Fiala, Nathan; Mangan, Cormac

Working Paper

Improving development effectiveness through R&D: Dynamic learning and evaluation

DIW Discussion Papers, No. 1325

Provided in Cooperation with:
German Institute for Economic Research (DIW Berlin)

Suggested Citation: Fiala, Nathan; Mangan, Cormac (2013) : Improving development effectiveness through R&D: Dynamic learning and evaluation, DIW Discussion Papers, No. 1325, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:
https://hdl.handle.net/10419/85062

DIW BERLIN

1325

Discussion Papers

Deutsches Institut für Wirtschaftsforschung 2013

Improving Development Effectiveness through R&D: Dynamic Learning and Evaluation

Nathan Fiala and Cormac Mangan

IMPRESSUM

DIW Berlin
German Institute for Economic Research
Mohrenstr. 58
10117 Berlin

Tel. +49 (30) 897 89-0
Fax +49 (30) 897 89-200
http://www.diw.de

ISSN print edition 1433-0210
ISSN electronic edition 1619-4535

Papers can be downloaded free of charge from the DIW Berlin website:
http://www.diw.de/discussionpapers

Discussion Papers of DIW Berlin are indexed in RePEc and SSRN:
http://ideas.repec.org/s/diw/diwwpp.html
http://www.ssrn.com/link/DIW-Berlin-German-Inst-Econ-Res.html

Improving Development Effectiveness through R&D: Dynamic Learning and Evaluation[1]

Nathan Fiala Cormac Mangan[2]

September 2013

Abstract

Research and development (R&D) is a common process in for-profit organizations. Despite the benefits, it is not routinely practiced in nonprofit organizations, in part because it is difficult to identify the effects of programs that are designed to involve individuals over long periods of time. This paper presents a process by which organizations looking to affect social outcomes can learn from their programs in both the short- and long-run in order to develop the most cost- and impact-effective programs. We call it Dynamic Learning and Evaluation (DLE). DLE is a multi-arm experimental approach to program development that encompasses all stages of the design and implementation process. It combines a clear model of the causal chain of a program with high quality monitoring and impact evaluation. During the initial program development, organizations randomly apply multiple implementation designs and test them against each other using qualitative and administrative data. Once the organization determines a combination of designs that hold the most potential, they then implement these designs in the field and estimate impacts using participant data collection processes. The organization then uses the results to inform the next round of program implementation. They repeat this process over multiple designs for the life of the program and organization. At no point in the lifespan of the organization is this learning process stopped: programs are continually updated using systematic and objective methods to improve their design and impact. We present this process in detail.

Keywords: Research and development, program development, non-governmental organizations, development strategies, impact evaluation, monitoring and evaluation

JEL Codes: O32, O19, D73, L31

[1] This project was inspired by a presentation by Chris Blattman to DFID, entitled Impact Evaluation 2.0.

[2] *Nathan Fiala* (corresponding author): German Institute for Economic Research, DIW Berlin, Mohrenstraße 58, 10117 Berlin, Germany, nfiala@diw.de; *Cormac Mangan*: cormac.mangan@barcelonagse.eu.

1 Introduction

Before most companies introduce a new product, they conduct extensive research and development (R&D) to determine whether that product works in a particular context, whether there is sufficient demand in a given market, and what is the best form of that product. R&D is a common process in for-profit organizations, yet it is absent in the majority of nonprofit organizations. While incentives to carry out R&D differ greatly, there are two common organizational barriers. First, while for-profit organizations target those who utilize the product, the consumer, nonprofits seek funding from donors, which distorts the incentive to learn about a program. Second, while for-profit businesses tie the results of R&D to their financial results, nonprofits are more concerned about whether their work has actually affected and benefited a targeted audience. These effects can take a long time to materialize, often making organizational learning a slow process. This paper addresses the issues inherent in both of these problems, and presents a process by which organizations, including local and international NGOs, governments, and social entrepreneurs, looking to affect social outcomes, can learn from their programs in both the short- and long-run in order to develop the most cost and impact effective programs. We call this process Dynamic Learning and Evaluation (DLE).

The DLE is an experimental approach to program design, and requires managers and staff at all levels of the organization to participate in the continuous development of programs. This is especially important as many organizations, in part due to donor funding processes, have a difficult time justifying ongoing program improvements. We contend that organizations and donors need to think critically about their programs and admit that no program, including their own, is immune to improvement.

DLE is a multi-arm experimental approach to program development that encompasses all stages of the design and implementation process. It combines a clear model of the causal chain of a program with high quality monitoring and evaluation (M&E) and impact evaluation (IE). Using DLE, organizations can improve their existing programs or develop new ones. During initial program development, an organization applies multiple implementation designs and evaluates them using existing administrative monitoring data. Once the organization selects a combination of successful designs, these designs are then implemented in the field and the organization compares impacts using participant data collection. The results

are then used to inform the next round of program implementation. This process is repeated over multiple designs for the life of the program and organization. At no point in the lifespan of the organization is this learning process stopped: programs are continually refined to improve their design and impact.

