

EXPERIENCES OF THE RESULTS AGENDA

DRAFT FINDINGS FOR DISCUSSION FROM THE CROWD-SOURCING SURVEY

Brendan Whitty

April 2013

www.bigpushforward.net



Cartoon by an AusAid staff member sketched at an M&E workshop in Canberra

CONTENTS

SUMMARY	iii
1. Aims and Approach.....	1
2. Methods	2
3. Quantitative Survey Data Analysis	4
3.1. Assessing the Extent of the Change.....	4
3.2. The Scorecard on Results	5
4. Qualitative Story Data Analysis	6
4.1. Artefacts and the Politics of Evidence	7
4.2. Strategies, Reactions.....	10
4.3. The effects: opposing perceptions	12
4.4. Conditions/circumstances	18
5. Findings for Discussion	22
Endnotes and References	24

SUMMARY

This paper discusses the data from an online survey, which invited visitors to the Big Push Forward website to give their perceptions of the impact of the results agenda on their working lives. It analyses the very different experiences and interpretations of the respondents as revealed through both 153 responses to the quantitative survey and 109 qualitative stories. Bearing in mind the limited and non-representative sample, the analysis discusses possible reasons for these variations, drawing on the concepts identified in a companion paper, *Uncovering the Politics of Evidence and Results*.

The study discusses the day-to-day practice of small-e evidence –results and targets in management of specific projects – rather than large-E evidence of establishing theories of change or broader development policies. The stories are about the nuts and bolts of the development processes and artefacts - the theories of change, results frameworks, reporting requirements and value for money rubrics. It is about what ‘e’ is being collected, how it is used, and to what effect. . Several points emerge from the analysis.

First, there are very different perceptions, both of the agenda overall and of individual artefacts. The quantitative data that showed while the impact of the results agenda was more positive than negative overall, a substantial minority reported mixed effects. Similarly, the stories presented conflicting perceptions of the same artefacts:

- if one story said that the pressure to articulate measurable results has promoted a desirable realism, then the next suggested it generated perverse incentives to pursue easy gains;
- one snippet may have shown the problems of collecting meaningless, over-simplified data, but it is immediately countered by another story which enthused about a new-found discipline in articulating results;
- for some, descriptions of donor-driven requirements to articulate causal chains encouraged useful discipline, while for others they curtailed transformative development.

The contradictory perceptions seem to be often in tension. Thus learning is often in tension with accountability; capturing the complexity in evaluation with harmonisation and reductionism; coordination of partners with constraining their freedom to adapt. An emphasis on one may exclude the other: thus, as one story says, “accountability trumps learning”.

Second, the devil seems to be in the detail. How the tensions resolve and the perceptions play out depends on how the artefact is communicated, managed and tailored to its context. The *fit* appears to be important: the fit of the artefact to the existing systems and capacity of the organisation, and also the fit of the artefact to the specifics of the intervention (e.g. its complexity, the number of partners).

Third, perceptions of an artefact seem to be affected by both people’s own circumstances and their relationship with others. The survey data suggests that those in M&E and management roles, who benefit from better data and more resources for their priorities, tended to be more positive than those in project implementation and mid-level roles (offering some support for the idea of the ‘squeezed middle’ of the companion paper). At the same time, the stories showed that the space for negotiation over the form and fit of the artefact seemed to shape perceptions. This space is inevitably a manifestation of power in the context of the specific relationships between donors and grantees.

These points are offered as material and research propositions for further discussion in the Politics of Evidence conference in Brighton on 23/24 April, 2013.

1. Aims and Approach

This is a background paper for the Conference ‘[The Politics of Evidence](#)’ that has been organised by the [Big Push Forward](#), an initiative seeking political space for development practitioners to assess transformative and locally owned development in appropriate ways.^{*} The paper presents the tentative results of a ‘Crowdsourcing Survey’ conducted between October 2012 and February 2013, which asked respondents to rate and assess the impacts of the results agenda through a quantitative survey and to provide qualitative stories and cases of impact. It presents the quantitative ratings of over 150 development practitioners therefore, of which 65 presented a total of 108 qualitative snippets of experiences or stories. It proposes interpretations of these results, for further discussion.

The background issues to this paper may be found in Rosalind Eyben’s companion framing paper, *Uncovering the Politics of Evidence and Results*, which sets out some of the issues of politics and power that form the backdrop to the results agenda, and seeks to identify space for “flexible and creative support of locally-generated and transformative change”.¹ The bureaucratic and political space in the development sector is partially defined by two separate but related administrative doctrines, whose history Eyben traces:

- on the one hand, Results Based Management (RBM), which developed from public sector reforms that use quantification and value for money to manage and control public sector interventions, and to drive for increased efficiency;
- and on the other hand, Evidence-Based Policy (EBP) with roots in the medical sector, and which seeks to develop de-politicised, technical best practices against which the medical profession could be held to account.

Eyben observes how these doctrines have been integrated into the development sector, and have been translated by a series of artefacts into the everyday realities of practice.

Although the two doctrines are intertwined, it is somewhat ironic that the evidence for the efficacy of RBM is at best mixed, both in the context of the development sector and more widely. Evidence of efficacy (costs cut and targets met) is often lacking, or is accompanied by dysfunctional effects and unfortunate surprises.² This paper draws on the perceptions of development practitioners about the effects of the results agenda.

In pursuing this goal, it is worth pointing out two issues about the scope of the paper. First, the survey invited responses on results-artefacts (which concern the planning, evaluation and monitoring of results) and evidence-artefacts (which are about finding out what works best) *in the context of particular interventions*. Very few of the stories reflect on knowledge processes that seek to inform development policy more broadly, outside a particular programme context. The discussion here focuses on evidence generated in the context of a particular programme, which dominates the majority of the stories (the top two quadrants of Figure 1).

^{*} I am grateful to Rosalind Eyben, Irene Guijt, Chris Roche and Cathy Shutt for comments on earlier drafts. We are also extremely grateful for all the survey respondents who took the time out of busy working lives to answer the survey, and provide such rich material for the discussion. I have tried to interpret the stories in the light in which they were presented. The opinions expressed are my own and do not necessarily reflect the views of the conference donors or of my colleagues.

Second, I am presenting perceptions from some individuals working in the sector who responded to a survey. This means that the stories are often relatively brief, with limited context. We at the Big Push Forward have discussed issues of power and the hesitation to discuss matters where funding is at stake, many of which are raised in Eyben's paper. The responses about perceived utility of certain tools may also suggest, from a more critical perspective, the 'internalisation of a self-disciplining agenda'. From a political economy standpoint it shows how the results agenda has given benefits to some groups within development, such as M&E staff, while being criticised by others. The paper will try to present both critical and 'flat' interpretations for discussions in the conference.

Figure 1: Scope of the Paper

	Results (reported upon / evaluated outcomes from a specific intervention)	Evidence (information on what works to solve a problem)
Management decisions relating to a specific intervention (survey focus)	Artefacts of RBM, accountability for results, learning	Evidence used to justify choice of intervention, including value for money
Policy decisions for class of interventions	Evaluations used to justify scaling-up from the specific (e.g. pilots) to general; for transferable lessons	Evidence used to inform government policies

Part 2 reviews the methodology used in the survey, in the qualitative and quantitative parts. Part 3 discusses the quantitative analysis, in relation to the effects of the results agenda and viewing it as positive or negative. Where relevant, this is broken down by respondent. Part 4 analyses the qualitative data, the stories, using the core conference questions as a framework (see Box 1). Part 5 offers concluding observations.

2. Methods

Crowd-sourcing experiences. The crowd-sourcing survey was launched on 31 October 2012 using the web-resource, SurveyMonkey. The survey had three parts: (1) basic information about the organisation and the respondents' role; (2) five multiple-choice questions to elicit a rating about the effects of the results agenda and the degree to which they were positive or negative; and (3) qualitative stories about 'what happened due to the results agenda and what were your experiences?' The full survey form is included in Annex 1. None of the questions were compulsory. The survey was advertised by email and through the internet, primarily using Big Push Forward contacts, but also through social media tools and linked to prominently on the Big Push Forward website. The survey was closed on 1 March 2013, during which time we received responses to parts 1 and 2 from a total of 153 respondents, 65 of which also provided at least one story in part 3. Since the quantitative questions in part 2 were not compulsory, not all of the 153 provided responses (147 responded to questions 11-13, and 143 responded to question 14, see Annex 1.)

Analysis. The survey responses were analysed in two parts, the quantitative (see Part 3) and the qualitative (see Part 4). There were 108 stories provided - 65 respondents provided at least one story (of these, 32 provided two stories and of these, 11 provided three). The stories varied greatly in length and in form. They were analysed using NVivo qualitative analysis software using the questions structuring the conference as the primary analytical framework and allowing further codes to emerge

(see Box 1). These codes were used to structure the subsequent discussion. The analytical process also included colleagues reading through the first draft and interrogating my assumptions and interpretations. The forthcoming analysis should be taken as triggering questions and fuelling interesting debate among participants to the Conference.

