Adaptive programming in practice: Implementing structured hypothesis-testing and experimentation

Report

Amir Allana
Anya Briggs
Mia Johnson
Jessica MacArthur
Leanne Rasmussen

April 2017

Introduction

The international development community has acknowledged the usefulness of adaptive management in navigating the uncertainty inherent in complex problems such as poverty and inequality. This is particularly true in market systems development programmes employing facilitative and light-touch approaches to change systems. Given that market forces, behaviours, and social and political norms are dynamic and often invisible, it is challenging to design interventions that influence them successfully and predictably. Doing so requires continual intelligence gathering, assessment of whether interventions are having the desired effect, and changing work plans and strategies accordingly.[1] In addition to this ability to react to new information, "adaptive management calls for programmes that use a systemic approach to take an additional leap into embracing a purposefully experimental, hypothesis-based approach."[2] In other words, it is not enough to learn and adapt; rather, learning and adapting must be deliberately directed and channeled to serve specific programme goals. Furthermore, learning and adapting is not about randomly 'trying' things to see what works; rather, it is about intentionally and strategically honing in on strategies that will yield specific, positive development outcomes.

This document provides guidance and case examples on taking a purposeful, hypothesis-based approach to managing intervention strategies within a development programme. It is based on the experience of market systems development practitioners that are adapting the scientific method[3] and the 'lean start-up' methodology of business model testing to improve the effectiveness of development programming.[4, 5] The case for taking this approach can be encapsulated into three principles:

1. **Invisible dynamics of complex adaptive systems are made visible by *doing* something and watching for a reaction.** The assumption is that because there are a wide variety of social, economic, and political factors that govern the behaviour of people and institutions, and because they are continually changing, static research and assessments have limited utility. The best way to gain insight is to *try* intervening.

2. **Multiple, small experiments that fail early is the cheapest, quickest way to maximise learning.** The interventions themselves become research and assessments, and learning is an explicit goal of the intervention activity (as opposed to outputs and outcomes).

3. **A small number of imperfect, easy-to-measure metrics are more useful than perfect, comprehensive ones.** It is always 'nice to have' more information, but to remain nimble the experiments must be subject to frequent monitoring; bogging the team down is a real risk.

This paper is divided into two sections:
- **Section A** provides guidance on implementing a purposeful, hypothesis-based approach to managing development programming using an illustrative example
- **Section B** highlights three case studies, which informed the guidance in Section A.

---

1 Allana, A. and T. Sparkman. 2014. Case Study. "Navigating complexity: adaptive management and organisational learning in a development project in Northern Uganda." Knowledge Management for Development Journal, 10(3): 101-112.https://beamexchange.org/resources/169/

2 Goeldner Byrne, K, T. Sparkman, B. Fowler. 2016. "The road to adaptive management knowledge, leadership, culture and rules." https://beamexchange.org/resources/776/

3 O'Reilly, B. 2014. "How to implement hypothesis-driven development." ThoughtWorks. Accessed online: http://thght.works/2oIQeG1

4 Rasmussen, L. 2014. White Paper. "Adapting Lean Thinking to Market Systems Development." Engineers Without Borders Canada. https://beamexchange.org/resources/199/

5 Murray, P. and S. Ma. 2015. "The promise of lean experimentation." Stanford Social Innovation Review. Accessed online: http://bit.ly/2pP9dfT