While R&D is necessary for businesses to engage in, is it really something a non-profit or public organization should do? We believe that for most organizations, the answer is clearly yes. This is due to a number of reasons. Most importantly, there is almost always room for improving a program. There are few programs that run perfectly and would not benefit from some kind of improvement. In fact, one of the easiest ways to identify an organization that is destined for failure is if the leaders of that organization claim their programs cannot be run better.

It's also hard to know exactly what can be made better. Within an organization there are likely some obvious bottlenecks that might hamper the ability of an organization to perform smoothly, but it is not always clear what is the most important bottleneck. There could also be hidden behaviours and characteristics of beneficiaries that make implementation difficult. More broadly is the question of whether the current programs are the best way to tackle a development issue.

Finally, R&D helps an organization to stay competitive in the world of donor financing. Donors have limited funds to spend on programs and prefer to spend their money where they have reason to believe there will be the most effect. In the past this has meant a convincing sales pitch that attracts a donor's interest. There is though increasing competitiveness within donor circles. Demonstrating credible impact and a desire to learn can help attract donor funding.

It must be recognized that R&D is not appropriate for all organizations, just as it is not appropriate for all industries. Industries that are new or are undergoing changes often invest heavily in R&D to stay ahead of competitors or keep up with new technologies, and so we argue must many non-profit organizations. Development is often a dynamic process. R&D is most appropriate for organizations working in changing, dynamic areas where constant learning is key for success.

Program development is also important to donors, who have finite resources available. Not only should the donor be concerned with whether his/her money is being used effi-

ciently and well, but whether it is being used more efficiently and better than it would be by another NGO, program, or version of the same program. Despite the growing interest in evidence-based policy, there is still relatively little knowledge available on program effectiveness, making such decisions difficult for the donor. Donors are also in a position to take the greatest advantage of an organizational structure focused on R&D through learning at scale. By funding multiple program types across a number of organizations, donors can leverage their ability to develop multiple types of programs and determine what works and what doesn't much faster than a single NGO.

DLE is a rigorous approach to program development and learning. We stress that the DLE process is best served through a theoretical approach to program development, rather than an a-theoretical, "trial and error" approach. Recent research has provided a number of examples where a small change in a program can lead to large unexpected outcomes, often due to misunderstanding the behavioral and incentive implications of a program. By focusing on a theoretical approach that informs the development of programs, the DLE can narrow the wide range of programs tested and inform not just the local project but similar programs in different contexts.

This paper contributes to the literature on impact evaluation, M&E and organizational development in three main ways. First, the DLE allows for identifying the most important steps along a theory of change and structural parameters faster than existing experimental methods, and can do so when faced with a variety of local contexts that may affect outcomes. This leads to a more rapid formulation or confirmation of economic theory and understanding of interventions.

Second, the DLE is a practical and beneficial organizational form for NGOs that presents a complementarity approach for both researchers and policy makers. The DLE makes it easier for organizations to adapt existing research to a particular context. This means the current gap that exists between research and implementation can be more easily eliminated. It also ensures that there is an appropriate contextual adaption for existing research. It is difficult for organizations to adopt new programs or adapt existing ones using new information. This difficulty results in a gap between research and implementation, with fast growth in practical research but less in implementation. Program implementation requires flexibility, yet

progress and impact measurement can still guide the process. We encourage a contextual revision of the research, with methods for adapting it.

Finally, the DLE formalizes a way of incorporating learning into an organization. While the DLE process that we lay out here is meant as a road map for conducting high quality learning, we realize that most organizations are not in the position to incorporate such a process. Nor would we expect an organization to try to implement the entire process in the short-run. Over time though, we believe most organizations should work towards adoption of a similar approach.

The DLE has implications then for many different audiences. For practitioners, it offers a solution to the erroneous assumption that the best project design can be known ex-ante. By utilizing several designs instead of just one and learning from them, better programs can be developed. It also means monitoring and other data collection must become integral to the organization, not simply a side thought. For donors, the DLE means that longer funding horizons are necessary. Results must be proven with real data, while also balancing an acceptance of failure when organizations are actively learning from the failure. Funding support for M&E is also critical to ensure proper learning is being done.

The implication of the DLE for researchers is also important. The DLE provides a direct option for replication and development of results obtained in other studies. Replication is a key tenant of science and needs to be more broadly accepted in the social sciences. The DLE can also provide important insights into mechanisms and "deep parameters" that affect program outcomes, often in very subtle ways. Balanced with good theory, the DLE can thus produce significant learning for the research community as well as practitioners.

The remainder of this paper is organized as follows. We begin in section 2 by discussing what we see as the future of development programming, starting with the rising interest in evaluation and experiments, and then describing the broader DLE system and goals. We then present each of the steps of the DLE in detail in section 3. Section 4 discusses the role of failure in organizations, with section 5 concluding by summarizing why the DLE presents a novel and important step forward for organizations.

