

How to find out what's really going on: understanding impact through participatory process evaluation

Article (Accepted Version)

Cornwall, Andrea and Aghajanian, Alia (2017) How to find out what's really going on: understanding impact through participatory process evaluation. *World Development*, 99. pp. 173-185. ISSN 0305-750X

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/69768/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

Copyright and reuse:

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Accepted by World Development, published July 2017

How to Find out What's Really Going On: Understanding Impact through Participatory Process Evaluation

Andrea Cornwall

Introduction

So-called 'gold standard' experimental and quasi-experimental approaches to impact evaluation have become increasingly prevalent in international development in the years following the rise of the results agenda (Vaessen 2010) and the influential report from the 'Evidence Gap Working Group' convened by the Washington-based Center for Global Development, entitled *When Will We Ever Learn: Improving Lives through Impact Evaluation* (2006). Harrison (2015) notes that of the 2500 impact evaluation studies in the database of the leading development impact evaluation organisation established in the wake of the CGD report, 3ie, less than 5% used a primarily qualitative approach. Indeed, studies using non-experimental methods accounted for less than a fifth of those commissioned by the World Bank's Development Impact Evaluation unit. For all that the principal proponents of these approaches acknowledge the value and importance of a mix of methods, one that includes qualitative methods (Banerjee and Duflo 2009; White 2009), 'rigour' has become a euphemism for 'quantitative': what counts is that which can be counted.

The debate on the potentials and limits of quantitative - experimental and quasi-experimental - approaches to impact evaluation takes us back to some basic questions about what kind of knowledge best serves the purposes of those who fund development initiatives. If such interventions are to stand any chance of being sustained, inspiring others to follow suit and being successful over the longer term, it is arguably important to understand *how* they work rather than just whether they have achieved those results that they set out to achieve and that are amenable to measurement (Cohen and Easterly 2009; Stern et al. 2009; White 2001).

This is not a new issue. Writing in the late 1990s, Pawson and Tilley (1997) contend project evaluation over the previous three decades in sociology, criminology and education has been largely ineffective precisely because it focused on *whether* projects worked, rather than *why* they work. As Thorbecke (2007) points out, a primary weaknesses of quantitative impact evaluation is that it can only answer the question of whether an intervention has an effect; Deaton (2009) suggests that even this question can only be answered effectively in particular, idealized, circumstances. And Jones et al. (2009) caution against 'proof' of ultimately spurious causalities. These shortcomings are well recognised within economics (Ravallion 2008; Banerjee and Duflo 2009; Deaton 2009), as well as beyond it (Heckman 1991; Pawson and Tilley 1997). This recognition has spurred attempts to address the *why* question in impact evaluation (Masset and White 2007).

Understanding *how* and *why* an intervention works involves more, however, than assessing whether and to what extent the intended, designed, impact of the intervention happened, even with the best laid Theory of Change at the outset. The intervention may end up having a range of other effects that contribute – positively as well as negatively – to outcomes. These effects may constitute discernible impacts in themselves. Bhola draws a useful distinction between three types of impact: ‘impact by design, impact by interaction and impact by emergence’ (2000:163). Conventional evaluation and impact assessment more generally, and quantitative impact evaluation in particular, tends to be restricted to ‘impact by design’. Yet there may be significant complications to attribution and causality that arise from ‘impact by interaction’, described by Bhola as resulting from ‘outcomes of an original intervention interacting with other concurrent interventions made by other agents and agencies, and thereby enhancing or inhibiting effects of the original intervention’ (2000:163-4).

The third category, ‘impact by emergence’, is especially vulnerable to being missed by evaluation and impact assessment approaches that concern themselves only with measuring intended and presumed effects, rather than with a more inductive approach to tracing change. Any methodology that is not sensitive to ‘impact by emergence’ may be missing key factors for success and sustainability (Bhola, 2000; Roche 2000). Bhola hints at the complexity of assessing these impacts, defining ‘impact by emergence’ as:

...unimagined outcomes emerging from the original intervention through its interactions with other interventions and its interfaces with historical and cultural processes in place but not easily discernable. (2000:161)

In this article, we explore Bhola's 'impact by interaction' and 'impact by emergence' through a case study of an evaluation of a nutrition intervention in which a fuller than usual range of assessment methods were used, including data from quantitative measurement of change in nutritional status. My focus is on one of the more unusual in the suite of evaluation approaches: participatory process evaluation. Process evaluation explores how implementation unfolds, gathering experiential accounts of what went on, what happened that was unexpected and how those involved reacted to those unexpected blights or opportunities (Moore et al. 2015).

Participatory process evaluation, we suggest here, provides a valuable complement to other approaches to impact assessment precisely because it brings into view all three kinds of impact identified by Bhola, along with insights into how change has come about and how that change might be sustained. It does so by using participatory methods to get to grips with the life of an intervention as it is lived, perceived and experienced by different kinds of people, including program or project personnel (Chambers 1997; Chambers and Mayoux 2005; Guijt 2008a and b; Leeuw and Vaessen 2009).

Participatory Process Evaluation: Principles and Practice

I tried and failed many times to explain to our donor organizations why processes had an importance beyond the results they achieved. The results-based framework within which we operated existed in the context of complex power relationships... Sometimes we found ourselves talking openly and finding support from among the donors, while at other times we had to conceal our true objectives and ensure that the results-based, logical framework outputs were achieved... We found ourselves adopting a language and a set of tools – technical activity reports, expenditure reports and products – quite distinct from the work we were actually doing... (León, 2010, cited in Eyben 2013)ⁱ

Mainstream international development's concern with ensuring that aid is well spent has led to the adoption of tools to produce what Rosalind Eyben (2013:7) calls 'results artefacts' and 'evidence artefacts'. The former consist of 'reporting, tracking and disbursement mechanisms' that include base-line data, performance measurement indicators, reports on results and logical framework analysis. 'Evidence artefacts' include randomized control trials (RCTs), systematic reviews, cost-effectiveness analysis, business cases and impact evaluation. Eyben observes:

According to 3ie - an organisation created by donor agencies in 2008 to enhance development effectiveness through the promotion of evidence-based policy making - high-quality impact evaluations that measure the net change in outcomes amongst a particular group, or groups, of people that can be attributed to a specific program. They have narrowed the

debate from the older OECD definition of 'impact' which is 'the positive and negative changes produced by a development intervention, directly or indirectly, intended or unintended'. Hence, increasingly impact evaluation (IE) tends to be associated with experimental and quasi-experimental methods because of their 'objectivity'. (2013:20)

This kind of narrowing of gaze is an example of the reductionism that characterized impact assessment until recent attempts to grapple with these limitations. One of the criticisms leveled at conventional quantitative impact assessment is that ultimately, all that counts as 'evidence' is that which can be counted (Eyben et al. 2015), even though measurable net outcomes, however they are counted, often tell us very little about *how* change happens. Indeed, Howard White, former executive director of 3ie, states 'measurement is not evaluation' (2011: 132). Angus Deaton argues:

RCTs of 'what works', even when done without error or contamination, are unlikely to be helpful for policy unless they tell us something about why it works, something to which they are often neither targeted nor well-suited. (2009:44)

What makes RCTs or quasi-experimental methods such as difference in difference analysis, propensity score matching or regression discontinuity analysis useful is that they are able to address the highly targeted question of 'does this intervention work in this place, with these people, at this time?'. From this, inferences might be drawn about whether an intervention would work

elsewhere, with other people, at another time. The rise in popularity of these methods has been associated with a concern about selection bias, through which project results end up being over-estimated where differences between groups are attributed entirely to the effects of the intervention when they may be due to already existing differences.ⁱⁱ It is often difficult, however, to draw conclusions from a highly controlled environment and yield useful policy recommendations that would be applicable in a different setting. As Deaton (2009) points out, another drawback of these approaches arises in the extent to which they can handle heterogeneity, as analysis is not always disaggregated enough to understand how interventions work differently for different people. Stern et al. (2012: 8) comment:

Jonas et al. (2009) observes that RCTs can answer ‘does it work?’ questions but not ‘how could it be improved?’. Cartwright and Munro (2010: 265) note that experiments can answer the question ‘did it work here?’ but not ‘will it work for us (elsewhere)?’