Respondents. As the analysis below notes, the position of the story-giver or respondent seems to be an important guide to their perceptions. In terms of the functional responsibilities in their jobs, most of the respondents identified themselves as being directly involved in implementing the artefacts of the results agenda: 55% of the respondents have monitoring and evaluation (M&E) functions while 22% described themselves as having learning functions (often in addition to M&E). A fifth (20%) identified themselves as having programme implementation functions.

In terms of organisational background, over half came from the NGO sector - 44% in total came from international NGOs and a further 8% from national NGOs, while 21% came from those providing services to the sector (consultancies 13%; independent consultants 8%). A further 8% of respondents came from universities and think-tanks, though it is unclear whether they are writing as observers of the agenda or as implementers of projects themselves (or both). Fewer come from the donor side, bilateral donors made up 10% of the responders, multilateral agencies 6% and foundations 1% (2 responses) – although as the forthcoming discussion shows, it is often not simple to make a simple donor/recipient divide, since many organisations are both.

The nature of this self-selecting sample of respondents is vital to understand the limitations of the scope of this paper. The group is dominated by M&E specialist and NGO-sector, hence the analysis should be read accordingly. Moreover, of the 153 respondents, 41% were subscribers to the Big Push Forward website, and therefore may be understood to be engaged, to some degree, in policy debates on the use of evidence.

Limitations. The nature of the respondents shape – and limit - the study. The responses are self-selected and despite a spread across different functions and parts, most respondents work in M&E and come from NGOs. Much fewer implementers and programme staff responded. none from the ‘ultimate beneficiaries’, very few from smaller civil society organisation at the ‘end’ of the aid chain or from governments receiving aid. This key limitation needs to be borne in mind when reading the observations below. There are limitations also in the nature of the claims that can be made on the basis of the data. The data is not representative of the development sector or any part of it.

Box 1. Conference Questions

Our aim is making participants more conscious of how power plays out in planning and evaluation processes; strengthen their capacity to deal with it; and gain the courage and confidence to navigate political space, maintaining or increasing options and putting pressure on the system to shift demands.

1. What do we mean by ‘the politics of evidence’ – factors, actors, artefacts? And why is it important?
2. What are the effects of these practices on transformative intentions and impacts?
3. Under what conditions do potentially useful approaches on evidence of and for change - such as theories of change or Value for Money – retain their utility rather than undermine transformational development efforts? What factors and relationships drive the less useful practices and protocols; and which enable evidence to be generated and used in ways that strengthen transformational development?
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?

Despite these limitations, the survey provides rich and arresting stories that offer propositions for discussion in the conference and for further research. This is what I have undertaken in the analysis here below.[†]

3. Quantitative Survey Data Analysis

The following sections outline the findings from the closed, multiple choice part of the survey. It focuses on two questions. Part 3.1 concerns whether there has been an impact from the results agenda, to what extent that impact has been felt, where and by whom. Part 3.2 asks whether that impact is positive or negative. Section 3 captures the input of 153 respondents.

The survey headlines can be summarised as follows. Those who responded appear to believe that the results agenda has had a significant effect and that it has been more positive than negative, particularly for learning but also on the ability of agencies to fulfil their mission. However, these general trends must be nuanced in various ways:

- Respondents noted significant mixed effects which are picked up on in the story analysis below. The overall picture is one of ambiguity and conflicting interpretations.
- Learning is not always translated into improvements in mission, confirming an old story for the sector and one the results agenda appears not to have been changed.
- There are interesting variations depending on the standpoint of the respondent: as a group, project staff report less change and are less positive than senior management or people with M&E functions.

There are various interpretations possible for these findings – one is that the results agenda is overall ‘a good but mixed thing’. But does a positive opinion depend on the respondent’s standpoint, given that management staff are empowered by the data and those benefiting from the influx of funds for their functions appear more positive than those subject to what may be experienced as increasingly stringent controls? And how far are the responses affected by the disciplining effects of the discourse which has swept the sector? Is it simply that the values have been internalised by many respondents, with those harbouring negative views forming a squeezed middle of truculent programme officers (as the framing paper suggests)? These interpretations are presented as propositions for further discussion in the debate.

3.1. Assessing the Extent of the Change

Respondents overwhelmingly agree that the recent change in emphasis have created a change within the development sector. Over half of the responses thought there had been *significant or greater change* in their daily work (56%) and a further 24% thought there had been *some change*. When asked whether there had been changes at an organisational level, only 3% reported *no discussion* about the results agenda; 15% though there had *been discussion but no change*; and 62% thought there was *significant discussion or change* in their organisation. These figures suggest that the **results agenda has had a significant impact** for respondents.

[†] I have a background as a practitioner and consultant in development and I am now doing a research degree on results-based management at the University of East Anglia, UK. I am not an outsider, but am caught up in the same power relationships as the respondents. I have conducted evaluations for donor agencies and worked for various other development actors.

We asked respondents to prioritise the primary and secondary aspects of their work where they thought change had happened, in order to understand what kinds of changes occurred. Answers indicated that changes have been mostly in the nature of the interaction with funders – reporting (25% most important, 12% second most important), proposal writing (21% most important, 8% second most important). Relatively few listed changes in the details of either how or what implementation takes place as their priority (7% first choice for both how and what implementation, 13% second choice for both). This suggests that respondents indicate **the main change to be in accountability and reporting rather than content of programming**.

We asked respondents to rate the extent of change to their daily work and within their organisation (divided by function, by seniority and by organisation). Respondents identifying themselves as having an M&E or learning function tended to see a greater change than those with project functions (respectively 61% of respondents with M&E function, 57% learning and 45% project function seeing considerable or greater changes).

From the perspective of the place in the organisational hierarchy, seniority appeared to affect those reporting significant or greater discussions and changes within their organisations: senior management reported (75%) and internal technical or policy advisers (71%) saw greater changes than external consultants (52%) or mid-ranking staff members (41%). When talking of the effect on daily work (rather than changes in their organisation as a whole), the standpoints were similar but the differences less pronounced (senior management reported 61% considerable or greater change; technical or policy advisers 58%; mid-ranking 53% and external consultants 47%).

Taking the functional and seniority survey data together, it suggests that **senior staff and those with M&E / learning functions believe there has been greater change, both within their daily work and - more pronouncedly - within their organisations, than mid-level staff or those with programme responsibilities**.

From an organisational perspective, it was difficult to pull out much of a pattern about the perspectives of the extent of change on the results agenda. There was a slightly greater tendency for respondents in think-tanks, consultants and national NGOs to report less change than INGOs, multilaterals, bilaterals. While this tendency is not particularly pronounced, it does suggest again that **those coordinating many grantees or partners found the agenda to have generated greater change than those being coordinated**. Even amongst those, the disjuncture amongst bilateral respondents between their personal work and the internal rhetoric raises a tentative research suggestion that there **may be more noise than change** (71% report significant internal discussions and change at the organisational level, but only 43% report significant or greater change to their daily work).

These are tentative findings forming research propositions for further research and which we hope will frame discussion to be taken forward in the conference.

3.2. The Scorecard on Results

In addition to most respondents widely agreeing that there had been an impact, they also noted that **the impact has been more positive than negative**: 42% of responders identified “some” or a “great” improvement in their ability to achieve their mission, while only 22% reported a negative impact. The impact on the ability to learn was reported as being even better: half (50%) reported some or great improvements in their organisation’s ability to learn, while only 12% reported negative impact. While the positive here outweighs the negative, two additional points are noteworthy: first, that the

benefit from learning is not always translated into a similar benefit for the mission; and second, that a significant number of respondents identified mixed effects (i.e. both positive and negative) for mission achievement and learning (33% and 29% respectively).

How are the positives and the negatives responses spread across function, seniority and organisation-type of the respondent?

Breaking down by function, M&E people were considerably more positive about the benefits of learning (63% positive, 26% mixed, 3% negative) than either those identifying themselves with responsibilities for learning in their jobs (48% positive, 38% mixed, 4% negative) or programme staff (41% positive, 32% mixed and 18% negative).

Similar trends are reflected in responses to the questions about the ability to achieve organisational mission. Looking at organisational hierarchy, senior management were more positive (52% positive, 41% mixed, 4% negative) than mid-level staff (43% positive, 35% mixed, 18% negative). Both were more positive than technical advisors, who were ambivalent (36% positive, 25% mixed, 36% negative) while external experts and consultants were the standout sceptics of all groups (16% positive, 26% mixed, 47% negative). All of the groups were more positive about their ability to learn as compared with the ability to achieve the organisational mission, but the similar relative distributions and variations were represented across the groups.

