THE BIG PUSH FORWARD

MAKING SPACE FOR FAIR ASSESSMENT FOR A FAIRER WORLD

The Politics of Evidence Conference Report

April 23-24, 2013 at Institute of Development Studies, UK

Convenors: Rosalind Eyben, Irene Guijt, Chris Roche, Cathy Shutt, Brendan Whitty

Summary

On April 23 and 24, 2013, one hundred development professionals debated 'the Politics of Evidence'. The conference provided an opportunity to share and strategize for people working on transformative development, and who are trying to reconcile their understanding of messy, unpredictable and risky pathways of societal transformation with bureaucracy driven protocols. The conference distinguished between the big 'E' (evidence of what works or not) and small 'e' (evidence about performance) and the interaction between these. The conference focused on participants' own cases – based on artefacts (the protocols, processes, mechanisms and tools) used by organisations to assess results and generate and use evidence.

Participants discussed four questions:

1. What is 'the politics of evidence' – factors, actors, artefacts? And why is it important?
2. What are the effects on transformative intentions and impacts of potentially useful approaches on evidence of and for change?
3. Under what conditions do these practices retain their utility rather than undermine transformational development efforts? What factors and relationships drive the less useful practices and protocols?
4. How are people strategizing to make the most of what the results and evidence agendas have to offer transformational development?

Participants discussed the accumulative interactions between the global trends that had given rise to the results and evidence agendas, such as the increasing marketisation of aid, changing resource flows, and shifts of paradigms including a stronger adherence to business as the route to economic growth.

Participants noted the main positive effect on transformative intentions was to encourage more critical reflection in planning programming. Negative effects included the questionable ethics of certain demands, the unclear utility of some artefacts, wasted resources, and 'sausage numbers'.

Participants shared strategies for reducing the perverse effects of evidence artefacts and for enhancing their use for more transformative effect. Recognising one's own power to make a difference, through either resistance or creative compliance, was considered a critical first step. Secondly, to understand the contexts that generate the promotion and use of evidence artefacts helps influence their effective use and critical reflection. Building collaborative relationships and stronger organisational capacities to engage meaningfully with evidence and results artefacts were also areas where participants had usefully invested efforts.

More evidence is needed about the 'politics of evidence', in particular how it is being experienced by grassroots workers and mid-level government staff. More examples about the utility of certain artefacts are also needed, as are ways to hold organisations to account about the utility and relevance of required protocols.

Table of Contents

Preamble

On April 23 and 24, 2013, one hundred thoughtful and engaged development professionals converged from across the globe, including those working on the ground, in head offices, in consultancies and research institutes. The Politics of Evidence conference provided an opportunity to share and strategize *for people working on transformative development*, and who are trying to reconcile their understanding of messy, unpredictable and risky pathways of societal transformation with bureaucracy driven protocols. They have struggled to make sense of the shifting sands of the results agenda – seeing the wisdom in some aspects while actively questioning its less useful, sometimes damaging consequences.

We designed the conference to make the most of participants' experiences and ideas. Hence everyone had the chance to share these, including documented case studies from about a third of the participants. As Lawrence Haddad comments in his blog on the conference, power pervaded these stories and we hope that the interactive conference process will have given participants courage and confidence to adopt and develop further the strategies and tactics (discussed and developed in break out groups and shared in the final plenary session) to make programming and evaluative practice more fitting for transformative development.

The Big Push Forward was initiated in 2010 as a platform to discuss a topic that many experience on a daily basis. December 2013 marks the end of its present phase. Conference participants expressed appreciation of the initiative and want to continue 'pushing forward' more strongly and suggested that BPF 2.0 could include:

- a virtual space for dialogue
- an email list such as the outcome mapping community
- creating groups for advocacy with specific donors
- spaces for collective advocacy
- thinking about how to influence donor discourses at various levels
- being able to connect with others when needs and questions arise
- other forums for messages to the broader development community
- PhD students linked through a shared interest in the topic.

Those wishing to initiate a Big Push Forward 2.0 will need to (1) harness people's energy to engage as well as their wish to vent; (2) foster supporters' confidence in personal agency and power; (3) provide resources; and (4) help identify how to make the most of what the results and evidence agendas have to offer. All these need to be supported by political and power analyses. BPF 2.0 can build on the discussions summarised in this report.

Rosalind Eyben, Irene Guijt, Chris Roche,
Cathy Shutt and Brendan Whitty
September 2013

1 Acknowledgements

Funders were crucial for this event and the entire Big Push Forward initiative that underpinned it.

3ie
Australian National University
Big Lottery, the United Kingdom
Comic Relief, the United Kingdom
Hivos, the Netherlands
ICCO, the Netherlands
In-Develop, Sweden
La Trobe University, Australia
Mercy Corps, United States
MISEREOR, Germany
Murdoch University, Australia
Oxfam America, United States
Oxfam International
SDC, Switzerland
Sida, Sweden
WaterAid, United Kingdom
Two grants from individuals who prefer to remain anonymous

We also gratefully acknowledge the administrative and communications support received from the Institute of Development Studies in hosting this conference and for the voluntary services offered by some of its students.

Eleven reporters gave their time to document discussions and together create the basis for this report. Their efforts were essential: Anna Roche (coordinator of the reporting team), Itil Asmon, Bodile Arensman, Therese Brolin, Rebecca Carter, Kate Newman, Chris Speed, Yang Ting, Josephine Tsui, Paola Velasco and Mike Wang.

Participants make or break conferences and the energy of those attending enabled much push back and push forward during sessions. Most participants shared rich, varied and frank cases, a small selection of which is summarised in this report. These cases provided the basis for identifying similarities and differences and taking us beyond simple generalizations about the results agenda. Many thanks for all your efforts and willingness to reflect critically on your own and each other's politics of evidence.

2 About the Conference

September 2010 saw the first gathering, at the Institute of Development Studies, of people concerned that the results agenda was reducing the space for assessing transformational development. From this, grew 'The Big Push Forward' in April 2011 expanding to include debates on design and reporting, and on the complex relationship between results and evidence (see Box 1).

In March 2012 convenors conceived the conference 'The Politics of Evidence', which was made possible through the generous support of donors who dared to take on the topic (see Acknowledgements). The conference focused on four questions:

1. What do we mean by '*the politics of evidence*' – factors, actors, artefacts? And why is it important?
2. What are the *effects on transformative intentions and impacts* of potentially useful approaches on evidence of and for change, such as Theory of Change or Value for Money?
3. Under what conditions do these practices *retain their utility* rather than undermine transformational development efforts? What *factors and relationships drive the less useful* practices and protocols?
4. How are *people engaging* with problematic practices and protocols? What are they accepting and doing, what are they resisting and how? *What alternatives* have they found to create spaces for approaches more aligned with transformational development?