It should be noted that more recent quantitative impact evaluations place more of an emphasis on uncovering “mechanisms”, i.e. the channels that the intervention has passed through in order to make a change. In the case of an RCT, this can be achieved through different randomised treatment arms, which randomize receiving different types of interventions. For example, through a randomised factorial design, Dupas (2006) finds that engaging teenagers with information on HIV/AIDS prevalence is much more effective at reducing risky sexual behaviour than just training school teachers – the main component of the program under study. However, unless accounted for in the design of an RCT, it

is quite difficult to uncover a mechanism during the analysis stage, making it quite difficult to evaluate unintended outcomes.

Quasi-experimental methods on the other hand, allow for more leniency in the analysis of mechanisms, as selection bias is controlled for in other ways. For example, Galiani, Gertler and Schragrodsky (2005) find that privatisation of water services reduces child mortality, and they then further decompose this result to find that the privatisation of water services is related to a reduction in deaths caused by water borne diseases, rather than those related to water conditions, such as accidents or congenital diseases. This is further corroborated by the fact that the reduction in child mortality is greatest in poorer areas. In addition, a few quantitative impact evaluation studies have paid attention to the heterogeneity of the intervention effect by the type of program participant. This is done by splitting the sample into different groups of relevance and comparing the effect size, which requires large datasets. For example, an evaluation of a nutrition program in Madagascar found that the effects of the program were much larger for children whose mothers were literate, educated and lived in villages with better infrastructure (Galasso and Umapathi 2009).

Conventional quantitative impact evaluation, however, provides us with few insights into how positive change can be *sustained* (White 2011), or indeed, into the temporal dimensions of impact, as Michael Woolcock notes:

Understanding the efficacy of development projects requires not only a plausible counterfactual, but an appropriate match between the shape of

impact trajectory over time and the deployment of a corresponding array of research tools capable of empirically discerning such a trajectory. At present, however, the development community knows very little, other than by implicit assumption, about the expected shape of the impact trajectory from *any* given sector or project type, and as such is prone to routinely making attribution errors. Randomisation per se does not solve this problem. (2009:2)

Woolcock directs us to looking more carefully at impact trajectories. He is roundly dismissive of the linearity implicit in prevailing experimental approaches to impact evaluation:

It is only the most ad hoc theorising or wishful thinking (or the overriding imperatives of domestic political cycles and the structure of career paths at development organisations) that could possibly substantiate an assumption that *all* project impacts are linear and monotonic. (2009:3)

Instead, he notes that despite the recognition that impacts may vary in different settings and with different groups of people, surprisingly little attention is paid to temporal variability.

What, then, might help answer these *how, why, when* and *where* questions?

Mixed methods are often advocated as a way forward. White (2009), for example, includes a repertoire of qualitative and participatory methods in his reflections on approaches to theory-based impact evaluation and other writings (see, for example, Carvalho and White 1997, White 2011). White and Masset (2007) use a theory-based evaluation to explore why a nutrition program in

Bangladesh does not work, mapping out all of the assumptions, inputs and dependencies and measuring elements of these. The authors use propensity score matching and find that while community-based counselling did affect women's knowledge of nutrition, it did not affect their practice. The reason for this was that women did not have enough time or resources to put the knowledge into practice. Perhaps most importantly, the mother-in-laws were in control of household food expenditures, and given that they were not sufficiently involved in the project, a significant impact was not observed. A 2011 World Bank guide to impact evaluation (Gertler et al. 2011) notes, 'impact evaluations conducted in isolation from other sources of information are vulnerable both technically and in terms of their potential effectiveness' (2011:15). Gertler et al. emphasise the need for complementary qualitative research that can 'examine *questions of process* that are critical to informing and interpreting the results from impact evaluations' (our emphasis, 2011:15).

Going back to the OECD definition cited by Eyben (2013), if 'impact' is taken to refer to 'the positive and negative changes produced by a development intervention, directly or indirectly, intended or unintended', an impact assessment methodology needs to be able to get behind the outcomes of an intervention to understand the nature of those changes. As well as reflecting on how intended and direct changes actually play out on the ground, such a methodology would be most useful if it could also identify changes that are *indirect* and *unintended*, as these offer often important insights into issues like sustainability, replicability and the potential for diffusion. And if such a methodology could also enable us to understand in some depth the patterns of

change over time that are giving rise to sought-after or unexpected outcomes, it might be the means for addressing the challenge of making impact evaluation properly rigorous. These are dimensions of impact assessment that experimental methods do not and indeed *cannot* adequately address (Ravallion 2009; Cartwright and Munro 2010).

Participatory process evaluation takes as its starting point an interest in precisely these *how, why, when* and *where* questions, and is explicitly concerned with mapping 'impact trajectories' over time in relation to different dimensions of an intervention and its outcomes. It is built on the recognition that perceptions of progress and success vary over time and between different stakeholders. It is premised on the understanding that at the outset, expectations of what the intervention could or should achieve may be quite different from how things end up once implementation is underway, and that along the way, views may change as a result of the process, and stakeholders may end up in quite a different place than they originally envisaged – and with a different understanding of what 'success' means.

Rather than simply examine the results of an intervention, participatory process evaluation seeks to contextualise the process of intervening in relationships and institutions, looking at shifting institutional frameworks and relational dynamics. Temporal narratives generated through the use of visualisation methods aimed at facilitating story telling provide opportunities to explore critical moments: crises, breakthroughs, conflict, gradual and sudden change. Analysis of the past and of different experiences of transition to the present open up opportunities to reflect collectively on desired changes for the future, working from the

perspectives of different actors in the process to facilitate a dialogue on futures possible. By seeking to understand how change happens from diverse perspectives, participatory process evaluation can generate a rich picture of the preconditions for positive change, as well as a deeper understanding of obstacles or barriers.

This approach looks for change not only in the places towards which deliberate efforts to bring about change have been directed, but also in the positive changes people report experiencing. Rather than measuring progress against a baseline according to a stock of indicators, this means understanding more inductively what has changed, how and why. Participatory process evaluation thus brings into view unexpected and unpredicted changes, as well as providing a basis for reflecting on why and how change happened - or did not. It provides a way of getting to grips with all three of Bhola's dimensions of impact assessment. As such, it is a valuable complement to quantitative methods of impact evaluation in general and RCTs in particular.

The account that follows draws on a case study of participatory process evaluation. Evidence consists of primary ethnographic data gathered by the first author in the form of transcripts of interviews and group discussions, pictures generated in the process of conducting those interviews and discussions and a field diary of observations and reported speech from informal interactions. Seeing an intervention through the lens of participatory process evaluation brings into view its internal dynamics and, with this, issues of some sensitivity: details of places and people are obscured accordingly.

Listening to Stories of Change

Two years into an integrated government nutrition program to address pervasive malnutrition and worrying levels of stunting in a rural community in central Kenya, the program's funders – a Nordic donor – decided it was time to take stock. A baseline survey had been carried out. This had been a bit of a catastrophe for the program, as it had, as we go on to narrate, alienated the local community. But partly as a result of this set-back, the program had employed a range of means of engaging people in reflecting on the nutritional status of children in the community, and in analysis and planning of ways in which the situation might be addressed. Social marketing had been used to engage community members with some hard facts – that a significant number of children were failing to thrive, that there was a serious stunting issue and that there was widespread malnourishment. Participatory methods had been used to engage community members in analyzing the situation that had led to such poor nutritional outcomes for their children, further raising awareness.

A cluster of interventions had been developed that included educating families on spotting the signs of malnutrition and preventive actions such as hand washing, dish racks and latrines, and making the leafy green vegetables that had come to be regarded as the food of the poorest into a desirable food for all.