1. Guidance

Based on practitioner know-how, we distilled the process of structuring hypothesis-based interventions into five steps. The goal is to 'learn by doing'. At the end of a learning cycle, depending on the insight yielded by experimentation, a programme may choose to pursue an intervention further or abandon it altogether. Alternatively, the programme may elect to develop a new set of learning goals and thus start the cycle over again.

```
                                              START
Scale-up or                                    HERE
abandon
intervention                              Define learning
                                          goals to guide
            Reflection,                   decision making
            analysis, and
            decision
                                                    Develop
                                                    time-bound
    Implement                                       experiments
    experiments and
    monitoring plan
                              Define
                              hypothesis
                              and metrics
```
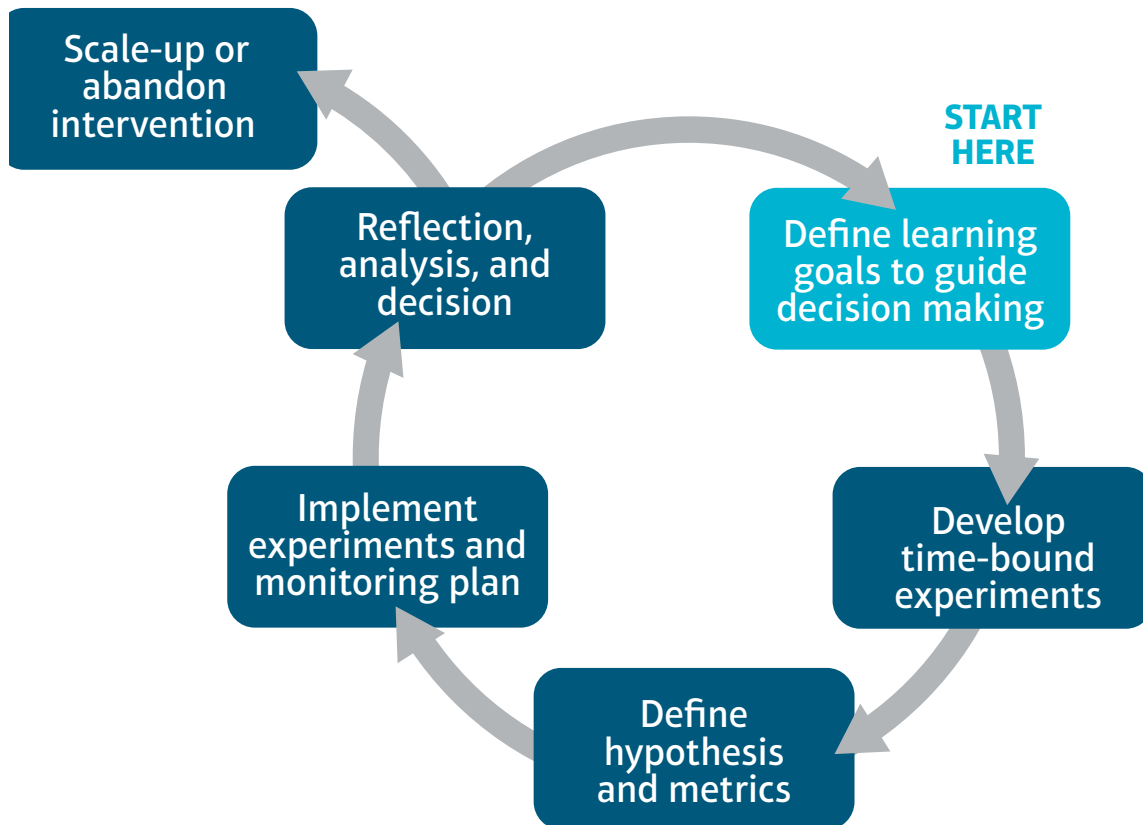
For the rest of this section, we describe each of these steps in more detail and with an illustrative example of a programme trying to intervene in a dairy market system.

> **Example**
> Project X is a market systems development programme that wants to improve participation of pastoralists in the dairy market. Project X has conducted research that indicates that there is an unmet demand in the country for local milk. They have also noticed a surplus of milk-producing, healthy herds in an economically poor region of the country. They have established a relationship with a respected national dairy company; the company has, with Project X's support, conducted outreach and marketing campaigns in the milk-producing region urging pastoralists to sell their milk. The communities express interest and excitement regarding the potential for earning an income off their milk. However, there has been little to no change in milk delivery.

Step 1 – Define learning goals to guide decision making

At this stage, a team must answer the questions: What is the purpose of the experimentation? What questions are you trying to answer? What specific actions will be guided by the end of the 'experimentation period'?

This seems obvious, but it's important to spend time thinking through what you're really trying to learn. We find that this type of structured experimentation is especially well-suited to learning about behaviour, which is inherently complex and difficult to predict.

Project X decides to investigate the collection and transport system, hypothesizing that this is where the problem lies.

Learning goal	Decision point	Possible decisions
e.g. "What milk collection system will instigate pastoralists to regularly sell their milk to the local dairy company?"	6 months	- Work with the dairy company to improve their milk collection and transport system - Determine that collection/transport is not the problem, and investigate alternative reasons for low participation - Abandon this intervention area

Step 2 – Develop time-bound experiments

At this stage, a team must decide: What 'tests' will they conduct? For how long? What is the logic behind these tests?