2 The Future of Development Programming

2.1 Impact Evaluation

Over the past decade the most prominent movement within development has been that of rigorous impact evaluation, primarily randomized controlled trials (RCTs). The profession has gained a wealth of knowledge about the nature of development programs from RCTs within a wide range of policy areas, from governance to education[1]. Governments and donors are enthusiastic about this progression, with many using the results to judge policy effectiveness and scale-up successful programs. The use of a more rigorous evidence-base of impact has also grown as a requirement for donors.

While the top-down modernist design approach to economic development has numerous well known shortcomings (Von Hayek, 1988; Easterly 2004), at an organizational level it persists, with both multilateral development organizations, NGOs and government departments designing and implementing interventions that lack an evidence base and contextual considerations. Legacy and inertia ensure that many of the development theories that found the basis of development organizations' work were conceived within a top-down development paradigm. Recently more systematic approaches have emerged that increasingly provide empirical quantitative evidence founded in economic theory. These approaches also disaggregate the development process and evidence requirements by donors. While there is a debate as to the suitability of such rigorous quantitative experimental evidence, the development community has broadly welcomed the evidence-based policy trend.

We will not present the case for running randomized controlled trials, as this has been done elsewhere[2]. We do note though that RCTs can play a key role in ensuring that programs can identify proper comparison groups, even at low sample sizes or using non-quantitative methods. Randomization is an efficient, fair and relatively easily applied method that can produce objective evidence of the value of certain programs, even at the design stage.

[1] Two new books by Banerjee and Duflo (2011) and Karlan and Appel (2011) present excellent overviews of recent work being done with RCTs.

[2] For an excellent technical review, see Duflo et al. (2010). Hempel and Fiala (2012) present a non-mathematical, simplified exposition.

While the RCT movement started simply, it has grown into a system that demands more than just identifying impact. The development economics literature is increasingly based on theoretical modelling that identifies the key variables and underlying structural processes behind interventions (Acemoglu, 2010). We are witnessing a progression from identifying what works, to focusing on what is driving the impact and explaining why a particular intervention is working. This increased level of theoretical understanding is important for building the capacity to model and predict the dynamics and impact of interventions under different conditions and environments.

However, there are limits to this methodology[3] and it is clear that the evaluation model needs to be extended: evidence of concentrated impact alone is not enough. We next introduce a framework for assessing program impact based on continual systematic analysis, and connect the evaluation to organizational decision-making processes to improve program impact as a result of the evaluation.

2.2 DLE System Description

We present in this section a quick overview of the DLE system and its stages. In the next section we discuss the specific steps in greater detail. Figure 1 presents a broad outline of the details discussed here.

The DLE can be applied at the level of individual aid programs that are carried out in complex, unpredictable environments within which it is inherently difficult to design effective programs. This converges with more recent thinking in complexity science regarding an "evolutionary" approach to development program design. We draw three core institutional features from Harford (2012) and Barder (2012), two prominent proponents of such an approach: systematic experimentation and innovation, the normalisation of small-scale failure, and effective timely feedback loops. One best-fit solution rarely exists in an organizational setting, but a range of solutions which requires organizations to prioritize amongst them. When deciding which to implement, organizations are often at an informational disadvantage, lacking full information on the impact, or even the probability of impact of a specific intervention. We view small-scale experimentation as a method to guide the decision-

[3] See for instance critiques from Deaton (2010) and Rodrik (2009).

making process and make more informed and effective decisions. The system itself is similar to some recent work, including the Problem Driven Iterative Adaption (Andrews et al. 2012) and experiential learning (Pritchett et al. 2012).

The DLE is an experimental and adaptive form of project design and evaluation that facilitates the identification of context-specific project variations in a changing environment. The proposed system facilitates a project design environment that encourages marginal and innovative experimentation alongside existing programs. This experimentation is institutionalized and embedded as a continuous part of the monitoring and project management structure, creating a tight feedback loop and allowing the best form of the program to emerge in an iterative process. Such a structure allows organizations to analyse alternative project designs through employing a monitoring system that actively provides feedback to quickly incorporate changes into the next phase of project design and implementation.

2.2.1 Initial Design

When designing an intervention that aims to model the key variables driving impact and the underlying structural processes, the project design phase is critical. We make the case for an increased rigorous and experimental focus at the early stages of program design. This can then also work to enhance later stage RCTs driven by generalizable theory built on consistent empirical observations across contexts. Researchers can identify the robustness of initial hypotheses at an early stage, and refine these hypotheses as the relative importance of project variables within each context becomes apparent. We possess the tools to assess impact through various economic evaluation methods; we now need the tools to understand processes and thereby design better impact.

This design is similar to what is presented in Pritchett et al. (2013). However, while they argue for a more expansive design phase where different designs are tested against each other, we expand on this idea by presenting a clear method with which to test these different designs using administrative data and a randomized methodology. We then explore how experimentation can be used along the entire program line to develop key information on the most important question for development practitioners: how will this impact people's lives?

