approach	Endorsement of the AAS Program by the Regional Development Council (RDC)	– RDC endorsements in Regions 7, 8, and 10 – Letter from the Region 10 Director of the Department of Science and Technology (DOST) to the Under-secretary – Minutes of RDC meetings
Partner organizations adopting and adapting aspects of the research in development approach	Different partners investing in activities that are oriented to tackle the hub development challenge (HDC)	– Partners' investments in activities to tackle the HDC – Memorandum of understanding, meeting report, and plan of work and budget of WorldFish—Philippine Council for Agriculture, Aquatic, and Natural Resources Research and Development Technical Working Group – Memoranda of agreement with department of agriculture projects for work on capacity building for AAS, climate change, and Tilapia – Memoranda of understanding with the DOST Regions 8, 9, and 10

Box 1 presents an excerpt of the final report, presenting evidence of the two specific outcomes from Table 1. The original report was extensively referenced with hyperlinks to process documentation held on an internal site.

> **Box 1:** Edited excerpts from final outcome evidencing report for the Philippines for the partnership outcome trajectory (AAS, 2014).

Endorsement of the AAS Program by the RDC

The RDC is the highest policy-making body and serves as the regional counterpart of the National Economic and Development Authority chaired by the President of the Republic. RDC's primary responsibility is to coordinate and set the direction of all economic and social development efforts in the region. It also serves as a forum, where local efforts can be related and integrated with national development activities.

The AAS Program has been endorsed by the RDCs of Region 7, 8, and 10. This was facilitated by our partners who are members of the RDC. Without our partners having sponsored the presentation of the AAS Program in the RDCs' sectoral committees which, in some occasions they head, our entry into the RDCs could have been difficult. The principles we shared with the Regional Offices of the Department of Science and Technology (DOST) facilitated our access into RDCs. In some

Table 3. Status of Endorsement of Aquatic Agricultural Systems (AAS) in the Regional Development Council (RDC).

Region	Status	Resolution	Sponsor
Region 7	AAS endorsed by RDC 7 Economic Development Committee (EDC)	RDC Resolution 1 (s. 2014), "Endorsing to Potential Partner Agencies and Convergence Groups in Central Visayas the Consultative Group of International Agricultural Research Program on Aquatic Agricultural Systems for Replication in Other Areas in the Region"	Regional Director of DOST 7
Region 8	AAS presented to full council of RDC 8	RDC VIII Resolution No. 21 (s. 2014), "Endorsing the Aquatic Agricultural Research Program to the National Government Agencies and Local Government Units"	Regional Director of DOST 8 and Chair of the RDC 8-Social Development Committee
Region 10	AAS endorsed by RDC 10-EDC	RDC X Resolution No. 33 (s. 2014), "Endorsing the Consultative Group on International Agricultural Research Program on Aquatic Agricultural Systems"	Regional Director of DOST 10

instances, DOST Regional Directors defended the program in full RDC sessions. Table 3 shows the status of endorsement. Regional development planning is necessary to address the uneven economic and socio development of the country, and these endorsements open the gates for AAS to engage as active participant in national development.

The evaluation team did not identify rival causal explanations and adjudicate between them. This is because they chose to substantiate progress indicators rather than outcomes, reflecting the early stage of the program in its intended life span. For example, with respect to Table 3, of interest is what endorsement of AAS by the RDC will lead to over time. Once identified, rival explanations of these changes can be developed and adjudicated between.

In Bangladesh, where the program had been implemented for a year longer, the evaluator sought to substantiate a farmer self-confidence and leadership trajectory among others. He used quotes from the logs of AAS staff working in the 16 communities as evidence. For example, in Bazarkhali village, a farmer claimed the program had led to the group's increased ability to take responsibilities for different work, improvement in their problem-solving abilities, and increase in social status for its members. He recommended further field work to substantiate these claims but did not have the resources to do so himself. His approach was to attempt to triangulate such claims through several sources and give a sense of the strength of the evidence as a result.