On Day One, participants' experiences were used to generate insights about questions 1 and 2 (see Figure 1). They shared a wide range of protocols, procedures and mechanisms, collectively referred to as *artefacts* and in group discussion highlighted consequences of these for their work. On Day Two participants explored the drivers of the context-specific positive and negative effects of these artefacts (Question 3). They shared details about tactics and strategies they are using to ensure the most utility from the results/evidence agendas (Question 4). This report shares the core insights from the discussions. As the conference operated under Chatham house rules, direct quotes as well as details in the illustrative cases are not provided.

Box 1. The results and evidence agendas

The pursuit of information on (intended) results and evidence of results to justify aid, improve aid and manage aid agencies through protocols, procedures and mechanisms for reporting, tracking, disbursement mechanisms, appraising, and evaluating effectiveness and impact

Section 2 introduces the 'politics of evidence', with Section 3 detailing the effects of the politics as shared by participants. Section 4 elaborates on the drivers of the effects. Section 5 summarises participants' strategies for making the most of the results and evidence agendas, before concluding comments in Section 6 about an agenda for the future.

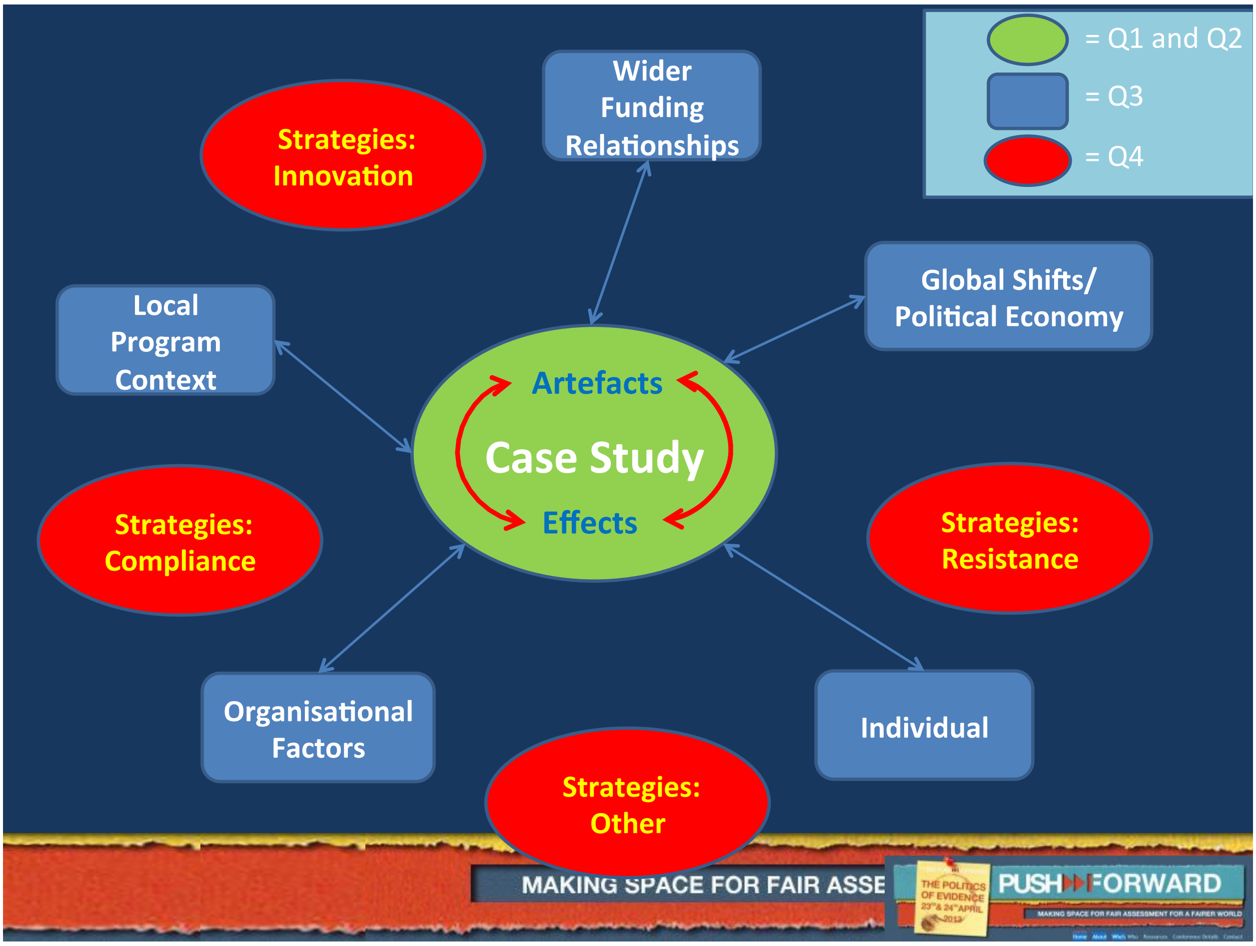

Figure 1. Framing the four conference questions

3 Introducing 'the politics of evidence'

Who decides what data is needed, how it should be collected – and why and how it is used as evidence? What drives their choices? Politics of what counts as good or credible evidence is deeply embedded in these questions.

What counts as good or credible evidence has not been a new discussion and has deep roots in philosophy and law. Yet in international development, debates continue to rage around 'which method is best' thus crowding out space for higher order questions of purpose, focus and participation in planning and assessment. Our conference focused on what underpins the debate on evidence, information and accountability, by considering the politics that shape what are counted as results or evidence, and the means by which they are generated.

At the conference, Rosalind Eyben introduced the notion of results and evidence 'artefacts' around which participants' cases focused.

> *"Results artefacts are used very widely within the sector for planning, implementation, monitoring and evaluation. They are reporting, tracking and disbursement mechanisms…. Evidence artefacts are … concerned with finding out*

what works best and therefore delivering value for money. These artefacts are used to document choice of intervention, for appraising proposals and for evaluating effectiveness and impact with respect to value for money." (Eyben 2013, p7)

Artefacts here refer to organisational processes and protocols based on formal and informal rules and norms. These artefacts are underpinned by assumptions of what counts as the right kind of result or evidence. The norms and values underpinning what is considered 'right' are rarely, if ever, explicit. Table 1 lists artefacts related to the results and evidence agendas commonly used in international development.