This involves identifying the key steps that drive the theory of change and drives success in a given context. These are often the most important items to focus on when adapting a program to an alternative context. The results of an intervention depend on certain conditions in the study context that may or may not be present in other contexts, such as the education level of the smallholder farmers or the quality of legal institutions in the country. The DLE enables an organization to identify these conditions at an early stage in the design process. This is closely tied to the concept of external validity of evaluation results, an issue which has been illuminated by numerous researchers. Program success differs according to regional, institutional, political, geographical and cultural opportunities and constraints.

We extend the reach of systematic evaluation and experimentation to the project design stage, and in doing so improve program design and effectiveness, while also tackling some of the criticisms levelled at RCTs. In addition to representing what we consider the most effective path for program evaluation, the "research and development" nature of the system we propose provides a framework from which implementing organizations can both improve their existing programs and, crucially, adapt the existing research evidence base to make this framework function within different contexts, increasing the relevance of development research. We consider this an under-acknowledged area within development economic research. With the growth in policy-focused experimentation, there have been commendable scale-up efforts; however, the gap between research and implementation remains significant, as a result of numerous capability, informational and contextual barriers. Closing this gap provides an opportunity for development organizations to make large marginal gains in the impact and efficiency of their projects through quantifying the impact of comparative programs.

To ensure we can design the most effective programs, we must identify the core structural variables driving the impact. We introduce a gradual learning phase that incorporates feedback from small-scale trials to identify the factors that are most effective for a given intervention. Instead of pre-supposing a best-fit program design, we extend the systematic experimental approach to the program design phase, prior to gathering rigorous evidence of impact through RCTs. The knowledge gathered from this process helps to form coherent economic theory, enabling us to identify the most important economic mechanisms. When combined with empirical results, it also provides guidance to examine the external validity,

the predicted dynamics of the same intervention in different environments and at a different scale.

Initial Design involves a systematic micro-experimentation process that simultaneously analyses alternative project designs, providing active feedback of quantitative and qualitative data. Researchers work closely with a wide range of stakeholders to formulate initial hypotheses and discuss theories of change, which leads to two types of initial project designs: marginal and innovative. These designs are initiated with monitoring systems providing continuous feedback on inputs, outputs and outcomes.

This data is used to quickly incorporate changes into the next phase of project design and experimentation. As this cycle is repeated, an iterative process emerges from which programs evolve, forming idiosyncratic solutions within dynamic, uncertain and complex contexts. A broad range of input in employed to formulate an initial design that is tested, identifying the most-efficient and least-efficient project components. The process gradually shifts, drawing the knowledge that shapes the process from internal sources; therefore, the program can adapt to the form best suited to its context. The system described above enhances existing monitoring systems by employing them as an analysis and learning tool as well as to track implementation progress and accountability. Although impact cannot be determined at this stage, a measure of effectiveness of inputs, activities and outputs for each variation can be used to identify more successful variations using the dynamic learning system. From these metrics, as well as qualitative reports, knowledge is extracted that plays into the feedback loop for the next phase of project design. Minimal project variation between and within consecutive phases allows us to continuously identify both the successful and unsuccessful project components. This cycle of 'innovate - experiment - test - learn - repeat' embeds itself within the organization. It is crucial that the project researchers normalize failure. With ongoing iterations, only a few will succeed, and it is necessary to acknowledge and document both. The role of failure is discussed on more depth at a later stage.

2.2.2 Impact learning

The next stage of the DLE process, impact learning, utilizes local project design successes that emerge from the dynamic learning process. This delivers a more robust causal interpretation and analysis of the successful project dynamics and underlying structural processes,

building broad network knowledge and contributing to a more generalizable understanding of the program impact. This involves more quantitative rigorous impact evaluations in the form of overlapping randomized controlled trials to better understand the underlying characteristics that drive impact. Both stages continuously work together to form a usable organizational knowledge platform that feeds into the research knowledge base utilising small-scale experimentation to adapt, improve and innovate. As RCTs take a longer time-period to carry out and information is continuously provided from the dynamic learning process, as soon as the next successful design iteration is identified, an overlapping RCT is initiated.

2.2.3 A network organization perspective

Many development organizations and their partners work in multiple contexts. This can make communicating implementation strategies difficult, especially when context is very important. The DLE can be used as a learning process for organizations that communicate well across a network of partners. Through building a diverse multi-faceted knowledge of what works in different environments, we can identify the components of programs that are consistently effective across contexts, and which components are not consistent. This contributes to the construction of the theoretical foundations necessary to design a randomized controlled trial that is rooted in investigating the structural dynamics behind the intervention.

At the early stages organizations implement network knowledge based on a theory of change, adapting it to the required context. At later stages we build and inform theory based on the network knowledge that contributes to understanding the underlying structural processes that drive intervention impact. DLE facilitates the adaption of the range of knowledge currently produced by impact evaluations in addition to network knowledge within large organizations, increasing implementation relevance. Successful innovations can be quickly transferred through the network and be replicated within a short period and at little cost in a different context, increasing the usability of the knowledge and building external validity on a small scale as different organizations test and adapt different variations. Rigorous evaluations serve to inform both network knowledge and broader academic knowledge by delivering a more robust causal interpretation and analysis of the project dy-

namics. They support network knowledge in that more generally applicable lessons can be drawn and applied to different organizational settings.