Outreach from a feeding centre and adult education in organic farming and income-generating skills was complemented with community-based participation activities and nutrition and hygiene education work with schools.

Small business, farming and health projects had been established as part of a holistic, multi-sector approach. Methods used for training and teaching adults

and children ranged from more didactic approaches, to the use of Participatory Educational Theatre, Child-to-Child and Participatory Rural Appraisal (PRA) to visualize the problem, examine the causes, publicise the consequences and create a space for public engagement in finding solutions to malnutrition and stunting.

At the time when the process evaluation was commissioned, two years into the initiative, monitoring of nutritional status was already showing results. An impact evaluation some three years later would demonstrate the success of the program in bringing about significant changes in nutritional status amongst children aged 12-60 months. This study, using experimental methods, showed statistically significant reductions in the levels of underweight children and stunting (Havermann et al. 2012). It didn't take a sophisticated anthropometric study, however, to demonstrate some of the more tangible results of the program. Any visitor to the area could observe gardens with leafy green vegetables, dish racks and latrines. If that visitor was an outsider and was taken to a school, they would be treated to displays of singing featuring hand-washing and other health promotion activities that could reduce the incidence of diarrheal disease. The old feeding centre used in the past to rehabilitate starving children had turned into a lively hub of activity supporting organic farming, income-generating projects and a network of community participation resource persons.

Things were clearly working. But how did the program achieve those results? How did it actually *work*? What made it tick, what made it successful? What was

there to be learnt about how things were done that might be useful for other places, other problems? What explained the program's success in enlisting community participation and engagement from local government? How sustainable was the injection of intensive effort accompanying the program? And what wider lessons might be learnt from the ways of working developed by the program, including their use of theatre and participatory visualization techniques? For this, the donor felt that what was needed was a process evaluation, to accompany a more conventional program review. And since participatory methodologies had been a powerful part of the learning journey in program implementation, the decision was made to commission a participatory process evaluation. A team was brought together consisting of two Kenyan participation experts, Charity Kibutha and John Gashigi, and two British consultants with experience in the use of participatory approaches, Tilly Sellers and myself.

Over the period of about ten days, we worked intensively together to develop and apply a methodology that could get to grips with the dynamics of change in all its complexity. We were especially interested in the interactive and relational dimensions of the program's ways of working and impact. Drawing on our accumulated collective experience of using participatory methods in evaluations of various kinds, we invented a sequence of activities that adapted existing methods to explore the questions in which we were interested. In what follows, we narrate the process of applying the methodology and reflect on what we learnt from this experience. We go on to explore the broader implications of

using participatory process evaluation in impact assessment to reach the parts that experimental impact evaluation approaches are largely unable to access.

Understanding “Success”: Designing and Applying a Methodology for Participatory Process Evaluation

It didn't take long after our arrival for us to show the program management team quite how different our approach to evaluation was going to be from a more conventional review. After a briefing by the program leader, who gave us some background on the nutrition situation in the area and the kinds of interventions that had been pursued to address it, we were informed of the fieldsites that had been selected for us to visit. Half an hour under a tree in the yard with an extension worker later that afternoon was all it took to elicit a comprehensive matrix ranking of sites, using criteria for success generated by the extension worker. On this matrix, three of the sites offered as a “range” of examples were clustered at the very top of the list, and one at the very bottom. A quick cross-check with a variety of program staff yielded confirmation of this assessment, along with some surprise at how we'd come about this information so quickly.

That we had been offered this pick of locations was to be expected. Evaluators are often shown the showcase, and having a basket case thrown in there for good measure allows success stories to shine more brilliantly. Even though there was no doubt in anyone's minds that this was an exceptionally successful program, our team had been appointed by the donor responsible for funding. Those responsible for the program were taking no chances. We were greeted with a

rather bewildered look of surprise by the program manager when we came up with our own list of site visits, selected from various parts of the ranked list, with our suggestion that we begin with these rather than going along with the ones we were originally scheduled to visit.

Instead of going straight to visit the fieldsites, we first spent more time at headquarters. We were interested in the perspectives of those involved with implementation, from managers to front-line field-staff. We wanted to know not only what the project had achieved, but also what people thought it *might* achieve when they first began, what they *hadn't expected* to happen and what had *surprised* them (cf. Guijt 2008a). Asking these kinds of questions gave us vital clues into how the program was working. It also allowed us to understand the institutional dynamics of the program in ways that simply could not have surfaced if we had only been interested in results. Rather than reduce the program to a set of outcome indicators that could be quantified, we actively sought people's versions of what had happened and of how the changes that mattered to them had come about. These 'stories of change' offered us a more robust, rigorous and reliable source of evidence than the single stories that conventional quantitative impact evaluations generate. And they also allowed us to explore a richer picture of what success might mean to different kinds of people, one that would provide opportunities along the way for reflection that could spur further action.

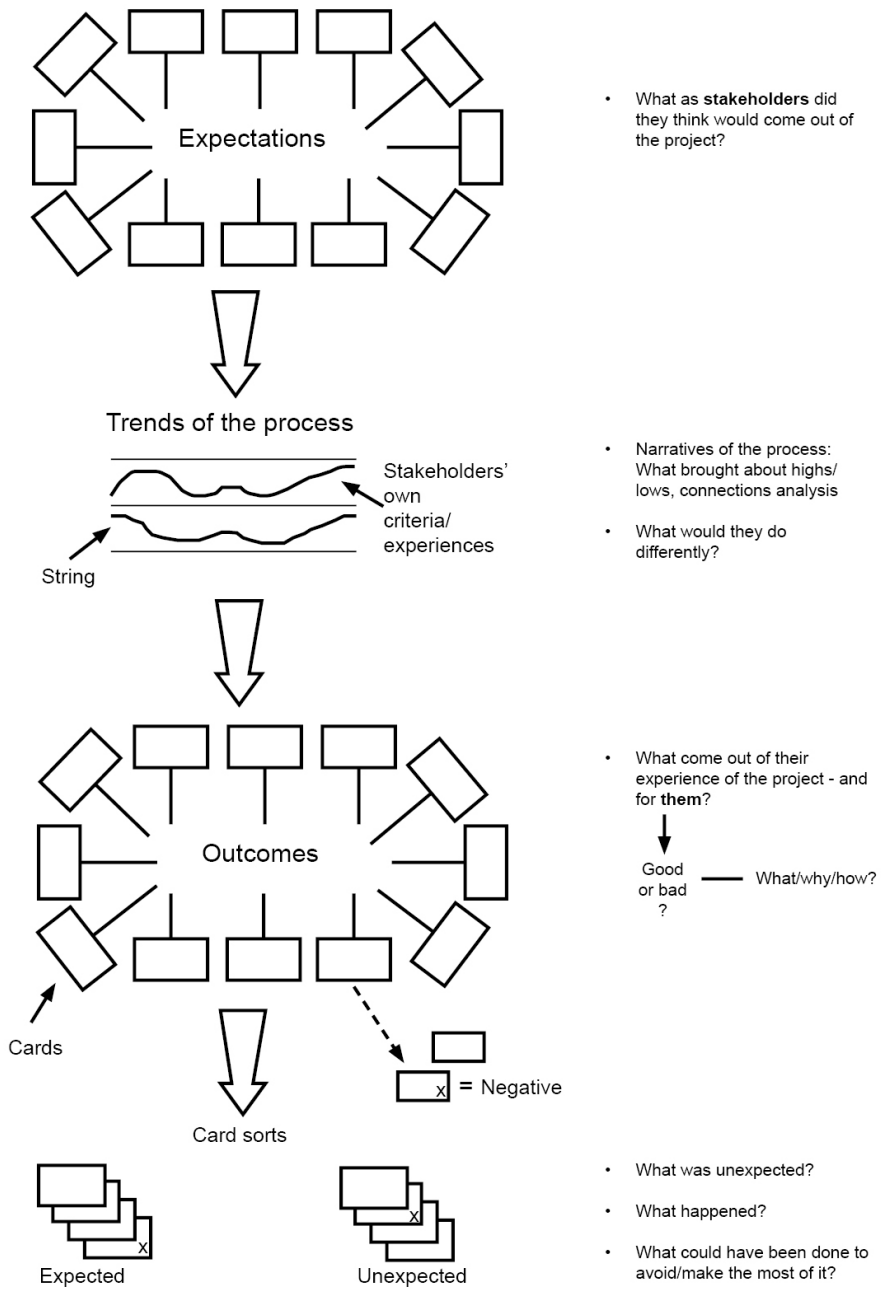
Developing a process methodology

Our methodology consisted of three basic parts. The first involved a stakeholder analysis to get a fuller picture of who was involved in the program. We were interested in hearing the perspectives not just of program beneficiaries, but of everyone with a role in the design, management and implementation of activities, from officials in the capital to teachers in local schools. This consisted of sessions in which we used coloured cards to generate a visual representation of everyone involved in the conception, design and delivery of the program. We also used these sessions to analyse perceptions of the relative importance and relationships between these stakeholders, using an adapted Venn diagram method for institutional and relationship mapping. On this basis, we made an initial selection of people and groups to interview.