The positive but ambiguous initial figures offer multiple possible interpretations. They hint that **senior management and M&E people tend to perceive the agenda as positive more than programme staff and those with a more analytical function** (learning, external consultants, technical advisers). This offers tentative support to the notion proposed in the framing paper of a more sceptical “squeezed middle” of programme or project officers who must react to the agenda. This interpretation suggests that the responses are shaped by where you sit within the sector.

An alternative interpretation, founded on a critical realist perspective might suggest that the positive responses are caused, in part at least, by respondents internalising the discourse and reshaping their perspectives of their work. Certainly the standout sceptics are those from research organisations, whose negativity about the agenda may be a function of a shared professionally-critical standpoint (or, conversely, to the well-known challenges in measuring the impact of research). At any rate, the findings appear to confirm that the agenda is not totalizing in its effects. Other factors such as respondents’ personalities, values, assumptions and politics may be shaping their interpretation of the results agenda.

4. Qualitative Story Data Analysis

The survey responses suggested the results agenda is perceived as having a mixed effect, broadly positive but with significant down-sides. The stories, unsurprisingly, also depict ambiguous interpretations of the artefacts and the agenda. As this paper aims to trigger questions, I present these ambiguities by juxtaposing conflicting interpretations. Part 4.1. discusses the main artefacts, which act to translate the politics of evidence into day-to-day practice, and how the varying interpretations made of these artefacts depend on relationships between donors and grantees. Part 4.2. elaborates on the broad range of strategies adopted by the story-tellers to the agenda, from internalisation to resistance. Part 4.3. considers the contradictory effects of the processes, setting up a series of oppositions capturing the ambiguity about the agenda in the stories. Finally, 4.4. looks at circumstances or

conditions which may explain what might trigger more positive or negative perceptions, beyond the standpoint of the respondent. I will refer to respondents here as ‘story tellers’ as it is the stories that form the focus of the analysis.

4.1. Artefacts and the Politics of Evidence

The stories describe a variety of processes, policies, protocols and practices, all of which concern gathering, demanding, providing and coordinating information. The following table is adapted from the framing paper, which divides artefacts into those pertaining to results and those to evidence.

Results artefacts	Evidence artefacts
<p>Used very widely within the sector for planning, implementation, monitoring and evaluation. They are reporting, tracking and disbursement mechanisms that include:</p> <ul style="list-style-type: none"> • Base-line data • Results reports • Progress reviews • Performance measurement indicators • Logical framework analysis • Risk register • Theories of Change • Payment by Results 	<p>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:</p> <ul style="list-style-type: none"> • Randomized control trials • Systematic reviews • Cost-effectiveness analysis • Option appraisal • Social return on investment • Business cases • Impact evaluation

Three stories concern the generation of evidence in a primarily research context i.e. where the purpose is not primarily the intervention itself (see Box 2, Big-E Evaluation). Thirteen other stories describe processes linking results to ‘inputs’ to provide an articulation of value for money (the artefacts in the right-hand column). The remainder of the stories concern primarily processes and protocols that require the identification, measurement and reporting of results or outcomes – those identified in the left-hand column.

The context of the artefact and the perspective from which it is viewed are, therefore, as important as the artefact itself in determining the story-tellers’ assessment. Consider the following two snippets:

“We are a small NGO working with donor agencies who regularly require us to produce quantitative results to be measured against a log frame or similar quantitative indicators. Actual quantitative results are regularly different from those originally planned, and our donors require us to explain the differences. In the past we sometimes felt guilty or incompetent during this process, having to delve into a project's operations to find out why things were slower, or less, than originally planned. Eventually however we realized that most of our late or insufficient outputs were due to one simple common reason. In fact they were not slow or insufficient at all. The reason they differed from the plans was because ... the plans were wrong.”

“We were in discussions with an organisation with a view to funding their work but when we pushed them on what their impact would be and, more pertinently, how they might measure it (without an overly quantitative focus) there was, in our view, "a great unravelling". The organisation couldn't clearly articulate what it sought to achieve and how it would know that it had succeeded (and to what degree) and despite much effort on our part to work it through with

them they were simply unable to articulate their work at that level. Ultimately we decided that the organisation didn't have the strategic ability or leadership to deliver on the (relatively straightforward) project and so we decided not to fund them."

Two points are worth noting. First, although both concern the importance of a plan and framework of results, the *interpretation* of the planning artefacts are very different. The stories imply different emphases on the role of quantification and perhaps on the expectations of accuracy. It should be noted that the donor in the story above describes in another snippet a relationship with a different grantee, where the grantees persuaded the donors to support a programme with much less ex ante planning, on the basis of more convincing assurances of the grantee's ability to deliver. This offers a third interpretation to that of the above two stories and suggests that interpretations of an artefact are often dependent on the specificities of the relationship.

In fact, artefacts are always embedded in the relationship between the staff of two or more organisations, and their interpretation will be a function of the negotiation and discussion within that relationship. Several stories described discussion around the terms of references and methodology of evaluations, and several more talked of the precise composition of the reporting formats. Take the following description of a reporting regime:

"Our organization received a generous multi-year core funding grant from the XXX Fund. While there were mostly positive impacts from this funding towards our program, the reporting requirements were rigid and based on the SMART format, which is heavily results-based. This was a very time consuming effort for our organization as there were three results-based documents to submit per year (annual plan, request for payment with highlights of activities implemented and annual report). ... Our main criticism is that the reporting format did not allow for a more analytical and nuanced assessment of the overarching outcomes and impact of our program. We felt frustrated and limited in not being able to adequately present our impact analysis according to our organizational M&E system which relies on the monitoring and evaluation of outcomes, priority groups and progress markers. We believe that that "impacts" are the result of a confluence of actors and circumstances for which no single organization can claim full credit. As such we prefer to talk about contributions, thereby acknowledging the importance of collaboration, the work of other actors as well as expected and unexpected circumstances. The restrictive results-based reporting format did not allow for this type of collaborative impact attribution, was unduly linear, and encompassed unrealistic assumptions about the pace of change in gender relations.

The point here is not to say that an artefact is 'good' or 'bad' (although there are clear statements to that effect) but rather that the artefacts change meaning depending on the interpretation put on them, and on the shape they take through the negotiations of the individuals involved. Interestingly, the above example concluded with a description of extremely positive discussions with partners as a result of their experience generating fresh interpretations through fresh negotiations in another context and a "unique tool" capable of "countering some of the 'magic bullet' approaches". Moreover, the starting point for the interpretations and for what is 'good' or 'bad' may be culturally specific:

"Although I have no formal work to support my claims, my daily experiences have illuminated how the results agenda is deeply incompatible with local Brazilian culture. This general incompatibility is certainly causally intertwined with the lack of this agenda's presence on the national scene. I, as a Global Northerner, also feel the paucity of what I would call an 'evaluation culture.' I am no anthropologist, but I find that Brazilian culture has a distinct way of going about assigning responsibility, discussing goals, and critique, all of which I am constantly learning. My company works with public contracts [...] I must subscribe to their long-term goal of instilling a more results-oriented mindset. The question is whether I can articulate convincingly enough how this message can mutate under the wrong conditions into

the "results agenda" (with its pejorative connotations)."

In short, artefacts, their use and implementation are therefore open to interpretations which are formed in part through negotiation, and partly by the prevailing discourse. As the framing paper notes, they are likely to shift, certain interpretations and artefacts rising and falling in favour. Section 4.3. juxtaposes positive and negative interpretations of a series of artefacts to illustrate this.

The second point is that artefacts are manifestations of power and control. Which interpretation a story-teller favours appears to depend on 'where she sits', and the relative power to interpret artefacts. This will shape the space for negotiation. Many of the respondents are located within the aid chain and must manage, mobilise and communicate information to be accountable, to persuade and to raise further resources.³ The artefacts shape the way these practice work and as such are direct, simple forms of power, in that their implementation and manipulation shapes who gets resources. For example:

"I believe that part of the reason we developed the Theory of Change for our organisation this year was because of the funding environment pressure to be able to explain our approach to development and why we specifically offer good 'value for money'."

The direct use of power by upstream actors curtails the processes of negotiation over methodologies and interpretation of artefacts mentioned above. The withdrawal of funds in the story above is one such example, but the stories presented others from both the donor and the recipient side. Story-tellers write of a broader change in the atmosphere, as they become "increasingly aware of pressures" stemming from the new emphasis on results. As one writes:

"I'd say the results agenda was never entirely absent. But over time a stronger 'language' has been introduced to become more specific."

While such direct manifestations of power are common (i.e. shaping behaviour under threat of withdrawing funding), they are not the only manifestation of power.