As well as learning about the most effective form and features of a particular program, we also gradually learn about the comparative impact of different programs. Thus, we can update our knowledge of the most-binding constraints.

3 Steps of the DLE

In this section we outline the steps of the DLE system alongside an example. These steps can be seen in whole and graphically in Figure 2 at the end of this section. The initial stages, while echoing the ideas presented in Pritchett et al. (2013), are an expanded methodological approach to design stage learning, while the later stages take advantage of randomized impact evaluations, sometimes overlapping but always developed with feedback loops, to increase learning on impact.

3.1 Define target issues and theory of change

Because there is extensive literature defining target issues and formulating theories of change (e.g. Bakewell 2004), we provide only a short overview. First, we must identify the problem at hand, the most constraining factor impacting the problem or "binding constraint", and a range of possible solutions that target this constraint with related theories of change. This takes a similar analytical approach and adopts the phrasing of Hausmann et al.'s (2005) "growth diagnostics" framework, except applied at the microeconomic level. The problem can first be defined locally. For example, "there is low micro-enterprise growth in a particular region".

The problem is then diagnosed: an initial investigative questionnaire is carried out alongside an extensive broad evidence gathering process. This serves to provide preliminary descriptive insights and evidence-based input regarding the source of the constraints and works toward identifying the most binding of those, i.e. determining which factor is most preventing micro-enterprise profit growth in region X. We aim to identify the most constraining factor in order to prioritize the targeting of intervention efforts and achieve the greatest impact. Is it that there is a lack of business management knowledge? Is it that there is no access to credit? Is it that business costs are inflated as a result of corruption?

After the binding constraint is identified, we delve deeper into the roots of the constraint to investigate the specific underlying distortions. In our example, if we identify business knowledge as the constraint, then we disaggregate the possible causes. Is it that the cost of education is too high? Is it that business owners don't know that education will help? Is it that there is a stigma to education? This evidence further narrows the range of possible solutions. Throughout this process the initial qualitative investigation is combined with existing evidence-based analysis of the problems and stakeholder knowledge to assess the source of the problem. The input of stakeholders familiar with the institutional and implementation environment is crucial in assessing the constraints as well as the range of solutions and their potential impact as they are in the best position to identify and predict contextual and political economy barriers.

The behavioural economics literature is replete with examples of rational intentions not translating into rational actions and simple solutions presented in the form of "nudges" that bridge this gap to improve development outcomes. As the evidence builds, it is clear that high-impact and cost-effective behavioural solutions exist to development problems; therefore, a behavioural analysis should be included in the early stages of program design. For example, self-control, attention and cognitive complexity can each lead to difficulties in implementing even the best designed programs. Something as basic as procrastination has been found to cause inefficiencies to the scale of 18% (Duflo, 2009). These counter-intuitive psychological insights can be difficult to generate, but the literature is growing. When designing projects, we must be aware of the possible behavioural barriers and resulting solutions.

3.2 Generate initial program designs

After identifying the most binding constraint, researchers seek solutions by soliciting a broad range of stakeholder and external input. The evidence for each solution should be examined and, where possible, a theory of change delineated as to the process through which the solution will be enacted and measured. To determine the probability of success of such an intervention and gage a rough assessment of effectiveness, preliminary small-scale qualitative research should include open questions regarding potential solutions. These should

always be designed with a causal chain in mind, though exploration can and should be encouraged.

Exploratory solutions are those for which there is a lack of hard evidence, but for which there are strong proponents or a supporting theory of change. They may seem counterintuitive and should by definition be relatively risky to enact due to a low probability of success; however, with high impact should they succeed. Designers should incentivize novel intervention design within the project design phase, with emphasis on the fact that failure is accepted as likely, but it is also contained and relatively cheap as a result of the micro-experimental approach. We discuss more about the role of failure below. The supporting evidence for such programs can be sourced from practitioner experience, reports, substantiated anecdotal evidence and behavioural solutions.

Marginal solutions are less risky and based on strong empirical evidence or existing successful project designs within the organization. They take two forms:

1. There already exists strong external evidence of impact. However, this evidence may need to be marginally adjusted to adapt to the context of the intervention. The existing knowledge base of academic theory and evidence helps to construct the theory of change to build off.

2. Starting with incremental programmatic changes to existing programs, organizations can gradually refine them. Marginal solutions should improve on the key parameters.

On average, the process is making gradual improvements to the program designs, with a smaller probability of large leaps should an effective exploratory innovation be identified.

3.3 First implementation phase

After the most promising marginal and exploratory project designs have been identified, organizations initiate the first micro-experimentation phase. The most promising program

designs are run concurrently, using the largest number of observations available so as to increase the comparability of the different designs[4]. At this stage, we do not employ a control group; therefore, we will be using the accompanying trials as each other's counterfactuals. It is crucial to randomize which participants or groups are assigned to each trial, in order to ensure that selection bias is avoided, increasing the relevance and applicability of the analysis (Duflo et al. 2007). Through randomization we can infer that depending on the number of participants, the participants across each trial will have the same characteristics on average, ensuring it is only the program components that are driving the results.