Table 1. Examples of artefacts to measure and assess (Eyben 2013; Whitty 2013)

Results artefacts	***Evidence artefacts***
Used very widely within the sector for planning, implementation, monitoring and evaluation. They are reporting, tracking and disbursement mechanisms that include: • Base-line data • Results reports • Progress reviews • Performance measurement indicators • Logical framework analysis • Risk register • Theories of Change • Payment by Results	Concerned with finding out what works best and therefore delivering value for money. These artefacts are used to document the choice of intervention, for appraising proposals and for evaluating effectiveness and impact with respect to value for money. Evidence artefacts include: • Randomized control trials • Systematic reviews • Cost-effectiveness analysis • Option appraisal • Social return on investment • Business cases • Impact evaluation

The politics come into play when certain artefacts are used, enforced, abused and distorted through power dynamics. Politics concern their interpretation, sometimes distorting the utility and intention of the artefacts beyond their original purpose. Politics imbue artefacts with a specific purpose based on explicit or implicit values of what counts as 'evidence' or 'results'. Politics are also present when certain kinds of artefacts are elevated above others, by describing them as more 'rigorous', 'robust', 'reliable' or 'scientific' – in the process disqualifying other kinds of artefacts.

The conference sought to unpack how artefacts and such words were used, defined and promoted by dominant players with particular values and norms.

Three panellists opened the conference by sharing examples of how certain artefacts became instruments of power. The first example was from a small NGO, required to hire consultants at its own cost to comply with mandatory proposal requirements - after approval of a concept note. The reworked proposal distorted original intentions to fit the donor's vision, unrealistic timeframes, and diverting resources destined for staff-focused learning. The project became 'a stranger' [sic] to the NGO.

The second panellist discussed an impact study costing over 1 million dollars brought the best of both the experimental and anthropological perspectives to bear on a micro-credit program. The study found it difficult to reconcile the two evidence streams. What was the 'truth' among diverse interpretations? As a result, managers at the agencies involved are finding it a challenge to change programming based on these findings, while the commissioning donor has reprioritised and no longer seems interested in this topic or the impact study findings.

By contrast, the third panellist spoke about eight randomised control trials (RCTs). Of the requests submitted to his organisation, only 3% were deemed appropriate for an RCT. A study of the afterlife of 8 RCTs found that uptake was influenced by factors such as: who was most interested in generating the evidence; choice made in research design; and whether efforts were made to translate evidence into policy-useful insights.

The opening panel highlighted some of the themes that were to return in many conversations, institutional dynamics, utility, personal costs, So how did other conference participants experience the politics of evidence embedded in their cases (see Box 2)?

Box 2. Examples of Cases Discussed

- Multi-year, mixed methods (RCT and ethnographies) by an international NGO consortium
- Quasi-experimental design to study impact of NGO sector
- Value for Money to manage a grant-making process for CSOs
- Building organisational M&E frameworks
- Evaluations – voluntary or imposed
- Activity review between partners
- Mid-term reviews
- Terms of Reference
- Building a community of practice around reflective results-based management
- Systematic review

People's experiences with artefacts ranged from the positive in terms of strengthening understanding and improving implementation, to the problematic in terms of distorting practice, harming relationships and disempowering those involved.

One positive example included an agriculture assessment system in which those involved had freedom to debate, was outcome-focused on a theory of change and process, and sought to understand contribution not attribution. In another, the person concerned was able to use the M&E tools given by the donors and turn these to positive use. Staff from one large health programme talked about how they were able to use their own contextually informed research and evidence about the challenges experienced by local women to challenge a donor's construction of 'the problem' and solution. Another positive example was from an NGO that collated an impact report from across the organisation, focusing on a theory of change and the steps in the process, involving communities in validating the results and using the BOND principles framework to assess what evidence was strong and weak.

Examples of problematic manifestations of the results and evidence agendas included that of a consultancy company working in one sector in a country with 240 separate M&E frameworks with 2000 indicators. When the consultant suggested one integrated M&E framework, the donor response was that 241 would be worse than

240 – each wanted to keep measuring in their own way. Examples from two other participants illustrated extremes in interest: one evaluation with a disinterested commissioner who only wanted a short slide presentation on programme results, and by contrast another programme that multiple partners constantly scrutinized with respect to the results achieved. A fourth example illustrated the tensions within organisations around an evaluation. While the donor and communities were keen on learning for improving the programme, the head office of the organisation was only interested in the randomised control trial. The evaluation manager was stuck in the middle trying to manage the different expectations.

While artefacts were the entry point for case presentation and debate, it was clear that the artefacts as such are not necessarily at fault. While some cases were clearly either positive or negative, many others revealed both pros and cons. Whether artefacts or tools create problems or benefits depends on how they are contextualized within projects, organisations or partnerships. Hence conference participants agreed that constructing the debate around a critique of specific artefacts was unhelpful, without more information about why and how they are used.

This section concludes with four observations that serve as context to the effects, drivers and strategies that are discussed subsequently.

1. Big 'E', small 'e' or the interaction of results and evidence. Participants confirmed the utility of distinguishing between the results agenda and the evidence agenda, as offered by both Eyben (2013) and Whitty (2013). Several participants commented that the organisations they know are often highly focused on 'small e' (evidence about performance and results monitoring) with high transaction costs, yet invest little in 'big E' (evidence of what works under which conditions for whom). Each has distinct uses, artefacts and audiences – yet interact within organisations in ways as yet not well understood.

2. Politics are everywhere. 'Politics' refers to more than just power relations; it drives organisational dynamics, social relations, trust, loyalty, strategic decisions and mindsets. Talking about power as only exercised by a 'results beast' and 'the donor as an evidence wolf' was considered unhelpful by participants, and incorrect. The results agenda is not driven solely by donors nor are donors homogenous. The forces at play (see Section 5, 'Drivers and Conditions') extend beyond donor relationships, and include public commitments set by Parliament, and the internal politics and interests within those receiving donor funds. Furthermore, being a donor is a function, not a certain type of organisation – each actor in the aid chain sees their funder as 'the donor'.

3. Development as complex. The extent to which development is considered complex influences how organisations and individuals see what needs to be known and done. Many participants believed too much time is spent trying to look for linear cause-and effect with artefacts designed on erroneous premise that development pathways and impacts are knowable in advance or definitely knowable ex-post. Participants voiced the need to keep pushing back on demands for attributable impact, evident in the desire to portray aid as directly delivering results. Participants also gave however examples of how artefacts can deal with the complexity of change.