A programme could offer slightly different versions of the product or interventions to see which yield better results. The differences and similarities between the tests, as we will illustrate in Step 3, act as variables and controls that will yield insight. The number of tests or experiments to conduct will depend on the capacity of the programme, its partner organisations, and communities involved. Too many initiatives could overwhelm resources or disinterest local businesses; too few would reduce learning and insight. It is thus important to clearly define the resource commitment and scope of the interventions in partnership with all involved agencies.

For Project X, the team signs a cost-share agreement with the dairy company to experiment with three different strategies for a period of six months. They divide the communities in the target region into three clusters with roughly the same total population; each cluster of communities will be targeted with a different strategy.

- **Strategy A:** hire a representative from each community to collect milk, and coordinate transport with the dairy company.

- **Strategy B**: hire a third-party transport van to visit each community at a set time every day to pick up milk.

- **Strategy C:** set up collection points within 5 km of each community where a representative from the dairy company would visit at a pre-scheduled day and time, 2-3 times a week.

Step 3 – Define hypotheses and metrics

Based on the tests and learning goals, a team would now explicitly define: What are the underlying assumptions and hypotheses? What would you expect to see if these hypotheses are correct? What are 1-3 simple, easy-to-measure indicators that will support or refute the hypotheses?

Project X and the dairy company decide to track one metric: volume of milk collected per week. Higher volumes indicate a positive response; lower volumes indicate a negative one.

Note that this deliberately excludes additional metrics that could explain variability between the

strategies, for example community differences and preferences. Measuring more complex metrics can bog down the experiment and fatigue partner organisations; they are better left to focus groups and traditional research methods. The results of the experiment may help guide follow-on studies that explore issues further, but care must be taken not to roll those additional questions into the experiment itself.

Project X defined the following hypotheses:

Hypothesis	Result that would corroborate
The pastoralists prefer milk collection to be close to their residence, and are unwilling to travel to deliver their milk *(convenience hypothesis)*	More positive response to strategies A and B, compared to strategy C
Pastoralists are willing to travel a small distance, but would rather deal directly with a dairy company representative than a third party *(middleman hypothesis)*	Positive response to strategy C, negative responses to strategies A and B
Pastoralists prefer to deal with someone they know within their community, rather than an unknown third party *(trust hypothesis)*	Positive response to strategy A, negative responses to strategies B and C
Milk collection and transport are not the problem; the pastoralists are not participating in the dairy market for some other reason *(null hypothesis)*	Negative response to all three strategies

The results of the experimental strategies will yield insight into which of these factors holds truer; this will inform Project X and the dairy company's strategies after the six months has elapsed.

Step 4 – Implement experiments and monitoring plan

The implementing team must put some structure around the experiments: How often will the metrics be measured? Who will be responsible for collecting the data? How often will the team meet to review progress? How formal vs informal will the process be?

Project X and the dairy company may decide that the data is collected weekly, by the dairy company, and reported every Friday via a phone call to the Project X team. This is once again the value of simple indicators and metrics: there are only three numbers to collect (volume of milk per strategy per week). The team could decide that they will re-evaluate the six-month trial in formal meetings every two months, offering early opportunities to address implementation-reated challenges or terminate the experiment if needed. These formal meetings would involve relevant management and field staff from both Project X and the dairy company, and include reviewing the data as well as informal observations and learnings.

Setting these expectations from the get-go regarding who will do what, and at what time, is vital to ensuring smooth execution of experimental interventions.

Step 5 – Reflect, analyse, and make decisions

Based on the outcomes of the experiments and 'tests', the programme or business should have data to answer the questions they had set out in their learning goals and make some decisions. Broadly, the decisions tend to follow one of three trajectories:

• Scale-up, crowd-in, or invest more in an intervention area that yielded positive results,

generally when one or more hypotheses are proven to be true
- Define a new set of learning goals, and thus a new set of experiments, based on unclear evidence or unexpected findings
- Abandon the intervention area entirely

It is also important to know what level these decisions can be made - is this something that field staff has the power to determine, or is this a senior-manager level decision. This can be different for different types of tests, but should be understood by all involved.

For Project X, if there is success in one or more of the strategies, the next steps may be to convince the dairy company to expand that approach to the entire region or to design an approach based on the hypotheses that were proven correct (e.g., convenience, middleman, trust). Furthermore, they may launch a new set of experiments to determine what payment schedules yield the greatest engagement from pastoralists. Alternatively, if there are negative results across all three strategies, Project X may need to reconsider its assumption that collection and transport are the barriers to engagement in the milk market.