3.4 Carry out monitoring and feedback

During the first phase of experimentation, the organization gathers feedback in order to assess the results of the trials. Organizations can then enhance existing monitoring systems so as to gather relevant information using a mixed methods, qualitative and quantitative, evaluation approach to gage the effectiveness of the trials, accumulating a broad evidence base for each arm (Deaton 2010; Rodrik 2010). Working in relatively small numbers, no method alone will hold enough explanatory power to infer clear causality. A mix of interpretive judgement and statistical analysis should guide the judgement of trial success. Although we cannot infer impact from the evidence base, if the assessment is consistently applied across trials we can gain a reliable assessment of project outcomes. This also avoids the pitfalls of relying on one sole method, providing additional insights into barriers and potential changes from different perspectives. With quantitative analysis, we measure the outcomes whereas with qualitative analysis we can better understand the processes and underlying issues (Rao and Woolcock, 2008). For example, a quantitative analysis alone may not identify political economy barriers, whereas an overly qualitative analysis may not accurately measure resulting changes in outcome variables. A qualitative approach allows feedback from the participants as to their experience and understanding, providing valuable input into the theory surrounding the intervention. Quantitative and qualitative data collection can be combined in a survey instrument across trial participants including both open-ended sub-

[4] Note that we do not argue for statistical power of the analysis at this time as the focus is less on analysis and more on broad learning,

jective and categorised quantitative questions. The range of feedback mechanisms used to measure the effectiveness of the trial phase can include the following:

Quantitative

- Quantitative survey instruments
- Regression analysis
- Descriptive statistics analysis

Qualitative

- Representative Interviews with participants and those implementing the program
- Observational analysis
- Focus group discussions
- Participatory techniques

3.5 Dynamic learning

This stage, as the name suggests, is the driver behind the overall DLE process. Employing the feedback gathered above, changes are incorporated into the next iteration of trials. The feedback provides evidence to identify which components within trials need to be adjusted, retained or cut. Researchers may also source new design parameters from the feedback include it in the next phase of trials. The feedback leads to a learning process whereby the gathered evidence is discussed in detail, the lessons are recorded and incorporated into the next trial design phase. It is important to accurately document both effective and ineffective trials and the proposed underlying reasons for ineffectiveness so as to ensure that the all the evidence gathered is recognised for future decision-making purposes.

Steps 2-5 are continuously repeated, with the programs becoming more effective and refined over time, and periodic successful exploratory innovations providing large gains in effectiveness.

When we reach significant decreasing marginal returns with one program, we can return to the approach introduced in Step 1 to identify the next most binding constraint on the issue at hand. This can be introduced to the iterative experimental cycle, while we continue to test and improve existing programs.

3.6 Rigorous Evaluation

After successful, effective programs emerge from the dynamic learning process, randomized controlled trials (RCTs) are carried out. RCTs provide a more reliable inference of causality between the program and its impact. The methods and advantages of the RCT approach to evaluation have been discussed previously and so will not be discussed at length here. While randomisation is consistently applied form the initial stages of experimentation, the primary methodological difference between an RCT and the small-scale trial approach in the dynamic learning process is that we utilise a control group as a counterfactual in an RCT, ensuring that selection bias is avoided. They help us to infer causality and model the empirically observed behaviour in order to use the evidence for predictive purposes.

RCTs thus provide something that has been missing from the previous stages: rigorous evidence of impact. Rather than relying on small-scale trials, a full impact evaluation is conducted at either partial or full scale. A partial scale test, such as a pilot, can provide low cost information on the impact of a program before the full resources have been used. Full scale tests allow for understanding impacts when there are larger organizational issues at stake.

Whatever scale the evaluation is conducted, a successful RCT relies on sufficient statistical power to ensure the effects are properly identified. In practice this means a certain minimum number of participants are needed to act as the treated and control groups. Depending on the program, this can range from a few hundred to over a thousand participants.

3.7 Midterm Review

RCTs can take years to fully identify impact, therefore we utilize mid-term results and the continuous feedback from the dynamic learning process to form new project designs and initiate another RCT. In some cases it may be possible to conduct overlapping evaluations, that is, begin a new RCT before the previous one has completely finished. This is possible when new learning from the experimental stages has led to an alternative design to test, but waiting for the full results of the on-going RCT is not necessary.

Feedback loops are the processes that facilitate the flow of information within a system in order to affect change. As demonstrated in Figure 2, the feedback loop in DLE is the process we harness in order to make alterations to improve, adapt or discontinue the project. Feedback processes transfer information back to its source, comparing information to an

identical feedback loop measuring outputs in another project. It should connect the intended beneficiaries to the actions of the organization. Feedback is commonly used to inform and improve programs however its use is rarely fully translated into action in a meaningful form, an error that is not unique to the development industry but is also apparent in the private sector as discussed by Markey et al. (2009). Monitoring information is commonly employed as a management and donor accountability tool rather than a source of constructive feedback to improve outcomes for the beneficiaries. Therefore we utilise the monitoring system in such a way that it is providing feedback regarding the outcomes for beneficiaries, and the resulting changes quickly implemented, increasing the relevance of the feedback mechanisms.