Case 1: A shift to simple numbers

Sida is asking for quantitative information – the minister for external affairs is particularly keen on this. Sida has changed a lot over the past 7 years. It wants evidence of the results of their intervention, e.g. with this £1b which they will spend in Zambia over 4 yrs they will provide X jobs, reduce child mortality etc. Sida is embracing management by indicators – top down. This version of accountability has completed shifted the results agenda towards Swedish tax payers.

4. Differences vary per type of activity and across countries. Some participants noted that the results agenda appears less extreme for campaigning, the domain of activists. Yet while not expected to demonstrate direct causation, the need for campaigning NGOs to focus on quantifiable results was changing how activists perceived campaigning. Participants also noted strong variation in relation to national politics, affecting whether the results agenda is ideological, or just a temporal reaction to the current recession. For example, the United Kingdom and Sweden are currently both characterised by a strong drive towards taxpayer accountability and a belief that results based management strengthens this.

4 Mapping effects: Helping or hindering?

Artefacts themselves (tools, methods, protocols) are not the problem. Participants' cases showed that artefacts have been used cleverly and ethically, as well as in ways that waste time and money and harm development intentions. Whitty's paper, based on a short crowd-sourcing effort, clearly highlights the mixed effects of the results and evidence agendas. Conference participants affirmed this in many of their cases, naming beneficial effects on organisations and their on-ground work, alongside hindering efforts. For example, in one southern African country, the government ministry introduced local performance assessment framework across all districts. Positive was the fostering of inter-district competition to achieve higher rankings and self-reflection in relation to peers. However, the participant presenting the case mentioned the risk of political abuse with low scores used to remove political rivals, and less attention to service delivery due to the focus on resource disbursement in the performance framework.

Notwithstanding the crowd-sourcing effort and participants' cases, conference participants commented on the relative paucity of information about the real, on-the-ground effects of the results agenda. Our collective understanding of effects was limited by the absence of two key groups in particular:

(1) grassroots workers in direct contact and implementing activities with citizens; and

(2) mid-level staff at various levels who bridge the gap between grassroots workers and senior management from whom results and evidence requests emanate.

4.1 Positive Effects

More reflection and *rigorous thought* were mentioned as positive effects in some of the cases. Greater emphasis on evidence has led to more intelligent consumption of data and use of artefacts by those in international development, which works in two directions. First, conference participants mentioned being more open to results and evidence, which fostered learning and helped to rethink strategies, priorities and practice. For example, several mentioned that working with 'theory of change' has helped clarify change intentions and explain these better. Some myths have been challenged or even debunked, although not always unambiguously or unequivocally. Methodological advances have been made and theoretical insights gained (also see Whitty 2013).

Case 2: Positive use of a baseline survey

A donor programme was supporting a West African government expand and increase access to health services. Programme implementers talked about how great the programme was, how girls had benefitted, but produced no evidence, and could not illustrate the social stratification of the girls who had benefitted. The M&E unit spotted a problem and an opportunity and used a baseline survey, very thoroughly, to illustrate how much social stratification mattered for health service use and the government.

Second, some participants found that being asked to use certain artefacts, sharpened their understanding of the artefacts in question, becoming clearer about their appropriateness and utility and more knowledgeably challenging imposed artefacts. Examples were given from 'value for money', quasi-experimental impact evaluation approaches, monitoring protocols, and systematic reviews.

More specific positive effects mentioned include:

- Creativity as a by-product of using certain artefacts, generating methodological innovations.
- Collection of cost data has allowed hitherto hidden costs to become visible;
- Process benefits, engaging with evaluative processes can strengthen relationships and 'create membership';
- More nuanced terminology and the benefits of connecting through language around the results/evidence agendas.

Participants mentioned two broader effects in discussions (also see 'Drivers' section): the rise of transparency and renewed interest in international development.

They noted that the rise of *transparency*, the ethics of tracking and sharing process and outcomes, had positive effects on internal and external accountability. However, the ascendancy of transparency is perceived as a double-edged sword. The obligatory nature of transparency also inhibits sharing of findings, particularly critical assessments, due to fear of reprisals and funding consequences. Participants mentioned the fabrication and sharing of 'sausage numbers' as a problematic manifestation of a generally positive trend.

Taking a broader perspective, some participants felt that more attention on results and the results agenda has *refreshed interest in international development* as a whole. International development has caught the attention of the media to give development a more global platform than before.

4.2 Problematic Effects

Losing sight of questions. Trying to generate and understand cost-effectiveness and demonstrate results over-rides questioning why development efforts were initiated and whom these were intended for. Losing sight of the critical questions is compounded by the tendency of people to latch onto specific favoured tools. Uncritical use of preferred tools can short-circuit thinking and obscure the lack of clarity of what exactly it is you are trying to understand.

A good example has been the recent uncritical interest in 'theory of change'. Some participants felt this had strengthened critical thinking about how change happens and how one's efforts seek to affect change but others noted that even with the surge of interest in theory of change as an artefact, thinking processes remain obscured.

Case 3: From meaningless aggregation to community validation

One organisation was under great internal pressure by the board, directors and donors, to look at global impact and was introducing global indicators. When these indicators were linked to each local programme, the findings were meaningless. They are now in a process structured around a theory of change that engages the community in validating results. This has led to some agreement and some challenging, with community opinions shared in reports. Community challenges generated far better organisational learning.

Transparency is not aided by 'sausage numbers' (see Box 3). The move towards transparency and openness in international development is important, but only if learning for improvement is not forgotten. Some of the politics play out when transparency trumps all and 'sausage numbers' are generated to meet demand, irrespective of accuracy or utility.

Box 3. Sausage numbers

Building on the adage that 'laws are like sausages — it is best not to see them being made', 'sausage numbers' became a metaphor during the conference for data of unknown origin and dubious provenance. People do not consciously understand what they are entering into a data system. 'Sausages' are the result, with users unclear about what has gone into producing them.

Ethically, many concerns were voiced related to the exclusion of staff from key processes, such as recounted by the first panellist (see section 2 above). Others confirmed this experience, mentioning imposed indicators that do not adequately reflect their processes and outcomes. Several participants also spoke of their deep concerns vis-à-vis their partners: 'We don't want to throw our reporting requirements over the fence to our partners.' In some cases, demands for global indicators, for example, from donors or senior management were not passed down to protect implementers. Intermediary staff or organisations took on the burden of translating these into more meaningful metrics.