The discourses of rigorous proof of impact constrain the forms of knowledge that are considered legitimate and which have shaped many stories whose starting point is the need to prove impacts. The values of the results agenda have permeated the sector and seem to have been widely internalised. Many recipients are responding in ways they think donors expect to be competitive and access funds becoming, in the words of the framing paper, "More Catholic than the Pope". This suggests a broader change in discourse and expectations for knowledge, hinting at a deeper seated power imbalance. It has also offers entry points to people who might not previously have been able to shape policies and resource allocations. One story emerged from an INGO HQ reacting to control its country offices, the country offices having started to generate theories of change; another was prompted by local civil society who demanded their international counterparts met

Box 2. Big-E Evaluation

The vast majority of the stories concerned data generated about a specific intervention, and used, and used in relation to the management and ongoing implementation of that project. This may be attributed to the questions asked and the formulation of the survey, and should not be attributed to the prioritisation of the respondents. However, it means that the discussion focuses exclusively on small-e evidence, rather than Big-E evidence.

Three stories focused on the use of RCTs, and the methodological strengths and weaknesses of these as methods. They raise questions about the claims made and the qualities of the numbers in two specific cases. The paper acknowledges the importance of these issues, but will not address them in depth.

the new requirements of clear articulation of results, and found champions within their counterparts who agreed. In short, the discourse has changed and people have seized onto the discourse to shape resource allocations.

The next section explores the different reactions and strategies to the results agenda, while 4.3 explores the differing interpretations and perceptions of the agenda.

4.2. Strategies, Reactions

The previous section indicated that the results agenda has caused the increased use of certain artefacts pertaining to articulating and reporting on results, on providing theories of change and on showing value for money. These have been interpreted differently, and the interpretations are shaped by negotiation and by power. This section discusses the strategies that have been used in response to the artefacts.

Some stories suggest that the agenda has been internalised. Such stories take as their starting point the key tenets of the results agenda – the need to integrate rigour in measuring attributable change. All of the stories evince some form of interest in learning about the effectiveness of their work and whether that contributes to impact. Moreover, the discourse has its own power – recipients are responding in ways they think donors expect to be competitive and access funds, which may not always be consistent with the donor expectations. Some of the stories suggest that champions have seized upon the agenda and used it as an instrument to drive internal change. For example:

“[...] most of the top management did not want to hear and embrace outcomes and impacts. In fact they believed that monitoring and evaluation was for economists and that it w[as] not their responsibility. [...] I started making presentations, whenever I would get a chance, on building a results-based research system. ... I thank God that even the "hardcore" scientists began to talk monitoring and evaluation and in fact focusing heavily on outcomes and impacts.”

Others have been able to use the results agenda to take forward learning and accountability agendas within their organisations. Regardless of the motivations, many of the stories indicate internal changes suggesting some form of compliance.

Several stories mention needing more resources to meet administrative requirements, whether via new staff (larger organisations), support from outside consultants, capacity support, training or simply the expectation that the same people will do additional work. Other stories talk of a change in internal processes, in an attempt to take the requirements seriously – of “new MEL guidelines and processes introduced that are more quantitative” as a positive change; of “improving ongoing monitoring systems [and] also putting a significant amount of resources in to 'proving' our impact”. They involve changing internal proposal and reporting pro-formas which can entail considerable effort and cost. One story involved a major overhaul of internal systems that “introduced new layers of budgeting and reporting requirements (financial codes/linkages to our MIS)”. This sought to link budgeting to their programme objectives, in accordance with donor reporting requirements, and their own grants to partners were linked to the same objectives, requiring their partners to change their reporting as well. Unfortunately, since activities and costs contributed to different objectives, the linking process was arbitrary and the information unhelpful. The tone of the story, in this last example, was deeply frustrated as the systems made little sense given the lack of attributable change to one budget code or another).

Many organisations have incorporated the agenda into their systems, before using similar processes themselves to coordinate across their own country-offices, partners, or grantees. The following experiences indicate the ambiguity felt by implementing these processes:

We have found this overall quite a healthy demand for the organisation. [...] It has not been without challenges. of course. All our offices have struggled to adapt planning and resourcing in the short term to increased demands, and we have been over-stretched at times on M&E and reporting. But over time we are budgeting better for that, and see a value in improved standards.

"[A] Philanthropic donor makes policy decision to become result-based in its giving - hires a senior evaluation manager. Existing staff less enthused about process of changing from reporting their succes [...] Uneven change process across the donor organisation leads to variance in expectations and quality of project implementation. Attempts of programmatic approaches by the donor, in thematically clustering disparately located and motivated recipients, presents challenges." [A precursor to a positive story of change.]

In short, the implementation is seen to have had good and bad effects. The paper argues elsewhere that what interpretation you provide depends on the detail of the artefact, on the fit and freedom to interpret it, and on what your position is and how it fits within your organisation.

Other stories indicated an uncomplicated strategy, simply involving a shift to the new requirements, without much effort and, perhaps, without requiring much of a change on the existing systems or people's ways of working. Of course, one person's effortless shift to a new process is another's indifference or box-ticking:

"[I] find the team with which I work almost without reluctance adopts new practices and gets on with the work it cares about."

"There is no desire to engage with the results agenda to influence organizational learning. Logic models have been developed recently for programmes but again this is only due to a donor requesting these and they do not feed into programme or organizational learning. The results agenda is only seen as another opportunity to access funding without any appreciation of the wider issues and benefits."

"And in the process of ticking all these boxes, some of them are losing sight of the quality and/or integrity of the process and not reflecting on their own real interests in evaluations and their findings. Once the contracts are signed, I have sometimes had the impression of a huge sigh of relief, especially if it is an M&E department responsible rather than programmes personnel, and of moving quickly on to the next task."

Each of these stories capture experiences that in form resemble compliance, but which in fact are empty performances that take up time and generate data that have little utility or meaning but further reinforce the results agenda. This supports the framing paper's observation that the agenda can trigger a costly 'performance to script' (going through the protocols demanded), without value to those collecting the information.

Other, more active, forms of resistance aim to *circumvent* the processes:

"staff have found ways to 'get around' the new consequences for poorly rated programs (eg framing a program re-structure as a 'new' program, thus giving the impression that the underperforming program has ended.)"

"The way we have gone about measuring against very very specific indicators in order to be able to prove and/or aggregate results/tell a story about our effectiveness has been challenging."

We have had to be so specific and so focused on a very limited number of indicators (and often only look at one of these indicators in any given project) on an even more limited number of projects that, eventually, what we learn about our impact is often irrelevant or insignificant in terms of what a programme or country team is, overall, trying to achieve. This means that country teams can and do, sometimes, choose to completely ignore the rigorous reviews that we need to get done and the conclusions that are drawn because they can fairly easily be written off as irrelevant or insignificant. Which perpetuates the feeling that they are purely extractive and can not/do not add value at a project level.

These stories indicate that some actors take on more active forms of resistance and circumvention of results artefacts. This is also not to say that the strategies of resistance are only negative. The stories indicate that some forms of resistance can have positive side-effects, leading to conversations with partners or with 'upstream' M&E staff to improve accountability processes, shifting existing views while at the same time rejecting the interpretation implied by the donor. Resistance in some instances produces useful learning for those upstream while some confident recipients – or, viewed another way, some flexible donors – have been able to resist restrictive results artefacts and to negotiate with their donors to use existing organisational systems. This depends on the relationship with the donor. The stories indicate that some claim to be more open to this than is assumed.

These stories show a variation in experiences that are strikingly similar to the experiences of reactions of publicly funded professions to earlier New Public Management reforms, which have included:

- Subjecting themselves to and internalising performance monitoring and evaluation processes, making considerable changes to the way people view their working lives and practices;
- Dissociating themselves from the reforms, either through hollow compliance without changing practice, or creating 'absorbing groups' to buffer the profession from the audit processes ;
- Using the changes as an instrument to colonise management roles or improve their professions' status – such as that of nurses – within the hierarchy.⁴

The expectation from the stories is that these strategies will continue to arise within the development sector as the results agenda approaches its adolescence. The interpretation adopted of an artefact and the strategy adopted in responses appears to depend – to some degree – on where you sit within an organisation, but also on personality, on perspective, on values perhaps on culture. It is to the different interpretations of particular artefacts that part 4.3. is dedicated.

4.3. The effects: opposing perceptions

The different interpretations of artefacts and the strategies adopted in response are illustrated by stories indicating contradictory and often polarised effects from the results agenda. The following analysis draws on the stories to present the different perceptions around different pivot points, with opposing or contrary views see-sawing around a key artefact of the results agenda.