Soliciting active feedback and promptly acting on the feedback can give a lead to productivity improvements within private organizations, as it should also within development organizations. Our model represents organizational M&E capability as a continuum running from basic monitoring to full randomized controlled trials. Feedback becomes more and more quantitatively rigorous over time, as capability grows. While it always stays central to the decision-making process, its composition evolves.

This design of feedback also helps to ensure access to funding. As the DLE constructs many short feedback loops in the form of small experiments, and comprehensive feedback in the form of RCTs, evidence is always being created to be both used for learning, and for proving impact to donors.

4 Failure

"I've not failed, I've just found 10,000 ways that won't work" – Thomas Eddison

The DLE emphasises the normalisation of failure as a necessity to create a continuous learning process. The structure mitigates the cost of failure, enabling us to halt implementation of ideas that are supported and enacted within an organization based on sound logical frameworks, but that are ineffective. This is a concept that forms the core of evidence-based policy. In order to learn, we must be willing to acknowledge what works and what does not work in each context, failure can be valuable in forcing us to review approaches, reshape programs, and ultimately improve services to beneficiaries. Evidence of program failure

should not be viewed as wasted resources but an opportunity to redirect them to areas that provide a higher return. Failure drives innovation, to innovate is to take risks through investing resources in relatively unproven designs and inherent in risk there is the probability of failure.

The concept of failure as a creative economic force has influenced economists since Schumpeter's (1942) description of the process of creative destruction. Better ideas dominate in an unimpeded free market where new ideas lead to the destruction of less efficient systems and products, often represented by the existing market players. The contribution of this system is in harnessing failure as a creative force within development organizations where the organizational structure previously acted as a barrier. So failure is normalised and contained in the DLE setting through providing a space where it can be identified and contained early on in the design process, keeping the costs of failure low, both monetarily and in organizational investment.

In the most innovative and profitable private sector industries, failure is acknowledged as inevitable on the path to success. Venture capitalists (to pick an extreme example) routinely witness high rates of failure, with rare but high-return successes. To encourage progress and the discovery of high-return activities this approach can be imitated by development organizations, as happens through the DLE exploratory design trials. While this comparison to venture capitalism is appealing, it isn't complete. One feature that separates the private sector from the public sector is the market price mechanism, essentially a highly efficient feedback loop (under the right conditions) that provides an accurate indicator of value to end-users. The level of value that a customer (beneficiary) derives from a product is reflected in the price. We do not have such a system in development programs, therefore we must construct effective feedback mechanisms to imitate this valuation mechanism. This is what the DLE attempts to do, to uncover and quantify the value that a program represents to a beneficiary, so that the highest return may be selected. Therefore at the core of the normalisation of failure, is the acknowledgement and analysis of failure, for which we need feedback loops.

Edmondson (2011) identifies such a form of failure as "Intelligent failures at the Frontier." The feedback of the failure results in an evolution away from the failed processes.

Note that we are identifying comparative failures through the DLE approach, not absolute failures.

There have been some impressive moves towards an acceptance of failure recently such as Engineers Without Borders' "Failure Reports", that promote failure as a necessary part of development projects and organizational learning through failure. However these are the exception, donors and management still need to overcome many barriers in order to instil a view of appropriate failure as a necessity in a growing, learning organization.

Resulting from the organizational cultural change brought about by the DLE, the design space from which potential programs and program features can be drawn is expanded. External evidence previous innovative but risky ideas can be tested, with failure accepted as a possibility and learning the central driving force. Thus the demand for ideas and learning grows within the organization given that the downside is eliminated as failures are contained at an early point in the design cycle, and the upside expanded. The reduction in implementation risk brought about by DLE compliments an increase in the returns to learning and the incentives to incorporate lessons, especially if innovation is rewarded.

5 What's different about DLE?

The steps that we have described here are not new to the for-profit business community. They do though present a novel way for non-profit organizations to learn and develop programs that produce the highest impact. We believe the DLE presents a novel and important step forward for organizations in seven ways.

First, from a methodological perspective, program designs are shaped within the relevant environment and thereby hold the most appropriate form within that context. The DLE does not just evaluate impact, it continually enhances impact. While contextual specificity is foremost a problem of adaption, through the DLE we can take the evidence from one context, and adapt it so that its success is replicated in another. Through this approach behavioural, political economy and other unanticipated barriers can be identified and overcome at an early stage.

Second, current approaches remain static throughout the evaluation, in a dynamic environment DLE actively adapts as is needed over time, continuously providing feedback and testing features that require adjusting alongside environmental changes.