Different Partners Investing in Activities That Are Oriented to Tackle the HDC

The HDC, and a strategic framework to tackle it, was agreed with stakeholders through a series of regional consultation workshops in 2012 culminating in the stakeholders' consultation workshop and design workshop in 2013. The collective development of both allowed stakeholder to explore collaboration including the support of the endeavors tackling the HDC. At least US$390,000 has been invested (both in cash and in kind) by at least nine partners since 2013.

Bangladesh and Zambia used an external evaluator to validate outcome trajectories. In Zambia, the external evaluator spent over 20 days in the field validating five outcome trajectories. One of the

critical outcomes he substantiated was the claim that the program had led to an increase in canal clearance, important for communities to grow more irrigated crops. He did this by visiting AAS focal communities, where the claim was being made and interviewed the community leaders and members involved in the task. He also double-checked their accounts against records kept of which villages had participated and the amount of time spent. Other claims were substantiated in a similar way.

In Step 9, the AAS team in the Philippines reflected on the results and came up with important learning and affirmation. For example, from the partnership outcome trajectory, they concluded that it is possible within a relatively short period (about 1 year) to facilitate research and development organizations to work toward a common goal through setting up a number of collaborative research initiatives. They realized that what it takes are communities that can organize and express their development requirements and an ''honest broker'' able to link communities' visions and dreams and organizational mandates. They concluded that research organizations can play this role because of the neutral space that research provides for people to work together and in so doing build trusting relationships. Outcome evidencing also helped them realize the effort required carrying out the honest broker role takes resources away from research and a challenge the team faces is getting the balance right when working for a research institution that is ambivalent about using resources for bridging and brokering.

AAS planned to repeat the outcome evidencing cycle in 2016 in all the hubs; however, AAS was unexpectedly closed down in early 2016 due to CGIAR funding cuts.

Differences Between Hubs

There were differences between how outcome evidencing was implemented in each hub due how local teams interpreted the instructions provided by the authors and how those instructions changed as we learned from the sequential pilots (Table 4). The most important differences occurred in Steps 2, 3, and 4. All hubs identified areas of change, but it was only in the Philippines, where program staff had spent extended periods of time in the field, that identified areas of change in terms of concrete collaborations. In Bangladesh, in contrast, the two areas of change were the more generic: community engagement and stakeholder engagement. Zambia and the Philippines clustered relatively large numbers of unprocessed outcomes in the outcome evidencing workshop, while the other hubs carried out some form of amalgamation, usually by the AAS team, before the workshop. Two hubs, Cambodia and Solomon Islands, used narratives as way of identifying and clustering outcomes. There was also a difference in whether hubs chose to use an external evaluator or use internal resources to verify the outcome trajectories. This choice was made largely on the basis of available budget.

Discussion

In the previous sections, we have tried to give a sense of the practicalities of developing and using a complexity-aware evaluation method in the field. The results suggest that outcome evidencing is fulfilling at least some of its potential as a complexity-aware rapid evaluation approach. Staff using the method were able to identify five areas of change in the Philippines little more than 1 year after the program began. In each of the five areas, AAS, through its facilitated engagement at community- and hub-scale, was helping create ''niche'' conditions by providing ''protected spaces that allow nurturing and experimentation with the co-evolution of technology, user practices, and regulatory structures'' (Schot & Geels, 2008, p. 538). The method then identified two emerging outcome trajectories as the main mechanisms by which the program was contributing to all five areas/niches. This helped the AAS team in the Philippines to take immediate and more considered steps to strengthen the outcome trajectories and thus, potentially, the niches. In this way, they were able

Table 4. Differences in the Outcomes Identification and Classification Processes in Hubs.