Research integrity was also felt to be compromised when artefacts were misused or abused, aggravated by reductionism and simplification of findings. Results are negated or denied. And who owns the data and who can publish it?

Participants repeatedly mentioned the *time and money they waste* in negotiating with funding agencies about the appropriateness of imposed artefacts and in being forced to generate sausage numbers, with limited time and energy left to engage with more important concerns. In several cases, imposition of inappropriate artefacts led to their flawed use, draining employee morale and straining partner relations. Ethics were strained in one case, as the choice of method (an RCT) led to debates about 'contaminating' the control group when a drought struck in the research area.

Critical in the discussions was the fear of a *narrowing focus of development,* due to the increased demand for certain kinds of results and evidence. Distortion of findings was also noted in contexts where evaluation or reporting processes made it impossible to convey local voices and needs. That which is easy to evaluate gets evaluated, crowding out that which is harder to ascertain. These missed opportunities to explore meaningful questions constitute a failure to advance the development agenda and leads as well to a narrower understanding of development.

The *inhibition of learning*, was repeatedly mentioned, a profound paradox given what some participants thought was the core intention of the results agenda to enhance learning. Limited or no learning was linked to inappropriateness of method and of metrics, as well as insufficient space and time for reflection. The distortion caused by the drive for aggregation of data led to decontextualized – and therefore meaningless – results.

Case 4: A defective M&E system

One organisation reviewed its agency-wide metric system and found that:

1) Numbers at aggregate level do not reflect what has actually been implemented; what those numbers mean remains in question. A gap exists between people's desire for data and their ability to make decisions with that data.
2) The system focuses on utility-based conversation with no evidence of *how* changes take place.
3) People do not want to see low numbers and are defensive about them, making it hard to get evidence for measurement.

Simplistic interpretations of important concepts with potential for strengthening efforts were mentioned, notably 'value for money' and 'theory of change'. In the worst case, value for money (VfM) has come to mean low cost, leading to frugalness over quality and adding to resource waste. Conceptually, certain principles of evaluation practice are not critiqued enough. Participants referred specifically to how '*independence'* (both for results data as well as impact evaluation) has become a non-negotiable criterion and synonymous with high quality. More recognition is needed that independence can have negative effects that hamper learning. In one case, a large impact study of the NGO sector was framed around a total separation between the commissioning NGOs and the commissioned researchers. The country-focused teams were not able to interact with the researchers contracted to undertake the synthesis, so each generated its own conclusions.

Utility of artefacts is forgotten. What is the cost-benefit of data production and evidence generation? Participants noted that many artefacts have not yet demonstrated their utility. Decisions about policy or programmatic priorities and resource allocation rarely seem connected to the request for results that demonstrate effectiveness.

Failure is painful and financially dangerous within international development, with participants noting that the results agenda has led to cases of *positive stacking of results.* A secure organisational environment with robust relationships was considered a condition for learning from mistakes. Even in organisations where rhetoric supports learning from failure, it is still highly risky to admit to failure.

The *standardisation of artefacts* within organisations was critiqued for hindering localised adaptation and utility. In theory, it should be possible to explore the systems and procedures that those receiving funding already have and then work with the systems. But the process of tailoring organisational systems to work rigorously and reflectively with results and evidence needs support, trust and space for negotiation. Hence, as reported in different cases, it is difficult and oftentimes those involved feel frustrated.

So what causes certain artefacts to hinder better learning, quality and accountability? Section 4 considers the drivers of the results and evidence agendas, and people's and organisational responses to the effects mentioned here.

5 Drivers and Conditions that Help and Hinder

Conference participants used the initial framing of five kinds of drivers and conditions (see Figure 1) to reflect on their own circumstances.

5.1 Shifting global political economy

Shifting money flows are changing dramatically and redefining power relations. Newcomers are strongly questioning dominant evidence cultures and artefacts, bringing in their own norms and data priorities. The financial crisis has increased pressure from others (including Parliament) to show that they are making a difference, with preference for simple metrics.

The traditional bilateral/multilateral/NGO triumvirate now represents a rapidly reducing percentage of national budgets in recipient countries. Bilateral aid budgets have seen significant drops in, for example, the Netherlands, although with increases in Australia and the UK. In all cases, aid budgets are much more intensely scrutinised, given the global financial crisis. Strengthened local economies in Africa, Asia and Latin America – and more fragile economies in Europe and North America – are shifting power relations. The large foundations such as the Bill and Melinda Gates Foundation, also play a strong role in shaping the results and evidence agendas. Commercial companies, riding on the wave of value-driven consumer demands, are being scrutinised for their impact on local wellbeing but are also influencing through what participants called a marketization of aid (for example 'value for money'). Remittances are three times larger than aid flows, while Brazil and China as 'aid newcomers' bring in different historical traditions of evidence and definitions of 'effectiveness'.

5.2 Organisational competition, performance pressure and data greed

Evidence and results artefacts are the gateway to next rounds of funding in a sector under fire, particularly for non-government organisations. Competition between NGOs due to reducing aid flows, the related fight for survival, and 'brand risk' converge to inhibit admitting failure and reducing sharing. Inter-organisational competition between NGOs in particular, has also reduced collective action to influence the agenda on their terms.

The culture of 'targets' and ongoing over-inflation of expectations maintains the pressure on NGOs to grow and succeed. They are unwilling to alienate large back donors and often do not push back when asked for unrealistic targets or inappropriate artefacts. These effects are compounded by: (a) the push to scale up and replicate; and (b) the need for aggregate results for sensemaking, communication and fundraising. Yet targets, evidence and results are crucial information for organisations with more proposals than they can fund.

Case 5. Lacking confidence to challenge

One INGO with funding from DFID to do programme work in nine countries required countries to develop their own logframe nested within a global one. DfID shared an example of a quantitative logframe despite the fund manager explaining that DfID were interested in qualitative data. The NGO undertook a baseline household survey, repeated at mid-term, and project end, each time with different households making the data hard to contextualize. The NG0 wanted to show DfID it had numbers in the logframe and country offices spent much time working on this. UK-based programme staff feel they do not know what is happening at the country level as they focus on info on budget and logframe information, rarely having time to talk to people involved. The NGO has been using the artefacts without creating space for learning and empowering in-country staff. DfID is not interested in learning and the NGO itself has adopted the results based ideas and logframe so seriously because of its own concerns that it doesn't have the evidence needed, yet doesn't really understand evidence, and is worried about this. People's lack of confidence makes them unprepared to challenge what is happening.