Third, the DLE promotes innovation through novel experimentation and the normalisation of small-scale failure. Ineffective innovations are contained at an early stage, successes are amplified and scaled up.

Fourth, the DLE incorporates also input from and is jointly enacted by the implementing organization, increasing organizational ownership and decreasing costs. Additionally, it is a less costly system overall as it reaches RCT scale after cost efficiencies have been accounted for in the dynamic learning phase. The relevance of the monitoring and evaluation system within each organization is also enhanced as the process is designed to continuously engage managers and complement project implementation. We move from accountability-focused monitoring systems to a participatory approach through the monitoring system takes into account the views of the field-staff and beneficiaries. This also helps to legitimise the new practices.

Fifth, successful innovations are quickly transferable through a knowledge network to be replicated within a short period and at little cost in a different context, increasing the usability of the knowledge and building external validity on a small scale as different organizations test and adapt different variations.

Sixth, the DLE is an efficient way of closing the gap between research and implementation, enabling relatively small organization to make research relevant, facilitating contextual adaption through small-scale experimentation.

Finally, the DLE can be used to fit within a broader knowledge framework, culminating with rigorous evaluations and contributing to academic knowledge that can be shared with other organizations. Global knowledge generation is a public good that we believe more organizations should be looking to develop.

References

Acemoglu, D. (2010). Theory, general equilibrium and political economy in development economics (No. w15944). National Bureau of Economic Research.

Andrews, M., Pritchett, L., & Woolcock, M. (2012). Escaping Capability Traps through Problem-Driven Iterative Adaptation (PDIA). Washington, DC: Center for Global Development Working Paper, (299).

Bakewell, O,. and Garbutt, A. (2005) The Use and Abuse of the Logical Framework Approach. Report for SIDA.

Banerjee, A., Banerjee, A. V., & Duflo, E. (2011). *Poor economics: A radical rethinking of the way to fight global poverty*. PublicAffairs.

Barrett, C.B. and M.R.Carter (2010).“The Power and Pitfalls of Experiments in Development Economics: Some Non-random Reflections,” Applied Economic Perspectives and Policy 32(4).

Bold, T., Kimenyi, M., & Mwabu, G. (2012). Pitfalls of Scaling-up Proven Education Interventions: Evidence from an RCT in Kenya.

Campos, F., Coville, A., Fernandes, A., Goldstein, M., & McKenzie, D. (2012). Learning from the experiments that never happened: lessons from trying to conduct randomized evaluations of matching grant programs in Africa. World Bank Policy Research Working Paper, (6296).

Datta, Saugato, and Sendhil Mullainathan. (2012) ‘Behavioral Design.’ Center for Global Development Policy paper 016.

Deaton, A. (2010). Instruments, randomization, and learning about development. Journal of Economic Literature, 48(2), 424-455.

Duflo, E., Kremer, M., & Robinson, J. (2009). Nudging farmers to use fertilizer: theory and experimental evidence from Kenya (No. w15131). National Bureau of Economic Research.

Duflo, E., Glennerster, R., & Kremer, M. (2007). Using randomization in development economics research: A toolkit. Handbook of development economics, 4, 3895-3962.

Easterly, William (2007). The White Man’s Burden.

Harford, Tim (2012). Adapt: Why Success Always Starts With Failure.

Hausmann, R., Rodrik, D., & Velasco, A. (2005). Growth diagnostics. The Washington Consensus Reconsidered: Towards a New Global Governance, 324-255.

Heckman, James J. 1992. Randomization and Social Policy Evaluation. In Charles F. Manski and Irwin Garfinkel, eds., Evaluating Welfare and Training Programs. Cambridge, MA: Harvard University Press.

Heckman, James J. 2010. Building Bridges Between Structural and Program Evaluation Approaches to Evaluating Policy. Journal of Economic Literature 48 (2): 356–98.

Kaur, S., Kremer, M., & Mullainathan, S. (2011). Self-control at work. working paper, Harvard University.

Karlan, D. S., and Appel, J. (2011). *More than good intentions*. New York: Dutton.

Leamer, Edward E. 2010. Tantalus on the Road to Asymptotia. Journal of Economic Perspectives 24(2):31-46.

Markey, R., Reichheld, F., & Dullweber, A. (2009). Closing the customer feedback loop. Harvard Business Review, 87(12), 43-47.

Pritchett, L., Samji, S., & Hammer, J. (2012). It's All About MeE.

Rao, V., & Woolcock, M. (2003). Integrating qualitative and quantitative approaches in program evaluation. The impact of economic policies on poverty and income distribution: Evaluation techniques and tools, 165-190.

Rodrik, Dani. (2009). The New Development Economics: We Shall Experiment, but How Shall We Learn? In What Works in Development: Thinking Big and Thinking Small. Jessica Cohen and William Easterly, eds. Washington, D.C.: Brookings Institution Press.

Schumpeter, J. A. (1942). Capitalism, socialism and democracy. Routledge.

Figure 1: Dynamic Learning and Evaluation Capability Continuum

Figure 2: The Full DLE System in Practice