“Accountability trumps learning”. This tension lies at the core of many of the stories. It hinges on two contrasting propositions –information is being provided to help allocation of resources; versus information is being used to let people hone and develop their ongoing work. Is data passed upstream for managers to make an allocation decision, or is power being devolved to the implementers? In shortened form: is it information up, or power down? This is a familiar challenge to the aid sector.⁵

Some stories insist that formal monitoring chokes organic processes of learning. As one snippet says:

“I have also learned that 'accountability trumps learning' and in trying to design an M&E

system that can do both you will kill off learning unless really careful to make space for it."

Another story concerns a research programme which is designed specifically to help a group of partner civil society organisations learn, while at the same time making these processes the object of study through monitoring and evaluation. The researchers in the programme "suggested that the best way to pursue results might be to work responsively with the outcomes of one's own previous efforts" – in other words, to meet the results agenda by reflecting on prior experiences and learning. As the programme continued, however, the story-teller saw a tension in the provision of support and making it the object of study and in the course of a meeting admitted to "*concern that we only have this beautiful sharing of experience and continual transparency of evolving experience because we have not objectified any of it*".

"So," says our researcher, "you don't think it's possible, after all, for a learning-oriented M&E system to meet accountability requirements in a results-driven era?" I am stumped."

The expression and balance between learning and accountability for results depends on the upstream requirements and the space they leave for learning. In practice where pressures on costs constrain the resources available for monitoring and evaluation and knowledge management, this space is likely to be choked, unless great efforts are taken to preserve it.

However, some of the stories suggested that in fact the dichotomy is false. For example, two experiences described the effects of the results agenda on their organisation as follows:

"...significantly more attention given to M&E, more consideration of what people should be measuring, and how. More consideration on what the programme actually wanted to achieve - increased clarity on what was realistic, strengthening the logic of the programme."

"...it is too early to say if it will work but I think that encouraging folk to think strategically, measure and monitor, report to a single template, that these are good things. We want to provide evidence, but that can be both qualitative as well as quantitative."

It is striking that many of the positive stories concern other peoples' learning within the same organisation, rather than the story-tellers' learning themselves. The question in these cases becomes who is learning, what are they learning, and how is that learning shaping the programme quality?

Reflect on cause vs. Log-frames on steroids. The development and refinement of a "theory of change" – an artefact that cropped up again and again – was core to many of the stories, both for reporting and for learning purposes. The theory of change concerns the articulation of a causal chain, which may be justified and supported to different degrees of robustness, and which may guide the design of a project and its measurement. It supports both learning and reporting, and many of the stories were positive about its utility:

"the process of writing the theory of change has been very positive and has inspired our Country Programmes to think about using 'theory of change' methodologies for specific projects and country strategies. Developing a Theory of Change has been one part of our ongoing journey to better understand what we do and why we do it..."

The results agenda has encouraged agencies to articulate in a more disciplined fashion the links between their work and the change they want to bring about, and thus supports them to learn. However, as one story-teller observed, what is important is not the artefact itself – be it an impact pathway or a theory of change – but rather how the artefact is used and to what purpose. He went on to say that the results agenda:

“has created space and motivation for greater use of theory of change which if done properly could lead to a more responsive [...] research agenda, with greater emphasis on reflection and learning. If ToC is used 'log-frames on steroids' then the effect could be opposite. We are right in the middle of this debate and it could go either way.

(This is the same story-teller as commented on learning trumping accountability). As the framing paper notes, the discourse around artefacts and the interpretation shifts. How a given tool is interpreted and implemented therefore will determine its use. Most artefacts have the potential for different uses, and must be moulded to fit a purpose. Similarly, evaluations and evaluation systems tend to have a vector, a particular function or use which the evaluator has in mind, which will shape how they are used. As with learning, the question is whether it is possible to have dual purpose systems. In what will become a familiar refrain, much depends on the upstream actors and their willingness to allow space and flexibility and adaptation to context. Here is another story-teller reflecting on experiences of two different donors:

“processes such as creating theories of change, for example, was more encouraged by [donor X] and there was flexibility in how to measure outcomes and impact in discussion with programme managers. Other donors could be more faceless and systems and procedures would determine that projects submitted for funding would need to be more focused on service delivery with clear indicators of output/outcome for monitoring and evaluation systems.”

Experiences with the latter were less positive and less meaningful in the data collected than with Donor X. A couple of stories from both donor and grantee perspective showed how responsive and considerate donors allowed a well-articulated and clearly justified theory of change convincing donors to depart from their pre-ordained structures and expectations and instead shape themselves to the grantee's strategy.

Useful Data vs. Reductionism. The discussions of theories of change and learning boiled down to 'contrasting interpretations of artefacts' which concerned the ability of an evaluation system to serve two functions. The third set of contrasting interpretations considers the value of numbers for resource management and accountability purposes. It is worth starting with the positives from the stories: some stories indicated that the pressure to articulate results had encouraged the generation of performance frameworks that capture the important information. Most of the stories of this kind focused on the discipline offered by a focus on value for money:

To respond to the value for money agenda, we tried to understand our costing by calculating the cost per beneficiaries in some countries where we work. This involved first developing a methodology to analyse the cost and then applying this in a consistent way across the interventions. As the results of this exercise we were able to estimate the cost per beneficiary, to identify local factors that affect the cost and the need to have reliable monitoring and reporting system in place to generate credible data on our achievements.

“... this requirement has helped bring greater discipline to our focus on VFM as Programme Managers and then we have supported our CSO Partners to make sure they are focussing on and controlling their cost drivers and how they provide value.”

Others suggest that the results agenda had prompted them to draw on a wider set of evidence in justifying their interventions:

[We] recently reviewed our country and regional strategies and tried to develop an evidence based model for determining working priorities. The model (Country Focus Framework) uses a combination of the Inequality HDI, Multidimensional Poverty Index down to district level if available, and in-country information provided by partners. It is hoped that this will improve

long term monitoring of impact, and also provide evidence-based reasons for choosing various priority countries and regions within countries

These stories indicate the possibility of generating useful information, introducing discipline into allocation decisions and into programme design. It's noticeable, however, that the positive stories appeared to emerge exclusively from those with responsibilities for allocating funds or coordinating measurement amongst grantees or partners.

The contrary view was also prevalent: that reporting 'upstream' on specific numbers was reductionist, took up a lot of time and resources producing meaningless numbers that are "*of little value to program improvement, and a not very credible approach to meeting upwards accountability requirements*". Indeed, reductionism was the single biggest problem highlighted in the course of the stories. Reductionism here suggests the dumbing down of reporting to meet standards, such that management received meaningless and simplified information:

"Lots of time spent across the organisation on an annual basis calculating and aggregating achievements against a whole of agency set of 'headline results' - of little value to program improvement, and a not very credible approach to meeting upwards accountability requirements..."

Several of the stories indicated that the numbers that they were generating were of very little use to the running of the programme, particularly for more complex, transformational interventions.

In fact, as the framing paper suggests, many appear to be driven by the need to report further upstream on aggregated numbers showing impact. These aggregates are produced to alleviate concerns from the domestic publics of bilateral donors, rather than the operational context, as the following two stories suggest:

The hope is that key indicators that are measured in every [bilateral donor-funded] project will help [the bilateral donor in question] to report on its world-wide achievements [...] in a "sexy way". Also, such reports are intended to be used for acquisition purposes. However, some serious dangers risk to be neglected in this discussion: [the] projects work with tailor-made approaches. Standard indicators fail to embrace the diversity of the projects. Also, by using indicators that are easily measurable, attention is given to activities that will contribute to "achieving the indicator" instead of activities that really matter but that are more difficult to measure.

The Results Framework [...] has partly become a focus on quan[t]itative numbers, i.e. how many direct and indirect beneficiaries per each funding agreement for a project or program. So that they can say that xxx numbers of people's lives were saved in the last year due to their funding, etc. Unfortunately, they themselves do not know what they mean by indirect beneficiaries and how we are meant to count these. The numbers are very rubbery and they don[']t mid that we seem to be only making a guesstimate of the nu[m]bers.

The last story goes on to explain that this means country teams resort to ignoring or subverting the systems. Such stories confirm the power dynamics at play: that results are at the centre of an accountability system, whereby information in the form of results are packaged up neatly and sent 'upstream'.

Again, the positioning of the story-teller seems to influence strongly the assessment of the artefact. Those managing programmes or coordinating the collection of data across several partners seem to be more likely to express positive experiences of the results agenda, while those at the receiving end are much more likely to react negatively.

"in most cases, 'value' is not clearly defined, leading to a tendency for this to collapse to thinking about income results/policy impacts - ie more tangible outcomes, at the expense of consideration of power dynamics, role of civil society, etc. We had an experience recently where we produced a mid-term review [...] This was basically a strategic analysis of how they might organise and strategise more effectively [...] But the entire questioning at the Senior Management Team was along the lines of, what in this report helps us make the decision whether to allocate funds to the campaign or not? (ie how to choose between funding the campaign and funding for programme work). I can understand the desire to answer these questions, but the same evaluation can't address both.