The next step involved a sequence devised to capture expectations and experiences over the life of the program, using a packet of coloured cards, a large piece of paper and pieces of string. It began with an open-ended question about what the person or group had expected to come out of the program. Each point was written on a coloured card, one point per card. Extensive open-ended prompting was used to elicit as many expectations as possible. Surrounded by cards, we then encouraged people to cluster them into themes, and name them, writing these themes on another set of cards. We then took those theme cards – labelled with words like ‘trust’, ‘confidence’, ‘resistance’ and so on – and moved to the next step of the analysis. In this step, we looked at trajectories over time. This was done using pieces of string to form a graphical representation, with the two-year time span on the x axis and points between two horizontal lines representing the highest and lowest points for each category on the y axis. We

were interested in the highs and lows of steady or rapid improvement or decline, and where things had stayed the same. We encouraged people to use this diagram as a way of telling the story of the program, probing for more detail where a positive or negative shift was reported. Fascinating details emerged. We then probed further to reveal dynamics, strategies, successes and crises that might otherwise not have come into view.

Figure 1: Methodological Design



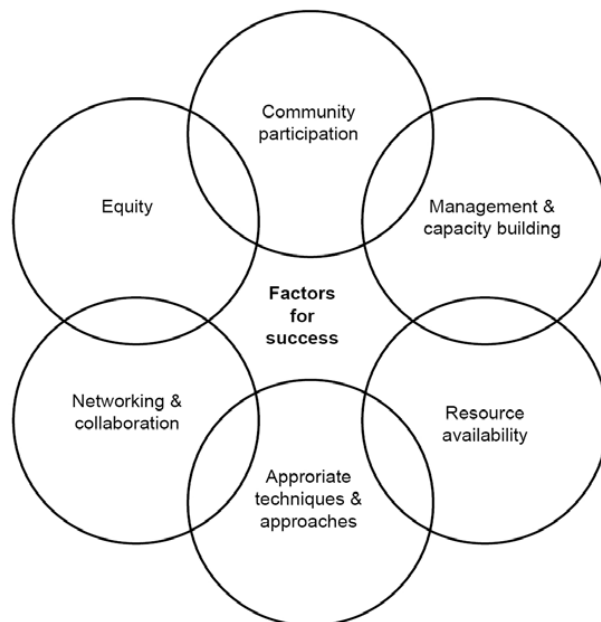
The third step was to use this data as a springboard to analyse and reflect on what had come out of the project. We did this by probing for positive and negative outcomes. These were written onto cards and sorted into two piles. We then reflected on what could have been done to avoid or make more of

unexpected outcomes. We also talked through gaps between people's expectations at the outset and what had actually happened. We kept people focused on their own experience rather than engaging in a more generalized assessment of the program, prompting them to think about what went well and what might be done differently if the program were to be run again in another place.

We complemented this sequence with contextual background research using unstructured interviews and Participatory Rural Appraisal methods with children and adults in villages in the area. We used social mapping to explore who was involved in the program and to map institutions. This opened up discussion on how the program was working at the village level, who was participating in it and who was perceived to benefit from it. Venn diagrams representing the period prior to the project and the present served as entry points for discussions about processes of institutional change, charting the dynamics between institutions over the program period. Network diagramming explored connections between individuals and institutions with some relationship to the program, in terms of communication, support, points of tension, competition and dysfunction. Time trends of community participation were used to generate discussion on how people have been involved, what community participation means to people, and to explore people's experiences of barriers, obstacles, conflicts and successes since the program began. Group discussions generated shared criteria representing what villagers felt to be vital to securing improvements in nutrition. These were then applied to ten villages: each criterion was scored and an overall score given.

From this we generated a diagrammatic representation of factors for success that was subsequently used to discuss and develop process indicators against which the program could be monitored and evaluated. We then used this as the basis from which to extend the analysis beyond analysing the change that had happened, to questions of sustainability and replication.

Figure 2: Factors for Success



Engaging Analysis

Rather than simply gathering and analysing the data ourselves, we enlisted the participation of different stakeholders in the process of sense-making of the emerging picture. This included the external team of consultants who had been contracted to carry out a conventional project review, who were there at the same time as us. Our paths crossed a couple of times. The first time, we

exchanged stories of what we'd been doing: theirs of being driven to meetings at chosen sites to meet the community and sitting politely through hours of singing and speeches, ours of sitting in people's homes or talking to them in their fields, hanging out with groups of children and government or project staff. The second encounter was close to the end of the review period, when they came to find out what we'd learnt. With the simple charts and diagrams we'd produced, retelling the story was swift – and safe for those who shared their confidences, the diagrams providing a cloak of anonymity and our synthesis of headlines the kind of discretion that allowed us to surface emerging challenges without revealing our sources. It was interesting to observe just how much more information our enquiry had revealed about the program than a conventional outputs-to-purpose style review.

Our way of working seemed to have left a particular impression with the district officials we spent time with. They were not often asked their views by the stream of outside visitors to the program, they told us, let alone invited to talk about their experiences and feelings. On our last night in the community, we were invited to have a drink with some of these officers. "We've never had visitors coming here who knew so much," one said to us. Another confided that it's easy enough to direct the usual kind of visitor towards the story that the program team wanted them to hear. Development tourists, after all, stay such a short time: "they're in such a rush, they go to a village and say they must leave for Nairobi by 3pm and they [the program staff] take them to all the best villages."

Apocryphal Tales and Lessons Learnt

Reflecting on the impact trajectory of the program from the distinctive vantage points of different stakeholders offered us fascinating insights. One of the most interesting twists in the tale, in many ways apocryphal, was an experience that took place right at the beginning of the program. As we note earlier, the program management team hit some problems when they sought to establish a baseline. They did what is so often done, and sent out a team of enumerators to gather data from a random sample of households. In this case, however, the enumerators did not just have questions. They had specially designed kits with which to gather basic anthropometric data such as height for weight and upper arm circumference measurements. At the time, a rumor was sweeping the area about a cult of devil-worshippers seeking children to sacrifice. The news spread like wildfire that the measurement kits used by the enumerators were designed to measure children for coffins. Families greeted the enumerators with hostility. In one place, they were chased with stones. To get the program off the ground again, the intercession of the area's chiefs was needed. This set in train efforts to re-engage alienated communities, and led the program implementers to the use of participatory methods, social marketing and an overall approach that emphasized engagement.

What was so striking about this story was that it simply had not occurred to the researchers that entering communities to measure small children might be perceived as problematic. The enumerators had done all the right things. They'd explained the purpose of the research and sought consent. They had their interviewing protocols. They had official documentation. But none of this

mattered. The incommensurability of knowledge systems, the lack of trust of outsiders and the failure to adequately countenance what was needed to establish the basis for informed consent provoke questions participatory researchers have raised for decades about the nature of 'extractive' research and the politics of 'expert' knowledge (Gaventa 1993, Tandon 1996). As one community member shared with us, "when X [the feeding centre people] came, many didn't like it because *they didn't know what was going to happen*".