Hub	Ways of Identifying Areas of Change, Outcomes, and Outcome Trajectories	Method of Verification
Southern Bangladesh Polder Zone	– Four areas of change identified by Aquatic Agricultural Systems (AAS) team – Fifty outcomes identified by change agents at community level in a workshop – Amalgamated into 16 outcome descriptions by AAS team – Four outcome trajectories identified in the workshop	External evaluator
Malaita—Solomon Islands	– Four areas of change identified by AAS team – Seventeen outcome descriptions identified through Most Significant Change (Dart & Davies, 2005) at community level complemented by other outcomes identified by AAS team – Five outcome trajectories identified in the workshop	Internal evaluator
Barotse—Zambia	– Four areas of change identified by AAS team – Seventy outcomes identified from learning reports produced by stakeholders and partner organizations – Four outcome trajectories identified in the workshop	External evaluator
Tonle Sap—Cambodia	– Three areas of change identified by AAS team – Nineteen outcome descriptions identified from learning reports from focal communities and then revised and verified by AAS team – Three outcome trajectories identified in the workshop	Internal and external evaluators
Visayas and Mindanao—Philippines	– Three areas of change identified by AAS team – Eighty outcomes identified in the workshop – Three outcome trajectories identified in the workshop, including 14 subtrajectories	Internal evaluators

to identify "beneficial coherence" (i.e., the outcome trajectory) and ways to stabilize and amplify it as Snowden (2010) suggests as a strategy for engaging in complex systems. They were also able to identify specific instances of outcomes identified in the outcome trajectories following Popper's (1992) logic. These outcomes were substantiated using the existing process information and triangulation. An internal evaluation team was able to start to build a case for the existence and program contribution to the trajectories, thus laying the groundwork for future impact evaluation that does not require control groups. Future cycles of outcome evidencing would have further interrogated the case by identifying and substantiating any causal links between the current outcome trajectories and broader impact on the lives of the program's target beneficiaries in due course.

What we learned from the pilot is that the method depends greatly on who and how the initial outcome "harvest" is carried out. The more change agents that can be involved the better. This produces large numbers of outcomes (>80) and potentially large numbers of outcome per cluster. It can be difficult to distill out the underlying outcome trajectory from a multicause diagram built from large numbers of outcomes and in practice some form of simplification is required. In the Philippines, this happened after a complicated diagram was drawn in the workshop. In Bangladesh, outcomes were grouped before clustering.

Outcome evidencing is an adaptation of outcome harvesting to meet a specific set of program requirements. The stepwise method we describe at the beginning is an ideal type constructed from learning from five pilots in five hubs. The adaptations to outcome harvesting include identifying areas of change to frame the collection of outcomes, the early involvement of change agents in identifying causal clusters of outcomes, the subsequent identification of the underlying causal pathway (called an outcome trajectory) using multicause diagramming, a specific step to look at

inclusion and winners and losers, subsequent evidencing of the outcome trajectories, and the use of results at the middle and the end rather than just at the end.

Outcome evidencing is still in its formative phase with some of its potential still to be proven. Whether it emerges as a new method in its own right or be seen as an adaptation of outcome harvesting remains to be seen. One claim that is yet unsubstantiated is that comparing outcome trajectories with existing program theory will allow the program to question its underlying causal premises and act accordingly. The outcome evidencing pilot did not lead to the AAS Program questioning its overarching program theory, although we do think it helped some key staff become clearer in the detail of that theory. For example, it helped the authors to identify "strategic niche management," and the idea that properly managed, such niches may act as as building blocks for broader societal changes, as a potentially useful addition to the AAS program theory. Questioning of the basic premises underlying the AAS program theory would have been more likely after a second round of outcome evidencing, which was originally planned. What the outcome evidencing pilot was able to do was to encourage staff and program participants to start thinking in terms of outcomes and outcome trajectories in the first place and to start to question hitherto implicit or unrecognized causal assumptions relating to their work on the ground. Staff and participants did act on immediate implications (Step 7).

Another area that needs further development is the identification of rival causal explanations for key outcomes, and the adjudication between them as an alternative to using control groups to making causal claims. There is little methodological advice in the literature as to how, and when to, construct and examine rival causal explanations. Most of the outcomes we looked at were intermediate in the sense they were given as evidence of an outcome trajectory that might plausibly lead to program goals. For these, our approach was to look for corroboration from more than one source and for the evaluator to provide some estimation of the strength of this "triangulation."