Participants noted the *organisational 'greed' for narrow and simplistic* data. Despite a greater interest among donors in social accountability, the results and evidence agendas are driven by an upward accountability dynamic in aid of decision-making and strategic thinking by senior management. Partner relations have suffered when partners are asked to collect data of no direct benefit to themselves.

Another driver is the *lack of knowledge or capacity* that accompanies data greed. The demand for information is compounded by what participants feel is a widespread lack of capacity, especially by senior management to explain the purpose of information and to use results and evidence wisely.

Participants mentioned the *gap between espoused theory and actual theory in use*, leading to mixed messages within organisations. Participants gave examples of conflicting sets of artefacts used in some organisations that required negotiations to establish which set would be prioritised. One increasingly common example is the tension between the need for much quick information on results, small E, and the time-consuming investment in impact evidence.

Finally, two other organisational drivers were mentioned:

- The lack of budgets for learning, either through core funding or dedicated program funding, with few donors valuing learning enough to pay for its real cost;
- The maturity of organisational relationships, particularly between funding and implementing organisations, with more difficulties related to results and evidence present when parties are new to each other.

5.3 Peculiarities of the development sector

The sector is caught in a vicious circle lead by *promise inflation, inflexible targets and fear of failure*. The non-negotiability of targets hinders adaptation en route, although results data may suggest the need for strategic rethinking. Data and results reports can make staff and policy makers feel in control. What is needed to make organisations less nervous and more flexible? Or are we dealing with human nature, as some participants suggested?

Despite a sector-wide motivation to advance thinking and practice, many development organisations find it *difficult to prioritise learning*. One recurring critique is the strong focus in results information (small 'e') is not matched by a strong interest in evidence on what works for whom under which conditions (big 'E') – although this is shifting rapidly with the explosive growth of impact studies. Multiple funding relationships, each with its unique set of artefacts, reduce time for learning. The more results data that needs to be pursued, following different protocols and calendars, the less time and space remains for reflecting on findings.

Case 6. Good use of data for managing strategic priorities

One example where results data is helpful for management was from Uganda, where one organisation has been collecting information on FGM through its case management process. It found that over the 10 years that it was working, levels of FGM had decreased so much they didn't need to work on it any more.

International development is ideas-driven with a penchant for *idea fads*. These fads relate to entire strategies, such as micro-finance or ICT with fluctuating donor interest not always evidence-based. Faddism also extends to results and evidence artefacts, e.g. the rise and fall of the similarly intentioned but distinct artefacts of structural analysis, logframe and 'theory of change'. The artefacts keep changing, yet their application becomes more standardised and mandatory.

5.4 The power of the individual

Participants spoke strongly to the power of the individual. They discussed this mainly in terms of their own agency to make the most of the results and evidence agendas (see next section). Some participants shared cases in which the power of individuals – as head of agencies, their evaluation units or their boards – forced the inappropriate or ineffective use of certain artefacts. But in other cases, the power of individuals was mentioned in a positive light, as instrumental in ensuring clear and creative use of certain artefacts or pushing back.

Consultants were mentioned as powerful in terms of shaping which artefacts were used and how. Problematic is the unregulated space they occupy in the results agenda, although members of evaluation associations are – in theory – bound by the association's professional standards.

The next section on 'Effective strategies' focused strongly on the power of individuals.

6 Effective Strategies

The diversity of cases, contexts and relationships surfaced an equally wide *diversity of strategies* to deal with the challenges discussed above. Participants had sometimes successfully adopted diametrically opposed strategies with similar cases. Thus this section serves as a source of ideas, not as a list of consensus 'should-dos'.

Some suggested strategizing around the image of *concentric circles of influence*. The *innermost circle* involves being more explicit about using our own agency and becoming more aware of the power we have and leveraging this for improvements. The *next circle* is about engaging one-on-one with others to develop creative relationships of trust. *Third* is the outer circle of organisational structures, involving influencing within organisations and through coalitions between organisations. But many other strategies work across these circles of influence as set out below.

6.1 Recognizing personal agency and power

The most mentioned strategy related to *personal agency* to shift internal discussions and (re)interpret artefacts. Our ability to use the results/evidence agendas positively makes us 'activists' in our relationships with organisations, funding agencies and partners. Our agency is sometimes held back by fear of the oft-mentioned funding cuts. But this might not be as strong as we make it, given that funding agencies also incur risks and shy away from the bureaucratic effort required to break contracts.

Personal agency brings us to *power*. To what extent are we blind to the power we have to make change; the power to challenge, agree, question, and create? Some participants wondered whether we are, at times, more complicit than blind. Or perhaps we do not understand our power to change or simply do not use our power appropriately?

Box 4. Wield expertise

Expertise, whether around small E or big E, comes in various forms, resides in different individuals, literature, and methodological options. Strategies mentioned included using the weight of literature that backs up a certain paradigm to strengthen arguments for certain approaches in certain contexts (for example the Stern et al 2012 paper is being used often for this very purpose). Other options mentioned included:

- Engage constrictively with people with different epistemologies by working with them
- Use our own close knowledge of examples of what worked and has not, adapting for context
- Identify and share good practice internally
- Build our own expertise if we are weak in key areas
- Back up the rigour, credibility and validity of any options you suggest, framing the arguments based on understanding of the position and language of others
- Borrow from academic research conventions that enhance quality and lay bare limitations
- Acknowledge doubts about data.

Power analysis – internally and externally – was mentioned often as part of improving our personal agency. Exploring one's own power, what power might need to be 'borrowed' (through relationships) and working with others to see and use their power to shift the politics of evidence production and use.

Exercising agency means *making conscious choices about how to respond* to the results and evidence agendas – resistance or engagement? Participants wondered whether this debate was another ephemeral development fad or if it warranted serious engagement. Deciding what is and is not worthwhile, what can and cannot be changed is important to for considering if and how to engage. Is it a matter of riding out the results/evidence wave, waiting for pivotal people to change and for artefacts to come and go? Or do we just need time to adjust, accepting perverse consequences and negative impact in the early stages?

Engagement enables us to influence to reduce the formulaic used of tools and the tick-box pressure. It gives us the option to challenge our own assumptions and learn, and when appropriate, try to shift power relationships and the way in which artefacts are used. Oftentimes it involves bridging worlds and becoming multi-lingual, adept at the terms and values of each world.