The tale of the baseline survey offers other lessons. Conventional research is rarely transparent in any way. It may not be at all evident to anyone why certain questions are being asked, why certain people have been chosen to answer them, what's going to happen to their answers and what the knock-on effects of taking part might be. It is quite reasonable, under these circumstances, to do as Wolf Bleek (1987) reports in his classic article 'Lying Informants': pull the wool over the enumerator's eyes. Participatory methods offer a degree of transparency and engagement that standard survey methods lack. This is not to say that the versions that emerge from the use of participatory methods are, like the answers to enumerators' questions, not also geared to what people think outsiders might wish to hear (Cornwall 1996). We were ourselves well aware of the kinds of biases that can prevail where practitioners of participatory approaches take versions crafted to meet the gaze of an outsider at face value, and of how our presence might be interpreted and received. But the very public nature of these processes and the kind of explanations of purpose that take place, as well as the opportunities for engagement in analysis that they present, does change the dynamic. Triangulation of methods and informants, cross-checking and

deliberation as part of this very public process all work to enhance the trustworthiness of knowledge produced using participatory methods, as judged according to well-established criteria for judging the quality of qualitative research (Chambers 1997; Pretty 2005).

What had come out of the program's participatory community-level work from the social marketing done after the baseline experience was striking. There was very little awareness amongst community members of quite how bad the nutrition situation was. The shock with which the data was received spurred further reflection amongst those who lived in these communities on what had gone so very wrong for so many children to be suffering in this way. This served as a powerful prompt for actions that could address a problem that was now shared. By using participatory diagnostic methods to uncover problems and generate solutions, then using participatory educational methods – Participatory Educational Theatre, Child-to-Child – to engage people, the program was able to effectively use participatory research as pedagogy. Community-led growth monitoring and a community evaluation process gave community members the opportunity to learn about other tools they could put to work in assessing progress.

For the community members involved in this process, used to donors coming in with material assistance rather than with participatory processes that relied on them coming up with their own solutions, it took a while for it to sink in that this broader problem analysis and work on systemic solutions could address the problems the community were experiencing. As one villager shared with us:

When we started with the activities, I thought what has this got to do with malnutrition – there is still no food. But later I came to realise that there are lots of causes of malnutrition like disease, water and not eating the right foods.

“We learnt many things we didn’t know,” we were told in one village. In another, a village resource person told us of the surprises this had brought, “we didn’t expect we would learn this much or there was so much that we didn’t know.” He went on, “the spread of information [now] is such that it penetrates every part, from one to the next person, there is no way we can go back now, just forward.” As another program worker recalled, “in some cases, we felt almost left out... it was so exciting to see them take over. The literacy level is not the problem, they’re able to understand and use the information themselves.”

Discussions with key co-ordination personnel in government, to whom the management of the program was due to devolve once the intermediary implementing agency withdrew, gave us further insights. A meta-analysis of the cards generated from the first step of the time-line exercise yielded six factors that were most significant in shaping their expectations and experience of the program: their sense of ownership of the program; how the community viewed them; the level of confidence they felt they had in the work; the level of demoralization associated with the program; the level of their own participation and engagement with it; the level of threat that they felt the program posed to them, as an externally induced activity that had the potential to both undermine

them and step on their toes. As can be seen from this list, this is not at all the conventional stuff of an evaluation that simply looks at what has been achieved in terms of outcomes 'on the ground'. Yet in terms of the continuity of the program, the story they told us was both important and interesting.

At the beginning, the government personnel said, they felt very little sense of ownership of the program. It took almost a year before they began to feel involved, and then things got stuck: two years down the line, things had not moved much further, and there were issues of access and control over the interventions that were clearly still fairly sticky. But there had been a dramatic change in how they felt the community viewed them, as government officials: from a low at the onset of the program, the line representing the confidence the community had in them on the graph shot up quickly and continued at a very high level. This had a knock-on effect on their morale. They felt more confident in their work, and their level of demoralization, which was very high at the outset, went right down, in incremental bounds. Riding on all of this was their own participation in the program. This rose exponentially and plateaued at a high level by the second year. And lastly, in inverse proportion to this, was the level of threat they felt, which began at the highest of levels and had all but disappeared just over a year into the program (see Figure 3).

Their colleagues at the divisional level had an equally positive story to tell, (Figure 4), although one that was tinged with similar frustrations to those experienced by community volunteers and resource people. Their themes were: negative or positive attitude to the project; morale; resources available to them;

standard of living (theirs). At the outset, they were pretty hostile to the program. It did not seem to be offering them much in the way of resources, especially for fuel for their vehicles to enable them to go out and about in the communities. They were concerned about rising expectations of them. This negativity fell steadily over the course of the program, changing direction in the last few months to a general positive sense of what the program offered them. Yet while at the start, their morale had risen as their negativity had dipped, this hadn't been sustained: the same ownership and control issues arose and the same resource issues that were raised by those working in the community also hit them. The resources graph registered a small rise and then a fall, representing allowances they were initially paid and the dissipating hope that further resources for their work would be available. Lastly, their standard of living registered an incremental, but small rise – far less than they had initially expected.