Conclusions

This article describes the early development of a rapid evaluation, complexity-aware approach called outcome evidencing, based on outcome harvesting. We developed the approach to meet learning and accountability requirements for an agricultural research program intervening in its geographic locations, which it understood to be complex systems. We made the adaptations because of a lack of an "off the shelf" approach that allows programs to regularly and critically review their program theory as a way of operating in complex systems. The approach identifies emerging clusters of outcomes, both expected and unexpected, happening within program areas of change. It then seeks to understand, describe, and verify these clusters as emerging causal pathways called outcome trajectories. The method is centered on a workshop in which change agents identify causal clusters and then uncover the underlying causality—the outcome trajectory—using a multicause diagram. The outcome trajectory is a theory of change that is subsequently substantiated. Comparing substantiated outcome trajectories against existing program theory allows the program to question its underlying causal premises. The method can be used for one-off evaluations that seek to answer questions about if, how, and in what contexts programs are working. However, it is likely to be most useful as a central part of a program monitoring, evaluation, and learning system. Repeated cycles of outcome evidencing build a case for program contribution over time that can be evaluated as part of any future impact assessment of the program or parts of it.

Declaration of Conflicting Interests

Funding

References

Aquatic Agricultural Systems. (2011). *CGIAR Research Program Aquatic Agricultural Systems: Program Proposal*. Penang, Malaysia. AAS-2012-07.

Aquatic Agricultural Systems. (2014). *Outcome Evidencing Report 2014*. VisMin Hub, Philippines. Retrieved May 2016, from https://goo.gl/vVwXC6

Agropolis International. (2015). *CGIAR Consortium headquarters in Montpellier*. Retrieved September 11, 2015, from www.agropolis.org/cooperation/headquarters-cgiar-consortium.php

Argyris, C. (1976). Single-loop and double-loop models in research on decision making. *Administrative Science Quarterly*, *21*, 363–375.

Argyris, C., & Schön, D. A. (1978). *Organizational learning: A theory of action perspective* (Vol. *173*). Reading, MA: Addison-Wesley.

Arkesteijn, M., van Mierlo, B., & Leeuwis, C. (2015). The need for reflexive evaluation approaches in development cooperation. *Evaluation*, *21*, 99–115.

Astbury, B., & Leeuw, F. L. (2010). Unpacking black boxes: Mechanisms and theory building in evaluation. *American Journal of Evaluation*, *31*, 363–381.

Britt, H., & Patsalides, M. (2013). *Complexity-aware monitoring*. Discussion Note, Monitoring and Evaluation Series, USAID, Washington, DC.

Brooks, S., & Loevinsohn, M. (2011). Shaping agricultural innovation systems responsive to food insecurity and climate change. *Natural Resources Forum*, *35*, 185–200.

Burge, S. (2013). *The Systems Thinking Tool Box*. Retrieved May 1, 2016, from http://www.burgehugheswalsh. co.uk/Uploaded/1/Documents/Multiple-Cause-Diagram-Tool-Box-v1.pdf

Collinson, M. P., & Tollens, E. (1994). The impact of the international agricultural centres: Measurement, quantification and interpretation. *Experimental Agriculture*, *30*, 395–419.

Dart, J., & Davies, R. (2003). A dialogical, story-based evaluation tool: The most significant change technique. *American Journal of Evaluation*, *24*, 137–155.

Douthwaite, B., Apgar, M., & Crissman, C. (2014). *Monitoring and evaluations strategy brief. CGIAR Research Program on Aquatic Agricultural Systems*. Penang, Malaysia. Program Brief: AAS-2014-04.

Douthwaite, B., Kuby, T., van de Fliert, E., & Schulz, S. (2003). Impact pathway evaluation: An approach for achieving and attributing impact in complex systems. *Agricultural Systems*, *78*, 243–265.