6.2 Creative compliance

Compliance, creative or otherwise, goes hand-in-hand with personal agency. Compliance was mentioned as 'opting out of challenging' or cynical compliance. But participants also referred to *creative compliance*. The strategy suggested repeatedly was 'use the label' but focus the content on critical reflection rather than the formulaic.

Focusing on what can be shifted, particularly in informal places, means accepting some ideas and requirements, even if ineffective. Particularly those working to change organisations or large programmes commented on the need to making the most of 'tipping points for opportunity' by working in nuanced, non-threatening ways. For example, if an organisation asks for a manual, do this while keeping focused on how to support more rigorous thinking. Use existing systems and standards to support your ideas (see Box 5).

Some participants consider seeming or partial compliance an essential starting point, a foot in the door for dialogue about assessing others recognized that compliance comes with a price and has an inherent tension. How do we avoid becoming complicit in the harmful manifestations of the results agenda, and blind to our own power to change it?

Box 5. Use agreed standards to comply creatively

Associations of *evaluation professionals* have agreed evaluation standards, professional ethics and values. These standards support a fair, useful and ethical assessment process. The BOND checklist of good evidence, formulated collaboratively with many UK NGOs, was found by some participants to be helpful and easy to use.

6.3 Understanding the dynamics of artefact use

Identifying the artefact(s) around which tension or contestation exist is relatively straightforward. But to know how to leverage for change, we need to disentangle the organisational factors and sectoral dynamics that shape how artefacts are used in context (see 'Drivers' above). All organisations have *lists of values* that drive priorities. Understanding organisational values and how these are – or are not – reflected in the artefacts allows one to question anomalies and to discuss possible adverse effects of certain artefacts. It requires knowing the organisation's readiness to adapt artefacts to match with its values and what room to manoeuvre exists. Understanding the dynamics of their use in a particular organisational context may allow us to reframe artefacts, or guide (more) intelligent adoption of aspects of the results/evidence agendas. Other options include switching from unhelpful tools to those more amenable to/supportive of transformative development, and that help institutionalise critical thinking processes.

6.4 Seeking and nurturing spaces for influencing and learning

Resistance is an important strategic option. Yet many participants considered *engagement and dialogue* as more consistent with reflexive practice, giving opportunities to learn and challenge one's own assumptions as well as space to push more innovative or appropriate thinking. The widespread interest in M&E within international developments suggests that there is space (sometimes, not always and perhaps not much) to negotiate on most useful options.

Central to dialogue and learning is 'space' for conversation, with participants referring to 'political space', 'bureaucratic space', 'reflective space' and 'safe space'. We need to become more adept at looking for these spaces, which constitute opportunities for reflection and creative solutions. One such space mentioned was the mid-term review of a program, often formative and thus a learning opportunity (also see Box 6). However, when such a review is treated as

Box 6. Use formal mechanisms to create space

Dialogue in organisations extends to managing the evaluation process, including careful formulation of Terms of Reference; it offers options for creating and shifting the space for appropriate artefacts and their use. Any TOR with prescriptive methods without a clear theory of change should be questioned. Use that first stage to question, clarify and negotiate.

a bureaucratic upward accountability mechanism, the space for learning – especially about weaknesses- and making operational decisions to increase accountability to poor and marginalized people is considerably compromised. Where spaces do not exist, they need to be created – more difficult as it means changing the system.

Box 7. Organisational Tai Chi

Identify and name a strong organisational interest and harness that positively. In one organisation interested in theory of change, this interest led to senior management being engaged in developing an agency-wide theory of change, which contributed to more rigorous thinking. Other examples talked about Tai Chi in terms of connecting with a problematic process and redirecting it towards a more positive one.

Asking questions – particularly on purpose, audience and use – emerged regularly as one strategy. One case involved a learning-oriented approach to assessment, querying what the organisation needed, adapting the initial request and thus opening the doors to a conversation with the requesting funding agency. Another approach is to discuss how certain procedures or protocols might trigger good practice in a better understanding of the change process (see Box 7). Yet another is to provide constructive feedback about an artefact's effects while offering alternatives and comparing costs and time requirements.

'*Create space and processes of critical reflection and learning*' was a recurring theme though most recognized doing this well is difficult. Some people and organisations flounder when given too much open space to learn; tools can give people confidence and skills and generate knowledge.

6.5 Build collaborative relationships

Protocols and processes are nested in *networks of relationships*. Effective upward accountability to funding agencies and downward accountability to communities is built on these. Ownership of evidence and results processes are important in these chains (see Box 8). Organisations want reassurance about results and appropriate measurements. If you can show, using their language, that this is the case, then space opens up for dialogue and negotiation.

Box 8. Engage those expected to use evidence

To use evidence, you need to value it and believe it. Where possible involve implementation staff in evaluations, especially if it is clear that it is not solely about proving merit but also about improving an intervention.

Local capacity, context and ownership are vital for transformative development, hence, the importance of talking with local partners and finding solutions together. Local staff are not averse to facts and statistics; they want to know about their achievements and have them valued. It is possible to explore the systems and procedures that grantees have and then work with their systems when these are in place.

Particularly when sensitive changes are being tracked, trust is critical to use data for improvement-oriented feedback. Asking partners to report on a particular issue is not intrinsically problematic, if this is accompanied by assistance to make required internal changes. Proactively sharing such processes with a funding agency also can make then more amenable to accept the approach.

Participants commented on the strong tendency for organisations to seek global indicators and expect local partners to report against those, often with little/no local utility. An alternative is to seek common parameters within which there is adaptability at the local level. Some mentioned the value of equipping communities and partners to tell their own story. Technology can contribute to exposing contradiction between official organisational stories to the public and what people involved in programmes are experiencing. Innovations in communication, e.g. Twaweza in East Africa, have much to offer by enabling communities to be able to voice what matters in their lives (also see Box 9).

Box 9. M&E as intervention

Make M&E and feedback part of the intervention, rather than a separate and additional process. If accountability is built into the logic of a ToC, for example, then it boosts the likelihood of it happening in practice.

Collaborations and collective efforts also figured in participants' cases. *Collective strategising* can help improve understanding of the drivers of the results and evidence agendas. *Strategic partnerships* can be part of the solution, such as long-term partnerships between researchers and policy makers to fund and evaluate innovative solutions for methodological challenges. The BPF platform was mentioned as one such collective strategy.