Figure 3: Key Central Government Co-ordination Personnel

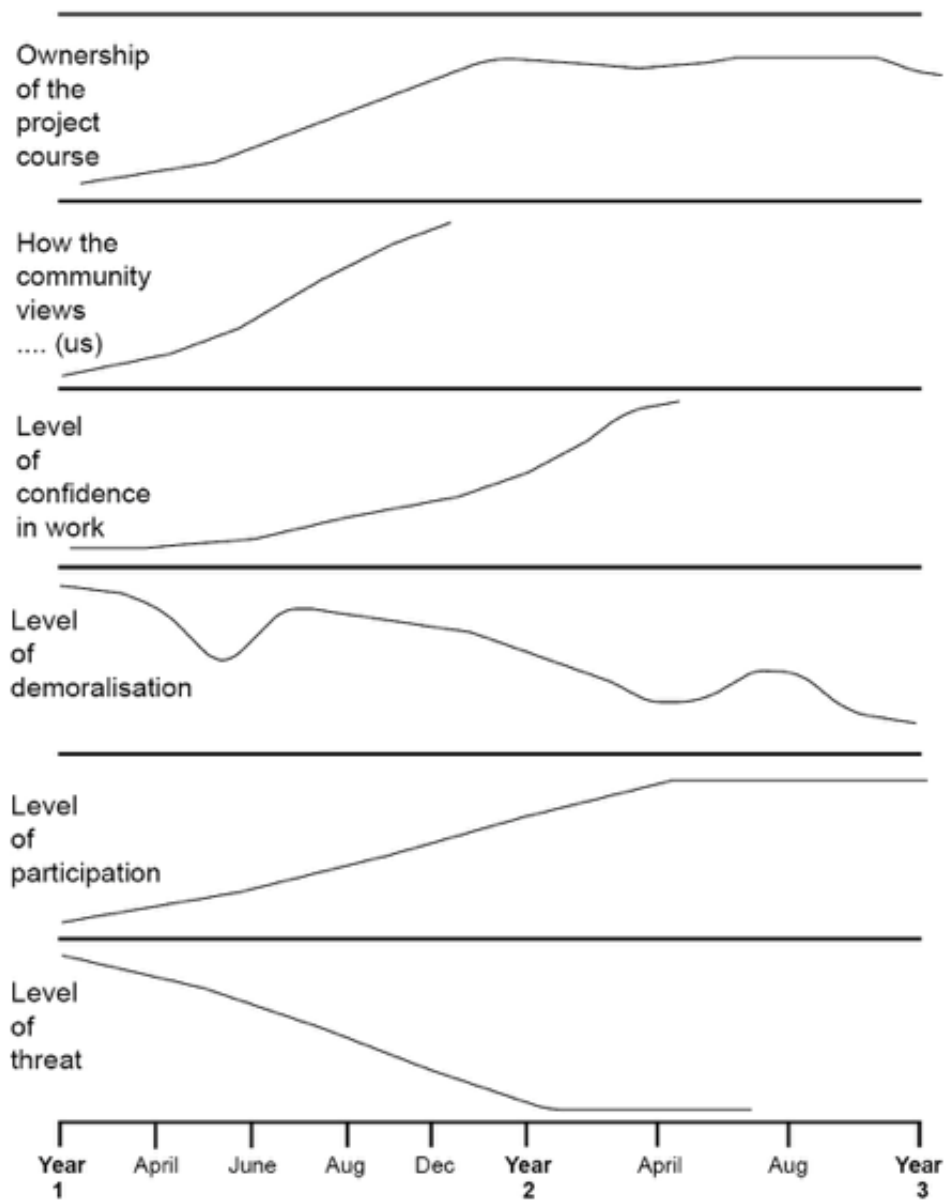
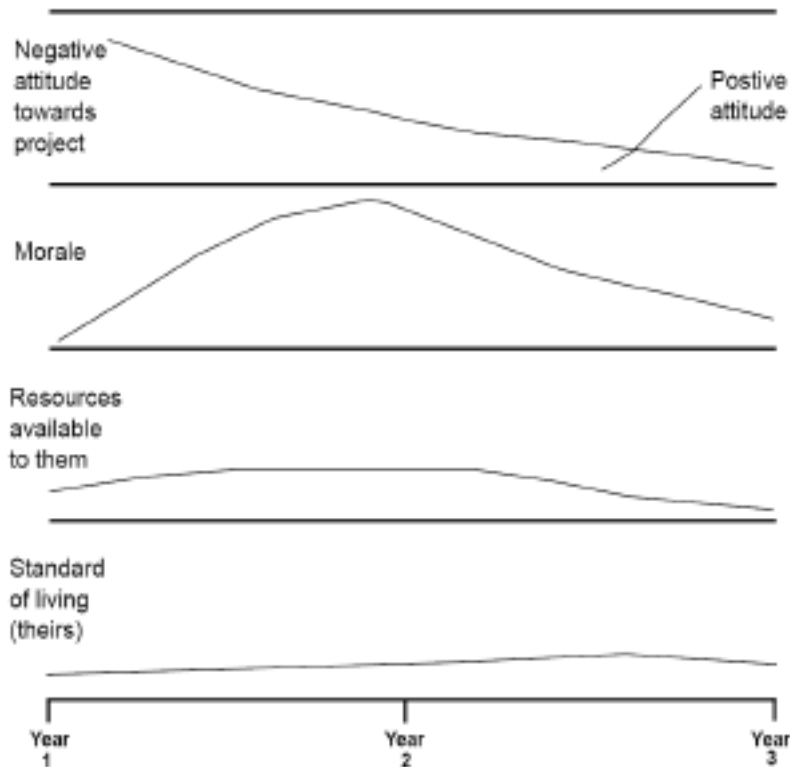


Figure 4: Government Divisional Officers



People opened up to us in surprising ways. The simplicity and visual nature of our methods and the intimacy that the small group discussions created meant that we were able to involve people of all ages and genders in the process of analysis; we did not anticipate, or experience, the kind of barriers to engaging women and girls that might be an issue either in more conventional interviewer-led research or more public participatory activities.

One theme that emerged in our discussions was how expectations of getting access to resources at the outset were stoked and maintained by a steady stream of white visitors. If we had been using conventional research methods, these very expectations might have inhibited what people were prepared to say to us given that two members of our team were white foreigners. But once our PRA methods got people focusing on the diagrams they were producing, they began telling us

their stories. For example, a time trends exercise with a community resource person revealed that his expectations of gaining something material from the program had been persistently thwarted, and remained a source of dissatisfaction. So when the program brought outside visitors and he had to spend the day without even being offered any food for the time he spent waiting for them to arrive and talking to them, he resented it. The day before we visited him in his home, he told us, he had been called to meet with the team that were conducting the review:

We hear about X [the donor] giving grants and support all over the country, and they give the government allowances and people see that, it is done in front of them, yet there is nothing for us. Like yesterday when we were out from 10am to 4pm without taking anything to eat for that review team.

The incident he narrated of being summoned to meet visitors is a development classic: in convening the obligatory performance of project success to visitors, scarce thought was given to the wellbeing of those putting on the performance; paradoxical indeed for a nutrition project to leave its participants hungry. We asked him: why carry on? At first, he told us, “we thought the Whites were bringing the money”. He had waited for that money, and waited some more and then just said to himself, “this is our country. Money or not money, let us get on”. But he was still waiting for some acknowledgement from the program, and for the material assistance that he continues to expect, for all that he sees the

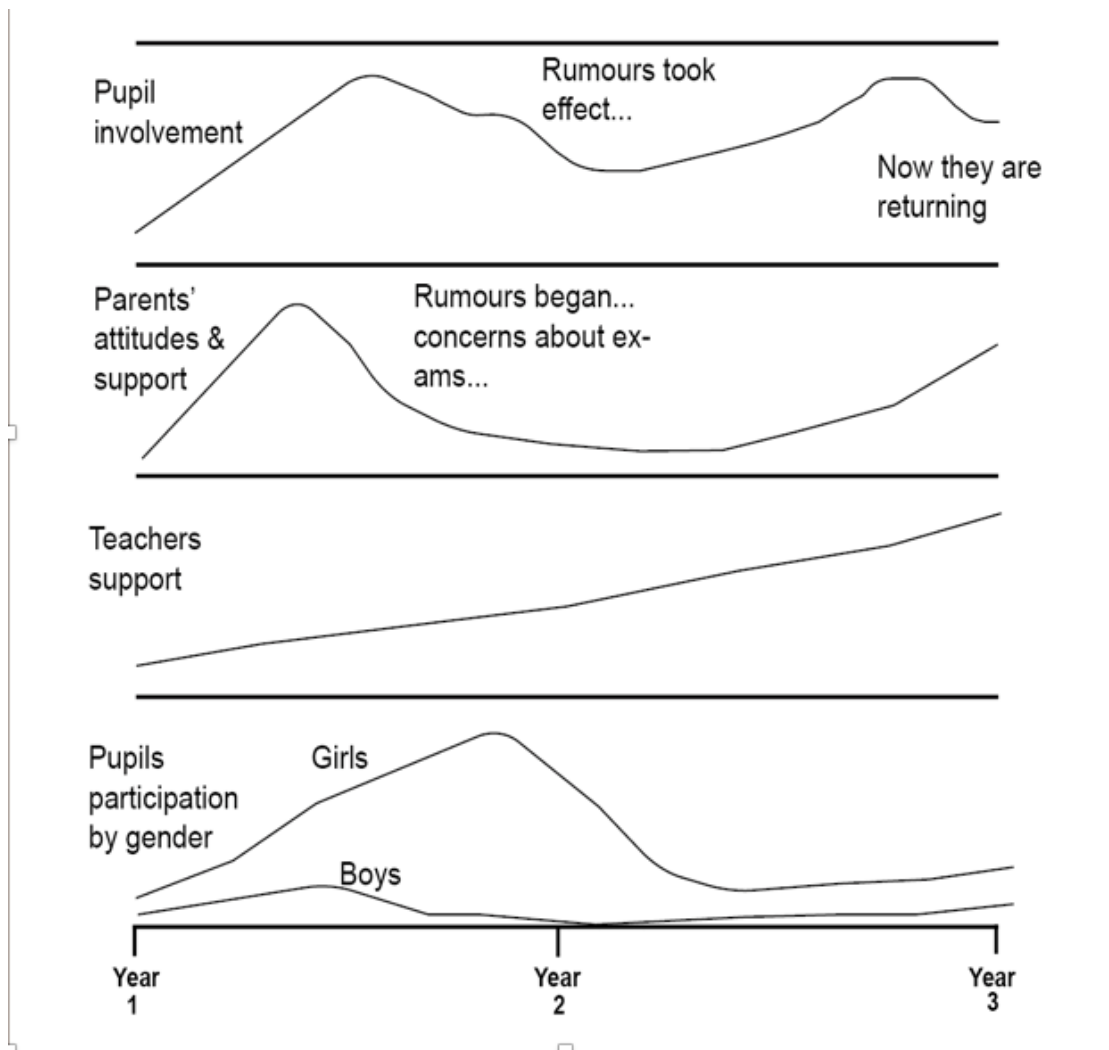
benefits of the program. Telling us about this provided him with another channel to make this evident to those running the program.