Dugan, P., Apgar, M., & Douthwaite, B. (2013). *Research in development: The Approach of AAS*. CGIAR Research Program on Aquatic Agricultural Systems. Penang, Malaysia (Working Paper). Retrieved October 3, 2016, from https://goo.gl/a99Zgk

Kemp, R., Schot, J., & Hoogma, R. (1998). Regime shifts to sustainability through processes of niche formation: The approach of strategic niche management. *Technology Analysis & Strategic Management*, *10*, 175–198.

Kennedy, P. (2009). *How to combine multiple research options: Practical Triangulation*. Retrieved July 26, 2016, from http://johnnyholland.org/2009/08/20/practical-triangulation

Klerkx, L., Van Mierlo, B., & Leeuwis, C. (2012). Evolution of systems approaches to agricultural innovation: Concepts, analysis and interventions. In I. Darnhofer, D. Gibbon, & B. Dedieu (Eds.), *Farming Systems Research into the 21st century: The new dynamic* (pp. 457–483). Netherlands: Springer.

Mayne, J. (2011). Contribution analysis: Addressing cause and effect. *Evaluating the Complex: Attribution, Contribution, and Beyond*, *18*, 53–96. Transaction.

Mayne, J., & Stern, E. (2013). *Impact evaluation of natural resource management research programs: A broader view* (ACIAR Impact Assessment Series Report No, 84). Canberra, Australia: Australian Center for International Agricultural Research..

Mohr, L. B. (1999). The qualitative method of impact analysis. *American Journal of Evaluation*, *20*, 69–84.

McNall, M., & Foster-Fishman, P. G. (2007). Methods of rapid evaluation, assessment, and appraisal. *American Journal of Evaluation*, *28*, 151–168.

Patton, M. Q. (2011). Developmental evaluation: Applying complexity concepts to enhance innovation and use. New York, NY: Guilford Press.

Pawson, R. (2013). *The science of evaluation* (Kindle ed.). London, England: Sage.

Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. London, England: Sage.

Popper, K. (1992). *The logic of scientific discovery*. London, England: Routledge.

Rogers, E. M. (2010). *Diffusion of innovations*. New York, NY: Simon and Schuster.

Rogers, P. J. (2008). Using programme theory to evaluate complicated and complex aspects of interventions. *Evaluation*, *14*, 29–48.

Schot, J., & Geels, F. W. (2008). Strategic niche management and sustainable innovation journeys: Theory, findings, research agenda, and policy. *Technology Analysis & Strategic Management*, *20*, 537–554.

Scriven, M. (1976). Maximizing the power of causal investigations: The modus operandi method. *Evaluation Studies Review Annual*, *1*, 101–118.

Scriven, M. (2008). A summative evaluation of RCT methodology: & An alternative approach to causal research. *Journal of Multidisciplinary Evaluation*, *5*, 11–24.

Snowden, D. (2010). Naturalizing sensemaking. In K. L. Mosier & U. M. Fischer (Eds.), *Informed by knowledge: Expert performance in complex situations* (pp. 223–234). New York, NY: Taylor & Francis.

Stame, N. (2004). Theory-based evaluation and types of complexity. *Evaluation*, *10*, 58–76.

Stern, E. (2015). Impact evaluation: A guide for commissioners and managers. *Bond*. Retrieved October 2015, from https://www.bond.org.uk/data/files/Impact_Evaluation_Guide_0515.pdf

van Mierlo, B., Arkesteijn, M., & Leeuwis, C. (2010). Enhancing the reflexivity of system innovation projects with system analyses. *American Journal of Evaluation*, *31*, 143–161.

Walters, C. (1986). *Adaptive management of renewable resources*. New York, NY: Macmillan.

Westhorp, G. (2014). *Realist evaluation: An introduction*. Methods Lab-ODI, London.

White, H., & Phillips, D. (2012). *Addressing attribution of cause and effect in small n impact evaluations: Towards an integrated framework* (Working Paper 15). New Delhi, India: International Initiative for Impact Evaluation.

Wilson-Grau, R., & Britt, H. (2012). *Outcome harvesting*. Cairo, Egypt: Ford Foundation. Retrieved October 20, 2012, from http://www.outcomemapping.ca/resource/outcome-harvesting