Employing such a purposeful, hypothesis-based approach may look different from pro-gramme-to-programme, depending on the programme's purpose and context. The next section highlights three different case studies of how organisations in Kenya and Bangladesh have been developing variations of this approach over time.

2. Case study examples

Case 1: Kuza Project, Kenya

Context

The Kuza programme takes a market-led approach to youth employment in an urban context, by working with local actors to create jobs in Coastal Kenya. In Kuza's micro-retail interventions, the team works with local entrepreneurs to establish micro-distribution companies that link large suppliers and manufacturers of fast-moving consumer goods (FMCGs) to small retail outlets. Youth access employment opportunities as sales agents, who collect orders and deliver products in a given geographical area.

In mid-2016, Kuza saw that micro-distributors had open sales agent positions that they could not fill. Quite often, they would hire a new sales agent who would remain in the position for less than a month, and sales positions would remain open for months at a time. This seemed counter-intuitive in a city with more than 40% youth unemployment. Wages for the sales agent positions were roughly equal to other opportunities available, but in contrast to many casual positions, offered a reliable monthly salary. The problem didn't seem to be compensation – so what was going on?

Hypothesis

After reflection, the Kuza team came up with the following possible reasons for why youth were not interested in the sales agent positions:

- Young people preferred to be paid daily or weekly, as opposed to monthly. Perhaps young people needed more consistent cash flows, rather than receiving a monthly lump sum.
- Young people may be more motivated by receiving commission for each product sold, rather than a flat salary. This was tied to an idea that young people may not have been feeling fairly compensated – instead feeling as though the entrepreneur was reaping most of the benefit for their hard work.
- Related to the point above, young people may prefer jobs where they felt a sense of ownership over, or inclusion in, the business. Kuza knows that many youth express a preference for 'being their own boss,' and perhaps the positions felt too restrictive in their hours, unrewarding in their outcomes, or impersonal in the relationship between the employees and business.

None of these factors were related to the actual amount of compensation. In that sense, Kuza's central hypothesis was that non-monetary factors were discouraging youth from taking up the sales jobs, and that different pay structures, or feeling more fairly compensated, could improve the retention of youth sales agents.

Tests and experiments

Kuza tried the following things:

- A sample of micro-distributors would offer weekly or daily compensation
- A sample of micro-distributors would shift a greater proportion of compensation to commission.
- In addition to the independent, entrepreneur-led, micro-distributors, Kuza would pilot micro-distribution with a number of youth groups. (In Coastal Kenya, locally-led youth groups often undertake income-generating activities with their members.)

Over the course of three months, Kuza would monitor whether employment retention improved, i.e. whether the micro-distributors ended up with more employees than they started off with by filling the vacant positions. For the newly-established youth groups, Kuza would monitor if they maintained their opening number of sales agents, or if youth group members quit at the same rate as they did with the independent entrepreneurs.

Measurement plan

The measurement plan could not have been simpler: every two weeks, a Kuza team member would call the micro-distributors and ask, "Hey, have you hired anyone? Fired anyone? Has anyone quit?" Answers were recorded in a spreadsheet that tracked the number of employees for each micro-distributor.

Thus, key metrics were: # of sales agents, # of vacancies, and # of staff turnover; monitored bi-weekly.

Implementation, results, and reflection

Amongst the independent entrepreneurs, the tests did not provide any conclusive evidence that pay structure (frequency + salary vs commission) had any systematic effect on attraction or retention of youth sales agents.

Instead, what stood out was the success of youth groups in running micro-distribution enterprises. Not a single group had any sales agents quit, and most even increased the number of sales agents during the three-month period. From talking with youth groups, the Kuza team believes that this is because youth feel invested in the success of their groups, and the positions feel more akin to 'being their own boss.' These sales agents also understand how much income the youth group is making from the venture, and feel that they are receiving a fair share of it.

When interviewing one of the independent micro-distributors who managed to increase his number of sales agents, Kuza uncovered similar dynamics. The entrepreneur was himself under 30 years old, and treated his staff more like co-entrepreneurs than employees; this was in contrast to feedback from other former sales agents who felt like they were not valued by their bosses. In addition to valuing his employees, this particular micro-distributor paid entirely in daily commission, and sales agents knew exactly how much profit the micro-distributor took versus how much they took. They felt like the arrangement was fair. Thus, in this case both monetary and non-monetary factors played a role.