Building on the last snippet, several stories revealed particular scepticism about the quality of information and of numbers generated to indicate value for money. As with other aspects of the agenda, perceptions are mixed. See for example the following comments, discussing an experience of an INGO managing several partners:

"Positive: this requirement has helped bring greater discipline to our focus on VFM as Programme Managers and then we have supported our CSO Partners to make sure they are focussing on and controlling their cost drivers and how they provide value. [...] Negative: we are concerned that the vfm metrics we produce will come to dominate the decision about the value of a particular grant. We have been constantly making the case these metrics cannot be used in isolation but as a part of a broader set of pieces of data (including qualitative) on which a reasoned judgement should be based. To be fair our donor counterparts have been receptive to this argument but we are not sure whether the debates are going in a direction which will allow space for reasoned judgement"

The concerns articulated note that VfM is particularly problematic when viewed mechanistically by managers or through a narrow economist perspective operating within a reductionist framework.

In short, the costs and frustrations with generating meaningless numbers targeted towards accountability and aggregated, inappropriate targets was the single greatest criticism of the results agenda. The results agenda is experienced by many of the respondents as unhelpful, particularly for transformational programmes when pursued through narrow choices of meaningless rigid quantitative indicators that fail to encourage learning by programme staff. They have opportunity costs for more productive knowledge management and were described as too often inhibiting the work and imposing significant costs on those generating the numbers. Like laws and sausages, such numbers may inspire rather too much confidence in their manufacture than is justified: the consumer would be advised not to consider how they were made.

Realism vs. Risk aversion. A further tension lies in the implications of the results agenda on planning. Consider the following two quotes:

"We deliver technical assistance projects for government reform [...] We are increasingly under pressure to "link payments to outputs" [...] - this means we have to think very carefully about the level of risk surrounding delivery of the outputs we chose to define in proposals and revised M&E frameworks/ logframes, and what proportion of payments we can absorb linked to those outputs. [...] this has forced us to analyse what is definitely achievable and what is not so certain, and to develop more robust strategies for achieving the former. Overall I think this has created more realistic proposals, better political and risk analysis within bids and projects, and a better understanding of what impact means on governance programmes. On the negative side, we are perhaps less ambitious than in the past [...]"

"This has built on and confirmed my deeply held suspicion that we have a tendency to rather 'overclaim' the case for our accountability to partners and communities. With the need to be increasingly rigorous in our measurement and circumspect in what we claim our results mean

we have had to accept that we can't measure 'accountability'."

In other words, the results agenda has forced a discipline in the planning and articulation of development programming and interventions, resulting in more credible and better-thought-through projects. The implication is that the sector previously lacked this discipline (see section 4.4 for a discussion of how different organisations have reacted to the agenda).

However, other stories suggested that the results had created an aversion to risk and a tendency to shy away from more adventurous changes which responded to the deeper causes of poverty:

"... by using indicators that are easily measurable, attention is given to activities that will contribute to "achieving the indicator" instead of activities that really matter but that are more difficult to measure. Thus, the introduction of standard indicators may set wrong organizational incentives that will undermine a meaningful development cooperation."

"A high level manager in an organisation funding organisations in the "Global South" told me they would only support projects that would produce quick tangible outcomes in the future. "

"We were developing our operational plan for the division (a regional division). Issues were very dumbed down and certain issues not selected because they were considered to be too hard to measure, and not 'speaking to' or contributing to 'corporate results reporting requirements'. I see this as a negative impact of the results agenda."

The results agenda, with its focus on reporting on discernible and pre-determined results, has in some cases reduced the space for more political, more complex and more risky interventions. Much depends on the funding relationship and the planning artefacts in use, and there is a question as to whether these negative examples are genuinely the 'fault' of the donors, or rather a shift in the surrounding discourse.

This question is more than a simple tension: it captures in a microcosm an ongoing debate about what development should be, between on the one hand transformative change which seeks deep-seated but uncertain and risky political change; and transactional change, which looks for outcomes linked to the inputs through shorter impact chains, which are assumed to be easier to control. Together with the wider issues of reductionism, it can incentivize simpler outcomes that are easier to measure, while discouraging donors and grantees from pursuing the risky or the transformative. The insistence on thinking things through and providing credible predictions of future outcomes, while beneficial in its need for discipline and proper discussion, can also throttle interventions that are less easy to control and measure.

Undermine relationships vs. Support and Organise Partners. The final tension to be raised in this section concerns the effect on external relationships of the results agenda. Again, the stories raised opposing views:

By taking control of the reporting process for [donor programme arrangement], and introducing a shared systematic process for identifying and describing outputs and results, we have introduced much greater clarity and consistency into our engagement with partners funded [through the programme arrangement].

The increasing monitoring and evaluation demands by donors are very much seen as a burden and framed in terms of upwards accountability and reporting, rather than learning, and partners have struggled with the lack of knowledge, capacity and resources to engage with them. For institutionally funded projects, log frames, although Programme Managers have found them a useful planning and management tool, they do not support or encourage the

establishment of practical M&E systems that encourage reflective thinking.

Both quotes come from INGOs staff members acting as intermediaries, channelling funds towards local partners and country offices. For one, the results agenda allowed the articulation and coordination of a range of partners, helping clarity of purpose and consistency. For the other, they are 'increased demands', and are seen as a 'burden'.

Concluding remarks. The discussion above shows different perceptions of similar artefacts. What explains the differences in the perspectives? Is it simply the location of the respondent in the aid chain - whether they are managers hungry for data or long-suffering project officers collecting the data? Or is it more concerned with the space for interpretation of the artefact, and the flexibility and adaptability offered? Certainly, many of the negative experiences appear to arise from a question of suitability – forcing a particular interpretation of an artefact as though squeezing “into a pair of trousers that doesn't seem to fit” (as one story put it). It is to these that I turn now.

4.4. Conditions/circumstances

The previous sections reviewed how the strategies and interpretations of the results agenda, this final section picks out three elements from the stories that seem to determine or shape the outcomes of the accountability processes, whether positive or negative. They are themes tentatively identified from the stories, proposed with a view for further discussion within the conference.

Where is the organisation starting from? A factor affecting perceptions of the results agenda concerns the starting point of the organisation: what level of capacity exists, what are the existing systems and processes in place? Many of the positive stories highlighted above seem to have emerged from internal champions who used the external pressure to leverage reform within an otherwise sluggish bureaucracy. For organisations doing little meaningful evaluation, resources and processes for monitoring will of course help improve learning. The agenda has enabled these individuals to get more resources for monitoring and learning, to shift internal processes, sometimes quite substantially. For them, accountability does not exclude learning.

However, the agenda appears to bite hardest on organisations who already take learning seriously, and who have existing capacity and existing systems. They are unlikely to gain the 'quick wins' generated by harnessing a push towards results, since they already have systems and commitments to learning, but at the same time the more damaging aspects of the agenda have a strong effect (reductionism, risk-aversion, standard-form bureaucracy). The following comments are examples:

“A very busy advocacy group in the "Global South" that I evaluated had been subjected to several rounds of mutually contradictory training courses on "results-based management", which took several full days of all senior managers' time. Attempting to classify their activities and outcomes according to a complex input - output - outcome - impact logic, the members of the advocacy group started to produce illegible progress reports that made them look quite unconvincing to the same donors who insisted on the training. The system was incompatible with their own, quite sophisticated internal monitoring and learning system.”

“We felt frustrated and limited in not being able to adequately present our impact analysis according to our organizational M&E system which relies on the monitoring and evaluation of outcomes, priority groups and progress markers.”

At the other end of the spectrum, are smaller organisations, often supported by intermediaries like international NGOs. The following responses are typical, often by exasperated INGO coordinators:

“Some partners do not respond - they have little or no resources or their IT infrastructure is

not reliable enough. Do we restrict/deny funding as an incentive to improve these areas or focus more on results? - Some do; the level of detail varies and whilst we do not want to be too prescriptive (results and impacts vary from programme to programme) it means that it's difficult to analyse these results consistently or efficiently considering our limited internal resources. - It is an added administrative burden for the partners themselves; is this the best or only way for us to monitor results?"

"We are a multi-donor funded organisation and have an implementation approach of working (funding etc) with local organisations to do the implementation. We are also trying to develop a results focus and setting up the M&E systems, processes and practices (our organisation is relatively new) to support this. However, working with many and varied partner organisations (public sector, civil society, other non-government organisations) who on average have weak monitoring and evaluation practices presents constraints to what it is possible to deliver in terms of being results focused. We have therefore developed a three tier approach - level 1 is a clear ToC and monitoring and evaluation plan through the logic from outputs to outcomes (short to longer term); level 3 is a clear ToC and monitoring plan to output level only; and level 2 is in between - a clear ToC and monitoring plan to short term outcomes."