Other ideas on collaboration, not detailed by the groups include: (1) the potential of collective organising of front line workers; and (2) strong community engagement that generates an appropriate evidence base and accountability.

6.6 Build capacity of senior managers and field staff

Collaborative relationships can be built by investing in capacity. Senior management was mentioned regularly as an important target group. This cadre of people is not always clear about what they need, the implications of what they request or how to use the evidence they receive. Participants gave many examples of the powerful influence of the evidence discourse that led senior management to pick up artefacts and insist their organisations use them without fully understanding them. As mentioned in Section 2, one of the opening panellists referred to a 97% rejection rate of proposals due to being inappropriate for an RCT. We can encourage more critical thinking by decision makers and budget holders. Using examples of success, stories of failure and theoretical frameworks have been used as strategies for building senior management capacity. One successful approach was to bring senior management face to face with front line workers and primary stakeholders.

Field staff skill development is also critical. However, capacity building is often equated with training, which in turn is often focused on using tools rather than encouraging the thought processes and questions. When questions are clear, then it becomes clearer when, if and how to use which artefacts.

Conversations and capacity building work well when supported by examples about better and different ways of working. Examples of artefacts that have been used to further rigorous thinking can be powerful, particularly if these have demonstrable utility, driving demand for more evaluative evidence.

If the sector is to learn deeply rather than proceed with business as usual, examples need to break out of the best practice straightjacket that is integral to the politics of evidence and extend to stories of failure (though conference participants were not keen on that particular term) (see Box 10). The problematic use of artefacts and of ineffective results and evidence processes also need documentation. One participant working in a funding agency commented "We are hypocrites when we ask organisations to be honest about failure but then we cut funding and we wonder why people aren't honest." Stories of failure can be useful internally as part of expectation management. One topic suggested for more documentation was on the cost-benefit of reporting and accounting and the effects on time for reflection and dialogue. Honest theories of change and what has (not) worked and why were also mentioned.

Box 10. Radical initiatives to confront inappropriate demands for results or evidence

1. Developing 'AID LEAKS' website to enable staff or partners to 'blow the whistle' on inappropriate demands and their negative effects on the development process. This idea recognized the difficulty in many agencies of having a more profound internal debate on some of the more problematic aspects of the results agenda.
2. Promoting a 'kite mark' for evaluators and purchasers of evaluations who ascribe to certain ethical, moral and developmental standards of monitoring, evaluation and impact assessment. This would seek to create a 'guild' who would promote these standards and in so doing publicall and confront poor evaluation processes, evaluators and in appropriate ToRs.

7 Future Agenda

Missing voices on effects. The Big Push Forward sought to initiate a discussion on the elephant in the room of assessment and planning, that is the politics surrounding how evidence on results, including impact, is generated and used in international development. However, we do not know enough. Participants commented on the need for more evidence on the effects of 'politics of evidence', in particular from those working in government and with citizens.

More focus on citizen feedback. Citizens are increasingly demanding information but how often is feedback given? Information is still demanded from the top-down. One convenor stressed persistent deeply flawed thinking about M&E, with organisations not seeking direct feedback from communities likely to lose credibility and support. Can this change with new technology and more experiences with social accountability?

Case 6. Firing the donor? The end of narrow accountability

Could rapid ICT developments and communication, herald a time when community members 'fire' their donors? For example, in Kenya there was a project which gave school-going girls cameras to take pictures in schools, and place these on the web. Photos of toilets were accompanied by comments such as 'DfID toilets are not so good'.

Understanding personal agency. Personal agency was evident in all cases and many strategies, yet its scope was not understood well enough. We are all part of a system in which results and evidence artefacts are created and required. Conference participants are expected to use these or guide their use. To what extent do we drive artefacts or do they drive us? What is our space for reflection and self-criticism, as players in the evidence and results field?

Learning and the link to evidence. If artefacts rarely fuel learning, what does? Can transparency be leveraged for learning, not just accountability? What changes and has changed as a result of knowing and using data? How can we learn how to use evidence better, suggesting the need for a 'politics of learning'? Conference participants suggested the importance of aligning who learns with who sets the learning questions.

Utility. The discussions on strategies illustrated people's ability to be rigorously relevant and creatively compliant. What is less clear is how the results and evidence agendas can make transformational development more effective. All people involved need to increase their capacity to understand the evidence being produced and to generate useful insights.

Data sharing. How can data sharing mechanisms lessen the burden on citizens and their organisations, and make knowledge generation more cost-effective? Are funding agencies prepared to share information and what are the barriers?

Ethics. Ethically, should those producing and providing the evidence on the ground be involved and benefit from the process? Few at the conference would argue against this. If so, what are the new ethics around evidence and results? Which values and which trade-offs are involved? How do we deal with the 'participation' concept that in the 1990s and 2000s focused on participation in planning?

Accountability about evidence quality and quality control. The quality of evidence is a new topic of interest, as is the need for high quality of quality control. Are donors that require certain protocols held to account about the supposed utility and relevance of such requirements? What processes do we have for holding donors to account when processes resulting from the results and evidence agendas are fraught, waste resources and negatively affect (transformational) development interventions?

New trends in development. International aid is changing extremely rapidly. What do new trends mean for the evidence and results agenda? How will the entry of certain philanthropic foundations and countries and their historical use of evidence influence what is allowed and valued? 'Value for money' is only the beginning of such shifts, with different understandings of accountability likely to affect the shape of evidence to come.

References

BOND 2012 *The NGO Evidence Principles* www.bond.org.uk/effectiveness/principles

Eyben, R. 2013. Uncovering the Politics of 'Evidence' and 'Results': A Framing Paper for Development Practice. Paper prepared for The Politics of Evidence Conference. http://bigpushforward.net/resources

Haddad, L. 2013 'The politics of evidence. A big step forward'. www.developmenthorizons.com/2013/04/the-politics-of-evidence-big-step

Roche, C. 2013. Organisational case study of an international NGO (INGO). Unpublished mimeo prepared for The Politics of Evidence Conference.

Shutt, C. 2013. Case study on value for money and the politics of evidence. Unpublished mimeo prepared for The Politics of Evidence Conference.

Stern, E et al. 2012 'Broadening the range of designs and methods for impact evaluations' **DFID** Working Paper 38 https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/67427/design-method-impact-eval.pdf

Whitty, B. 2013. Experiences of the Results Agenda. Draft Findings for Discussion from the Crowd-Sourcing Survey. http://bigpushforward.net/resources