These findings suggested that Kuza's hypothesis was partly true: non-monetary factors were responsible for youth's interest in the sales agent positions, with concepts of fairness and inclusion standing out as the most powerful non-monetary motivators.

Decision

As a result of the tests, Kuza has shifted its focus in micro-retail to working with youth groups to distribute a growing variety of products in different neighbourhoods. This pivot has enabled Kuza to create an increasing number of jobs for youth in the sector. Although Kuza is still willing to work with independent entrepreneurs, the team is now more aware of the dynamics needed for independent micro-distributorships to thrive and offer valuable employment opportunities.

Case 2: PROOFS Project, Bangladesh

Context

In the second half of the four-year Profitable Opportunities for Food Security (PROOFS)[6] project, iDE has worked towards institutionalising a micro-level adaptive management approach. PROOFS has worked to improve the food security of rural smallholder farmers through interventions in agriculture, WASH and nutrition.

Within the agricultural component of PROOFS, the project identified and trained 640 Farm Business Advisors (FBAs): individual smallholder farmers that provide agricultural information, products and services to farmer groups. The goal was to improve the flow of relevant information and services from the agricultural industry to farmers.

PROOFS worked on training FBAs to increase their businesses, however over the course of the project it became apparent that FBAs were not homogenous and that capacity development interventions were not meeting individual needs.

As a result of these findings, PROOFS shifted approach to design a FBA profile that was based on business size, experience and motivations to develop tailored capacity building for each FBA. This framework was developed in collaboration with field staff looking at both historical data from the first two years of the project and at live data. Under this new framework, project field staff would themselves analyse individual FBA needs, and develop tailored interventions to meet them.

By empowering field staff to have the responsibility and authority to test what will work on their own while reporting results to mid-level project management, we intended to foster a state of healthy experimentation. This shift is significant in a culture of traditionally hierarchical project management where capacity development comes in 3-day workshop packages with shiny workbooks.

Hypotheses and implementation

iDE started by building the capacity of field officers to undertake intensive assessments of FBAs. These assessments formed the basis for FBA profiles which included data on performance indicators such as sales, skills, attitudes, as well as business sales history and areas for potential growth. From these assessments, field officers created hypotheses of what type of training each FBA requires to expand his/her business specifically to the targeted farmer groups. For example, after an assessment, if a project officer identifies that an FBA is lacking in inventory management skills, the officer tests to see what the best approach would be: e.g., a peer learning event, a direct training. If the FBA already had a direct training which did not result in increased sales, the officer would facilitate a peer-to-peer learning visit for the FBA visiting another FBA who is known to have good inventory management skills. If after a peer-to-peer learning there was still not an increase in sales, the officer may move to one-on-one coaching based on conversations with other officers. This process was documented and helped to create a rubric of what types of training to use, when, where, and with whom.

Measurement plan

This iterative testing of different types of capacity development to fill different sized holes in capacity deficits (group cross-learning visits, peer-to-peer visits, one-on-one coaching, direct training on business elements and technical knowledge) helped to identify trends and categories of FBAs. A matrix of different capacity development methods against sales data and field observations for different types of FBAs helped to reduce the testing time leading to

---

6 Funded by the Kingdom of the Netherlands, implemented by ICCO, iDE and BoP Inc. (2014-2017)

faster capacity interventions. These were then discussed as a larger team during the *bi-monthly coordination meeting* to help harmonise the approach across all seven of the project districts with all 640 of FBAs.

Reflection and analysis

The benefits of this experimental capacity development approach is the institutionalising of regular market actor performance assessments, which in turn creates a culture of evidence based decision making. iDE has learnt that each market actor has different needs and by understanding these and having the ability to try different approaches, business growth is enhanced.

Case 3: SanMark, Bangladesh

Context

Scale. That little buzzword found all over development literature sounds fresh, impressive, and confident in a project's potential. When you redefine that word as *getting hundreds of masons to standardize their work practices given differences in knowledge, skills, terrain, and local resources, knowing that the slightest mistake could be the difference between public health safety and public health problem*, suddenly what seemed exciting and motivating to a project team becomes overwhelming and maybe even a little bit scary.

Sanitation Marketing (SanMark) in Bangladesh has long been characterised by training local entrepreneurs in how to make and market latrines. The approach has helped many gain access to sanitation, though there are still a number of kinks to work out. When consumers demonstrated an interest in offset pits, which are better for waste management, the setup proved quite tricky to install. Common designs included the construction of concrete waste collection boxes situated between the point of defecation and the latrine pit. Inconsistency in the installation depth and poorly mixed concrete made these additional structures prone to cracking and leaking. In a context so defined by training, the natural response was to simply re-train latrine producers on installation best practices. The SanMark Pilot[7] project team realised that they could try to control the actions of hundreds of local masons or they could try a different approach.