"[We work] entirely through local [...] organisations in 14 countries, who have varying degrees of capacity, but all of which have considerable resource constraints. [...] partners have struggled with the lack of knowledge, capacity and resources to engage with them. For institutionally funded projects, log frames, although Programme Managers have found them a useful planning and management tool, they do not support or encourage the establishment of practical M&E systems that encourage reflective thinking. Both partners and Programme Managers struggle to take large indicators and break them down into concrete measures of progress that can practically use to support their own project management."

In several stories, INGO staff presented their local partners as having low administrative capacity. These representations are convenient, and often appear to indicate a familiar disjuncture in perspectives on what counts as valid knowledge. The 'failures' are often in the inability of the local partner to translate or articulate achievements in written form, rather than in a lack of capacity or the ability to implement programmes. Demanding a blanket standard or language of reporting often means cutting off funding to these organisations who cannot make the translations – or (as is the case) the intermediaries ramping up significantly the 'brokering' support they provide at the expense of coordination resources.⁶

The stories help to distinguish between actors at different points in the trajectories: many of the positive stories highlighted above emerged from a middle range of 'sluggish bureaucracies' who hitherto had not focused seriously on learning and for whom the results agenda has been an important wake-up call. The agenda has enabled them to get more resources for monitoring and learning, to shift internal processes, sometimes quite substantially. One other set of organisations arguably comprises the learning-sophisticated organisations who already have systems and capacity in place for learning. For them, the results agenda has been an additional burden, reductionist and damaging. A third set is the smaller, often national level organisation described as having limited administrative capacity – although not low capacity overall – who struggle with making the translations required between their expertise and knowledge and the formal requirements of the results agenda, and who risk having their funding reduced through failing to make these translations.

For the last of these, a debate is core to wider perspectives on development: do you support organisations with low administrative capacity in the belief that such organisations are vital to a healthy development process and further that their lack of administrative capacity does not necessarily indicate poor work; or do you insist on channelling funds only to those who can show that they are

competitive and effective, in the belief that you are duty bound to have as much short term impact as possible? This problem is well-established in aid debates: however, the results agenda has intensified the issue through the increased stringency and insistence of the requirements of translation.

How has internal reform been implemented and communicated? The detail of *how* internal reforms have been implemented is the second significant element in the stories. One key issue lies in who designs the programmes, and how, vitally the programme staff and the M&E staff are interlinked.

“The M&E officer was responsible for developing the M&E system we would use based on our theory of change. We, the programmes team, had to be proactive to be involved but it was clearly seen as 'his responsibility'. Our partners and country offices had no input. When the document arrived, I could hear the sighs around the desk as my team read it. It was hard to read for us, native English speakers, and when it was clear that, while it could be described as 'good' in M&E terms, it would not be possible to implement. It would take too long, was difficult to understand and, while it was not just about numbers, focused on collecting 'evidence' rather than listening to people's experience and learning from it. There was some good stuff in there - for the first time we would be able to have a more holistic view of what we were doing and what was happening as a result of it.”

Inevitably, organisations have internal divisions. Previous sections indicate that there are a range of strategies available to actors, from adopting to subverting to resisting. The strategies adopted within an organisation, and the effects of the results agenda, will depend on who, internally, is pushing for the reform, what are the expectations, and how the expectations are managed of the users of the data.

“We had an experience recently where we produced a mid-term review of a global campaign for an INGO. This was basically a strategic analysis of how they might organise and strategise more effectively (eg around national-global linkages, working in partnership etc) which, according to feedback, was useful to the campaign managers. But the entire questioning at the Senior Management Team was along the lines of, what in this report helps us make the decision whether to allocate funds to the campaign or not?”

A common focus on the use of the evaluation mechanisms is therefore vital. Unfortunately, resources for measurement are limited, and collecting some data may mean you cannot collect others. As noted above, an insistence on collecting data for accountability purposes will squeeze the space for learning. To some degree evaluations and knowledge management is a zero-sum game:

“We have a big picture as an overarching framework that outlining the process of evidence inputs, the evidence validation, processing and packaging - tailoring to key internal and external stakeholders, and the use of the evidence for programming improvement/changing practices, information for policy advocacy, resources mobilization as well as the broader accountability and governance goals. The implementation of the framework and approaches will be challenging, as it requires a decision to cut some reporting regimes, to exercise more analytical and synthesis skills than describing, and the needs to strengthen solid data collections. But it has also potential to simplify process and to increase professional satisfaction”

Changing internal knowledge management, this suggests, is political and requires bringing the users of previous “reporting regimes” on board, and clarity of needs between different departments and between the coordinating body and the grantees. Fit of the artefact and its interpretation to the organisation is clearly vital. I will leave this with some snippets emphasising the importance of communication:

“I have learned that consistent, clear and simple messaging from leadership is key; that

institutional change processes will take years to effect, and will require dedicated leaders to promote, facilitate and be responsive to the change process experiences.“

“A donor came to visit one of our projects, and, while very impressed with the outcomes we were demonstrating, told us that in the next 6 months they were going to start investigating our 'value for money' as well, even though the project had not been designed to do so. The previous week I had recieved an email from a different donor with a summary of their new focus on value for money, which was a different approach and understanding of the concept. “

As a result of the contractual requirement in our PPA to publish information on DFID-funded projects, we were able to use this requirement as a catalyst to focus our SMT's minds on the value and practical implications of transparency, and secured a commitment and practical plan of action to publish information on all of our organisations' projects, not just DFID-funded ones. We were able to plan to do this without causing too much angst among staff as the deadline for doing it was quite generous.

Does the measurement fit the aim? A third factor with some explanatory power is a familiar one: the complexity of the intervention and the objects. Predictably, none of the story-tellers described their intervention as being simple – complexity arose consistently as an issue with which accountability regimes must grapple. However, despite the challenges complexity presents to measurement of development processes (as outlined in the framing paper), the stories suggest that the interpretation of the artefacts has not kept up.

Three key factors contribute to the frustration. The first was unrealistic demands for attributable change or proven value for money in complex programming contexts. In one story, an organisation brought in well-respected consultants to review their work on Value For Money:

“[...b]ut in the end, they concluded it wouldn't be worthwhile them even doing the work, because the nature of the work we do - supporting African governments with complex, messy reform processes - was not amenable to straightforward VFM analysis and any paper they produced would be too vague to be useful. This was disappointing, because it reinforced the sense among some in the organisation that the Results Agenda is like saying you have to pass an exam and then not letting you sit it.

Several further stories indicated that the demand was not only unrealistic, but also wasteful of resources in complex programming contexts:

“We are also trying to measure behaviour change which is notoriously difficult. We are small with limited staff and resources. I think we have ended up spending too much time and effort on M&E when we could be putting that effort into improving our model rather than trying to demonstrate something that's almost impossible to prove with any certainty.”

“Our main criticism is that the reporting format did not allow for a more analytical and nuanced assessment of the overarching outcomes and impact of our program. [...] We believe that that “impacts” are the result of a confluence of actors and circumstances for which no single organization can claim full credit.”

Demands to prove value for money triggered several complaints of wasted time: “[l]ots of time spent across the organisation ... of very little value to program improvement”; “we and our partners have spent increasing amounts of time justifying our work [...] which has little perceivable benefit so far;” and the agenda “[...] has diverted staff and partner time away from other activities, and it's not clear what the added benefit of completing a 'business plan' in addition to an application form should be.”

The second source of frustration was the requirement to harmonise reporting to pre-defined targets in contexts of complex programming:

“As part of a new institutional contract, we have been asked to divide our programme budget by the objectives included in our programme proposal, and by direct/indirect costs. However the division is quite arbitrary, as all our objectives relate to increasing civil society space and demand and respect for human rights, and are interlinked.”

“RBM is implemented in such a way that our strategies make grand leaps in logic (the program impact will increase peace and security [...] even though the overall program is only a few million). This in turn, forces project log frames to do the same. That means a small project like refurbishing police stations (which is needed) looks like it lead to increased security for the whole country. This frustrates implementers, because they don't know how to fill out the log frame and feel we're making them jump through evaluation hoops without a real learning purpose.

As mentioned above, the need to report to pre-defined results is often driven by the need to indicate performance to domestic audiences. For those implementing the artefacts, the requirements are unhelpful and costly.