In another example of the frankness that participatory methods permitted, children at one of the schools told us that boys complained and didn't want to take part in the program at the beginning because it involved having to sing to visitors. The boys didn't think this was something boys should be doing; it was "girls work", they said, and in any case, they were shy about singing in public. They also told us how children had stopped coming to the club run by the program because many parents were worried about the associations that rumour had made between the program and devil worship. When we talked to the teachers and program staff, another lesson became evident: the program staff had realized that it wasn't enough to seek children's participation through the schools and to regard the children as a conduit to changing behaviour at home. Their parents had to be brought on board. The decision to do so reversed the dip in involvement, and had other effects as parents began to notice changes in their children and attribute them to the program. There are broader lessons from this about the value and importance of children's participation that resonate with the literature (see, for example, Johnson 2015).

These changes also proved decisive in convincing teachers that the Child-to-Child clubs supported by the program were worth their support. Using diagrams and cards to have a conversation opened up a space for teachers to tell us about their initial scepticism, if not downright hostility, to the initiative; they feared it would distract the children from their studies and waste their time. One of the

teachers confided that at the beginning, they thought they would be given books, equipment and other things, and that the teachers would get money and become rich. These hand-outs didn't arrive. But the Child-to-Child activities began to have tangible, visible, effects on the children and the neighbouring community. Teachers told us how, slowly, they'd come around. They didn't expect the children to get so involved, they said, nor to see the kind of changes that were emerging, changes that exceeded anything they could have expected at the outset. Something that especially surprised them was seeing the children teaching themselves and others. "The children have become quicker to learn for themselves", one teacher commented, so it became easier to teach them. One of the parents remarked on how she'd seen the self-confidence of the children grow, as they learnt and taught each other: she'd seen her daughter teaching her younger son, and both of them begin to help in the house to make some of the changes the program was teaching them about.

Figure 5: Time-Trends: School Teachers



Many of our richest insights came from the application of this simple time-line diagram sequence.ⁱⁱⁱ It offered a means of structuring an otherwise open-ended narrative biographical process in ways that generated comparable visual artefacts across different groups that could then become part of a meta-analysis of emerging themes. Putting these time trends diagrams side by side and looking for patterns, we could engage the stakeholder groups we worked with in reflecting on the complex causalities bound up with the changes that had taken place – negative, as well as positive, unexpected as well as expected.

Our commitment to confidentiality, the sensitivities that emerged and our respect for the intimacy of confidences shared in these encounters all meant that bringing our informants themselves together to compare versions across the different groups was not an option we wanted to pursue. The task of this part of the analysis – comparing their versions, sifting through what these different stakeholders had told us, and forming conclusions - was, ultimately, ours. We worked through our materials with painstaking care, juxtaposing the diagrams and transcribed commentaries and looking inductively for patterns, then focusing on understanding the nuance of the situated versions told to us by different stakeholders. In this way, we formed a collective version of the stories of change we were hearing.

These stories of change provided insights into how different dimensions of the unfolding story intersected with each other. The simple line diagrams told their own stories about the interactions of impacts and the changes taking place amongst different stakeholders in difficult-to-quantify areas like trust, morale and enthusiasm. This painted a complex picture in which the path of change was far from regular or linear. But our methods also brought to light some quite straightforward impacts, including some that would have been completely hidden from view in a more conventional evaluation. This included arguably one of the most impressive achievements of the program. Hearing stories that a neighbouring district had copied some of the practices introduced by the program, we requested a vehicle to see for ourselves. The program manager was hesitant: what did this have to do with the evaluation? *Everything*, we said. A

bumpy ride later, we came upon a scene of garden plots of leafy green vegetables, dish racks and latrines, all ideas borrowed from the program area and diffused through people's own networks beyond the geographical boundaries of the program: all evidence of 'impact by interaction' and 'impact by emergence' that might otherwise have remained invisible or considered to be inconvenient contamination.

Conclusion

Participatory process evaluation offers not only a means through which to better understand 'impact by design'. It also, as we hope this article has shown, lends an opportunity to understand 'impact by interaction' and 'impact by emergence' (Bhola 2000). With this, it can help to better answer questions about a program or policy, such as 'how could it be improved' (Jones et al. 2009), and 'will it work for us [elsewhere]' (Cartwright and Munro 2010), as well as some of the other questions Stern et al. (2012) highlight, for which experimental methods cannot provide us with credible answers.

In the case described here, participatory process evaluation revealed both dynamics and effects that would have been obscured if all that was of interest were whether a given intervention had the expected results. Arguably, these unanticipated impacts were more important to the success of the program in the longer term than anyone might have imagined at the outset. Some came about through elements of the process that had not been given thought in the design of the program. We came to learn, for example, how important rivalry and conflict

were as a motivating force and in accounting for the program's success. It is hard to imagine this being deliberately built into the program; nor would it have provided the basis for measurable indicators. And we also came to see tangible signs of sustainability of those successes in the diffusion of practices beyond the project area that a conventional approach to evaluation would simply have kept out of view.

A theory-based evaluation approach (White 2009) could have arrived at a similar focus on factors for success. But what emerges from this case is the value of a more inductive and dialogic approach. Multi-stakeholder narrative biographies of a given policy or program offer insights not only into *what* changes – potentially, as in this case, revealing changes that might not find their way into a Theory of Change – but also, importantly, into *how* and *why* change happens. Involving those who are engaged at different levels in program implementation in exploring 'impact trajectories' (Woolcock 2009) in a way that is sensitive to positionality sheds light on dimensions of impact that are otherwise difficult to access. For these and other insights into how interventions unfold, it offers a more thoroughgoing methodology for understanding the dynamics of change than impact evaluation that relies on quantification alone.

Transferable lessons from this evaluation are primarily methodological. The sequence of methods that were developed as part of the participatory process evaluation are simple, easy to train people to use, cheap and quick to implement and could be applied to a wide diversity of programs and projects. But there are also some wider lessons. One is the importance of cultural sensitivity in the

application of methods of evaluation that involve anthropometric or other forms of measurement. The fear that the use of physical measurement of infants generated took the program months to recover from. Yet one of the most positive unintended consequences of this was that much more effort had to be made to build rapport with people in the communities, and this helped deepen the participatory elements of the approach that was developed to program implementation. There are lessons, too, from the ways in which the program engaged people in activities. The child-to-child activities had some very positive outcomes, for example, as did the agricultural extension and health promotion activities which succeeded in transforming a food of last resort into a popular, widely-grown and eaten vegetable. Perhaps most of all, the participatory process evaluation revealed the benefits of emergent program design that is able to respond to changing circumstances and take up emerging opportunities.

Where participatory process evaluation is perhaps of most interest to contemporary international development practice, however, is as a complement to widely-used forms of impact evaluation for what it offers evaluators: a way to reach the parts that quantitative and quasi-experimental methods find especially hard to reach. Doing so can be especially helpful in enabling evaluators to address issues of external validity (Moore et al. 2015). Oakley et al. (2006:414) make the case for enhancing randomized control trials by making process evaluation integral to them, suggesting that ‘additional costs (such as collecting and analysing qualitative data) would probably be balanced by greater explanatory power and understanding of the generalisability of the intervention’ (2006: 415). Complex interventions call for modes of evaluation that need to be

as sensitive to process as to results if they are to provide generalizable conclusions (Moore et al. 2015). As such, participatory process evaluation has a valuable contribution to make to improving the rigour of impact evaluation.