Hypothesis

The team hypothesized that most or all of the challenges involved in the installation of off-set latrines could be eliminated through a latrine design with simpler, more intuitive installation methods that would eliminate the need for further training.

Experiment

The team worked with a national manufacturing firm to design a new latrine model, called "SanBox". This new design eliminated the need for a separate structure altogether by attaching the collection box directly to the pan, at the point of defecation. SanBox was also made of strong plastic, which could be mass produced and eliminated the need for latrine producers to mix and make the old concrete models.

The team distributed 30 sets to a range of latrine producers with varying skill sets located across four sub-national districts. The team did not provide any additional installation guidance but instead allowed latrine producers the freedom to "figure-it-out". They assessed the installed SanBoxes for sanitation safety and documented obvious variances.

Measurement plan and results

Successful installation was based on an internally developed risk assessment for basic sanitation, with ten simple yes/no indicator questions. A 'no' answer indicated lower risk and risk scores were summed for each latrine. A score of ten would indicate a very risky latrine. Questions included some of the following: Is fecal sludge leaching out of the latrine system? Are there flies visible around the latrine pan? Is the latrine less than five metres from a water source? Risk results from all the latrine producers were compiled and reviewed by the project team. The team discovered that an improved design was only half the battle.

While SanBox was more durable and easier to install than the previous model, latrine producers were still showing variance in how SanBox was situated and secured. These inconsistencies could

---

7 Funded by the Swiss Development Cooperation and the World Bank; implemented by iDE (2012-2014)

have implications on latrine cost and safety.

Switch to multiple hypotheses

The team determined to pivot their approach. While the overall hypothesis that scaling challenges could be reduced through better design proved valid, the team realized other factors, such as packaging of latrine components, might also be a crucial factor for improving installation. Furthermore, the SanBox was originally designed to allow for the inner component to "pop" out for easy replacement. Consumer feedback suggested that this left a gap between SanBox's inner and outer components allowing insects to crawl inside. The team recognized the need to experiment in these dimensions.

Now in the follow-on project, SanMark Systems, the team is taking advantage of multiple field locations to test multiple hypotheses. Ongoing work to streamline the SanBox design continues. In some areas, latrine producers are casting the plastic into a concrete platform (slab) before sale; in other areas, masons are placing the SanBox on a perimeter of brick and filling the gap with sand to hold SanBox in place, while yet in other areas, the team is testing a version of SanBox that has inner and outer components welded together, further reducing the number of pieces. These tests are currently in process and results are not yet clear, however the pre-casting of the plastic SanBox in the concrete slab appears to be the best method so far. Alongside these pilots, SanMark Systems is also testing marketing strategies. Ongoing tests include comparing the effectiveness of marketing tactics such as live courtyard theatre, video ads at local tea stalls, and customer hosted testimonial events. Other experiments include testing consumer preferences around aesthetics as a purchase incentive, through the sale of painted or tiled slabs or coloured SanBoxes.

The use of adaptive management in iDE's WASH portfolio has grown proportionately with the increasing size of its target population and in the face of resource constraints. It is an approach that SanMark Systems has been embracing more and more, not for the sake of conclusive research but out of necessity. With these fast experiments attribution is not always something you can analyse with a regression line but the idea is let go of the academics of experimentation and turn it into something more pragmatic. Then with a little strategic flexibility the task of scaling is no longer scary, but exciting.

3. Concluding remarks

As principles of adaptive management – budgetary flexibility, changes to work plans, and an emphasis on learning cultures – become more widely adopted, development organisations must find ways to direct that learning and experimentation process. Learning and adapting is not about randomly 'trying' things to see what works; rather, it is about intentionally and strategically honing in on strategies that will yield specific, positive development outcomes. This may look different from programme-to-programme, depending on the programme's purpose and context.

Purposeful, hypothesis-based experimentation can take many forms, as illustrated by the plethora of examples in this paper. In all cases, however, this approach helped reduce wasted time and effort on strategies that don't work, and quickly identify the complex dynamics at play around a given intervention or market system. Taking such an approach, thus, holds promise for market systems development programmes seeking to operationalise adaptive management.