References

Bamberger, M., Vaessen, J. and Raimondo, E. eds. (2016) *Dealing with Complexity in Development Evaluation: A Practical Approach*. London: Sage Publications.

Banerjee, Abhijit and Esther Duflo (2009) 'The Experimental Approach to Development Economics', *Annual Review of Economics*, 1: 151-178.

Bhola, H. S. (2000) 'A Discourse on Impact Evaluation: A Model and its Application to a Literacy Intervention in Ghana', *Evaluation* 6(2): 161-178.

Bleek, Wolf (1987) 'Lying Informants: A fieldwork experience from Ghana', *Population and Development Review*, 13(2): 314-322

Cartwright, Nancy and Eileen Munro (2010) 'The limitations of randomized controlled trials in predicting effectiveness.' *Journal of Evaluation in Clinical Practice* 16: 260-266.

Carvalho, Soniya and Howard White (1997) *Combining the quantitative and qualitative approaches to poverty measurement and analysis. The practice and the potential*. World Bank Technical Paper 336, Washington, DC: World Bank.

Center for Global Development (2006) *When Will We Ever Learn: Improving Lives through Impact Evaluation*, Washington: CGD.

Chambers, Robert (1997) *Whose Reality Counts: Putting the First Last*, London: Intermediate Technology Publications.

Chambers, Robert and Linda Mayoux (2005) 'Reversing the paradigm: quantification, participatory methods and pro-poor impact assessment', *Journal*

of International Development, 17: 271–298.

Cohen, Jessica and William Easterly eds. (2009) *What Works in Development: Thinking Big and Thinking Small*, Washington: Brookings Institution

Deaton, Angus (2009) 'Instruments of development: Randomization in the tropics, and the search for the elusive keys to economic development', The Keynes Lecture, British Academy, October 9th, 2008 (published with revisions in 2009).

Dupas, Pascaline (2011) 'Do teenagers respond to HIV risk information? Evidence from a field experiment in Kenya', *American Economic Journal: Applied Economics* 3(1): 1-34.

European Evaluation Society (2007) *The importance of a methodologically diverse approach to impact assessment*.

http://www.europeanevaluation.org/images/file/Member_section/EES%20Statement%20on%20Methodological%20Diversity/EES%20Statement.pdf

Eyben, Rosalind (2013) *Uncovering the Politics of 'Evidence' and 'Results'. A Framing Paper for Development Practitioners*, Brighton: Big Push Forward.

Eyben, Rosalind, Guijt, Irene, Roche, Chris and Shutt, Catherine (2015) *The Politics of Results and Evidence in International Development: Playing the Game to Change the Rules?* Rugby: Practical Action Publishing.

Galiani, Sebastian, Paul Gertler, and Ernesto Schargrotsky (2005) 'Water for life: The impact of the privatization of water services on child mortality', *Journal of political economy* 113(1): 83-120.

Galasso, Emanuela, and Nithin Umapathi (2009) 'Improving nutritional status through behavioural change: lessons from Madagascar', *Journal of Development Effectiveness* 1 (1): 60-85.

Gaventa, John (1993) 'The powerful, the powerless and the experts: knowledge struggles in a information age' in Peter Park, Budd Hall and Ted Jackson (eds.), *Participatory Research in North America*, Amherst, MA: Bergin and Hadley.

Gertler, Paul, Martinez, Sebastian, Premand, Patrick, Rawlings, Laura and Christel Vermeersch (2011) *Impact Evaluation in Practice*, Washington: World Bank.

Guijt, Irene (2008a) *Seeking Surprise: Rethinking Monitoring for Collective Learning in Rural Resource Management*, PhD Thesis, Wageningen University, Holland.

Guijt, Irene (2008b) *Critical Readings on Assessing and Learning for Social Change: A Review*, IDS Development Bibliography 21, Brighton: IDS.

Harrison, Elizabeth (2015) 'Anthropology and impact evaluation: a critical commentary', *Journal of Development Effectiveness*, 7(2): 146–159.

Havemann, K., Pridmore, P., Tomkins, A., Garn, K. D. (2012) 'What works and why? Evaluation of a community nutrition programme in Kenya', *Public Health Nutrition*, 19: 1-8.

Heckman, James (1991) 'Randomization and social policy evaluation', National Bureau of Economic Research Technical Working Paper No 107, Cambridge, MA: NBER.

Johnson, Victoria (2015) 'Valuing children's knowledge: the politics of listening', in Eyben, R., Guijt, I., Roche, C. and Shutt, C., eds. *The politics of results and evidence in international development: playing the game to change the rules?* Rugby: Practical Action Publications.

Jones, Harry (2009) 'The 'gold standard' is not a silver bullet', *ODI Opinion Piece* 127, London: Overseas Development Institute.

Jones, Nicola, Jones, Harry, Steer Liesbet and Ajoy Datta (2009). *Improving Impact Evaluation production and use*. Occasional Paper 300. London: Overseas Development Institute (ODI). <http://www.odi.org.uk/resources/docs/4158.pdf>

Leeuw, F. L. and Vaessen, J. (2009) *Impact evaluations and development: NONIE guidance on impact evaluation*, Washington: World Bank

Mackay, Ronald and Douglas Horton (2003) 'Expanding the Use of Impact Assessment and Evaluation in Agricultural Research and Development', *Agricultural Systems* 78(2): 143-165.

Moore, G., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., Moore, L., O'Cathain, A., Tinati, T., Wight, D. and Baird, J. (2015) 'Process evaluation of complex interventions: Medical Research Council guidance', *British Medical Journal*, 350: h1258.

Pawson, Ray and Nick Tilley (1997) *Realistic Evaluation*, London: Sage.

Oakley, A., Strange, V., Bonell, C. Allen, E., Stephenson, J. (2006) 'Process evaluation in randomised controlled trials of complex interventions', *British Medical Journal*, February 18; 332(7538): 413-6.

Ravallion, Martin (2009) 'Should the randomistas rule?' *The Economists' Voice*, 6 (2): 1-5.

Roche, Chris (2000) 'Impact Assessment: Seeing the Wood and the Trees'. *Development in Practice* 10(3&4): 543-555.

Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., & Befani, B. (2012). *Broadening the range of designs and methods for impact evaluations*, Working Paper No. 38, London: Department for International Development.

Tandon, Rajesh (1996) 'The Historical Roots and Contemporary Tendencies in Participatory Research: Implications for Health Care' in Korrie de Koning and Marion Martin (editors) *Participatory Research in Health: Issues and Experiences*. London: Zed Books.

Vaessen, Jos (2010) *Challenges in impact evaluation of development interventions: opportunities and limitations for randomized experiments*, Discussion Paper 2010/01, University of Antwerp, Department of Development Policy and Management.

White, Howard (2011) 'Achieving high-quality impact evaluation design through mixed methods: the case of infrastructure', *Journal of Development Effectiveness*, Vol. 3, No. 1, March 2011, 131–144.

White, Howard (2009) 'Theory-based impact evaluation: principles and practice', *Journal of development effectiveness* 1(3): 271-284.

White, Howard (2008) 'Of Probits and Participation: The Use of Mixed Methods in Quantitative Impact Evaluation', *IDS Bulletin* 39(1): 98-109.

White, Howard, and Edoardo Masset (2007). 'Assessing interventions to improve child nutrition: a theory-based impact evaluation of the Bangladesh Integrated Nutrition Project', *Journal of International Development* 19(5): 627-652.

Woolcock, Michael (2009). *Towards a Plurality of Methods in Project Evaluation: A Contextualized Approach to Understanding Impact Trajectories and Efficacy*. Working Paper 73, University of Manchester: Brooks World Poverty Institute.

Notes

ⁱ <http://www.thebrokeronline.eu/Series/Stories-from-Aidland/The-ghost-in-the-aid-machine>, December 7, 2010.

ⁱⁱ We are grateful to one of our reviewers for making this important point.

ⁱⁱⁱ For a time-line of the application of these methods, see Figure 1, which sets out the sequence used.