The third is the Procrustean reporting requirements and methodological straight-jackets, such as log-frames, which do not allow for nuance in the description of causal chains (after the Greek Procrustes who, upon inviting them to stay the night, either stretched them or cut them down to fit a single-size iron bed)⁷:

Methodological restrictions have been set up from the beginning whilst evaluation standards should be the starting point for methodological choices. This is also reflected in the evaluation of reports that we need to submit: a standard format for all [case studies] is to be used and in terms of describing the methodology we are requested to tick a box 'indicate what methods you have used for the evaluation' and the possible options range from RCT, difference in difference, to focus group discussions, interviews and force field analysis. Different methods for different purposes all lumped together. And nothing about an overall approach to impact assessment of complex issues (like capacity of an organisation or civil society). Wonder how they will score us on this[?] Synthesis will be done by economists with no understanding of assessing complex and dynamic issues like organisational capacity. We had to use standard indicators for all organisations, but how will the analysis by the synthesis team be done? Giving a mean score per indicator?

The restrictive results-based reporting format did not allow for [...] collaborative impact attribution, was unduly linear, and encompassed unrealistic assumptions about the pace of change in gender relations

From this, the space for the development of systems allowing people to learn and develop better interventions is therefore based on three factors: first, the degree to which the donor or other upstream actor insists upon quixotic requirements for attribution; and second, the degree to which programmes are constrained by the need to report against aggregated results or can select results that fit their programming; and third, how far the methodologies fit their programme..

5. Findings for Discussion

Respondents to the quantitative questions indicated not only that the results agenda has had a considerable impact, but that these impacts are more positive than negative overall. However, it also

shows that there are a considerable number of mixed and negative experiences. The challenge of the survey is to unpack the different interpretations of these relatively flat figures.

It does so with reference in particular to the stories presented, understood in the light of the conceptual framework offered by the framing paper for the conference. The stories describe a sector experiencing the growth of three artefacts in particular: the use of theories of change to plan projects and to identify results; the reporting on quantified results; and the linking of outcomes or results to the inputs and costs (value for money). These artefacts generate ‘small-e’ evidence in the context of a particular project are used for resource allocation and decision-making as well as to show ‘performance of the will to govern’ to donors’ domestic stakeholders, as indicated in the framing paper.

These artefacts are perceived, and have been interpreted, very differently in the stories. As the framing paper notes, the agenda is powerful and has shaped the discourse, but it is not totalizing. There remains room for manoeuvre, and the stories show – as with prior experiences in other sectors under NPM reforms – different reactions. Respondents’ strategies span from outright resistance to keen adoption, with the agenda being wielded as a powerful lever by some to make internal changes.

This has led to quite different interpretations. For some, the new artefacts have strengthened learning, for others the space for learning has been choked off; some have found the data generated useful in making informed decisions, while others gripe that it is reductionist, meaningless and costly. For some, it has improved coordination, for others it has undermined relationships with others; it has either helped a disciplined articulation of programmes, or tied interventions to inappropriate models and risk-averse programming.

What can we make of these conflicting experiences? One interpretation is that people’s views are shaped in part on their *positioning*. For example, the survey suggests that management and M&E are more positive than programme staff – is this because they benefit from the data and resource allocation to M&E functions? Programme staff, in contrast, are being held to account more robustly and not always in useful ways. This offers some qualified support to the notion of a ‘squeezed middle’ raised in the framing paper.

However, three other factors with some explanatory power seem to emerge from the stories:

1. *The capacity of the implementer.* The stories suggest three broad classes: those with resources who have been sluggish in monitoring and evaluation; those who have existing sophisticated systems who take learning seriously; and those smaller organisation with limited capacity. The first group in particular seem to have benefitted from the agenda, the last two less so.
2. *Fit to organisation and implementation.* The ability of an organisation’s management and its donors to steer reforms through in a clear, well-communicated and adequately resourced fashion in a manner that fits the mission and values of the organisation.
3. *Fit to programming.* The flexibility of the donor, and willingness to accept as a starting point the intervention and its circumstances: in particular, story-tellers looked poorly on demands for arbitrary results based on domestic politics, a rigid insistence on attribution and methodological narrow-mindedness.

These factors offer starting points for discussion and questions that may be useful to take forward into the conference. Do these possible explanations for different interpretations resonate? Under which circumstances are the artefacts accountability mechanisms used to control and constrain, and when are

they useful guides for action and reflection? What other interpretations or circumstances should be considered? Is it fair to say that peoples' interpretations differ depending on their position in the sector, or are there more important factors?

Endnotes and References

¹ Eyben, R. (2013). *Uncovering The Politics of 'Evidence' and 'Results': A Framing Paper for Development Practitioners*.

² Pollitt, C., & Bouckaert, G. (2011). *Public Management Reform: A Comparative Analysis - New Public Management, Governance, and the Neo-Weberian State*. Oxford: Oxford University Press;

Newman, K., Capillo, A., Famurewa, A., Nath, C., & Siyanbola, W. (2012). What is the evidence on evidence-informed policy making? Lessons from the International Conference on Evidence-Informed Policy Making. Oxford: INASP.

Nutley, S. M., Walter, I., & Davies, H. T. O. (2007). *Using Evidence: How research can inform public services*. Bristol: The Policy Press, University of Bristol; Hood, C., & Peters, G. (2004). The Middle Aging of New Public Management: Into the Age of Paradox? *Journal of Public Administration Research and Theory*, 14(3), 267-282.; Lapsley, I. (2009). New Public Management: The Cruellest Invention of the Human Spirit. *ABACUS*, 45(1), 1-21.

³ Wallace, T., Bornstein, L., & Chapman, J. (2007). *The aid chain: Coercion and commitment in development NGOs*. Practical Action Pub..

⁴ See Bezes, P., Demaziere, D., Le Bianic, T., Paradeise, C., Normand, R., Benamouzig, D., Evetts, J. (2012). New public management and professionals in the public sector. What new patterns beyond opposition? *Sociologie du travail*, 54, e1-e55; Broadbent, J., Jacobs, K., & Laughlin, R. (2001). Organisational resistance strategies to unwanted accounting and finance changes: the case of general medical practice in the UK.

Accounting, Auditing and Accountability Journal, 14(5), 565-586; Shore, C., & Wright, S. (2000). Coercive accountability: the rise of audit culture in higher education. In M. Strathern (Ed.), *Audit Cultures: Anthropological studies in accountability, ethics and the academy*. London and New York: Routledge.

Shore, C., & Wright, S. (Eds.). (1997). *Anthropology of Policy: Critical perspectives on governance and power*. Thomas, R., & Davies, A. (2005). Theorizing the Micro-politics of Resistance: New Public Management and Managerial Identities in the UK Public Services. *Organization Studies*, 26(5), 683-706.; Power, M. (1996). *The Audit Explosion*. London: DEMOS; Power, M. (1997). *The Audit Society: Rituals of Verification*. Oxford: Oxford University Press.

⁵ Wallace et al (2007), op. cit.

⁶ See for the concepts and terms 'translation' and 'brokerage': Mosse, David, and David Lewis. "Theoretical Approaches to Brokerage and Translation in Development." Chap. 1 In *Development Brokers and Translators: The Ethnography of Aid and Agencies*, edited by David Lewis and David Mosse, 1-26. Bloomfield.: Kumarian Press, Inc., 2006; Mosse, David. *Cultivating Development: An Ethnography of Aid Policy and Practice*. London and New York: Pluto Press, 2005; see also Latour, Bruno. *Reassembling the Social - an Introduction to Actor-Network-Theory*. Oxford: Oxford University Press, 2005.

⁷ I am indebted for the classical reference to Funnell, S. C., & Rogers, P. J. (2011). *Purposeful program theory: Effective use of theories of change and logic models* (Vol. 31): Jossey-Bass.

Annex 1. Survey Content

Experiences from the results agenda: Big Push Forward Survey

PAGE 1

This survey has been designed by the organisers of the Big Push Forward in preparation for next April's conference on the Politics of Evidence. The results agenda is much discussed, but we lack a body of evidence about practical experiences to understand its effects on achieving development goals. We are therefore seeking to learn more about the positive and negative experiences in relation to the 'results agenda' from people working in the development sector.

The term 'results agenda' is used here to describe planning, monitoring and evaluation processes that focus on the measurement of the impact, results and/or value for money of development projects. The results agenda leads to a strong orientation around specific measurements (intended / actual) as required by a donor, headquarters or a 'senior' partner organisation. The results agenda influences all stages in the decision-making, design, implementation and evaluation processes of planned development activities.

We are interested in any practical experiences, including both constructive suggestions and examples of effective practice as well as negative experiences. While this is directed to those receiving funds to implement development projects, we're interested in experiences of good/bad practice from donors, consultants and others with experiences of the results agenda too.

Any information provided will be kept confidential, unless otherwise signalled. The Big Push Forward will not release any information about specific examples without permission, although the aggregated information will be used in the course of the politics of evidence conference to inform the types of practical positive and negative experiences from results-based management. We understand that these issues are sensitive. Please provide answers to any of the questions to the extent that you are comfortable. We also include check-boxes on p.1 (questions 7 and 8) which allow you to agree for us to contact you.

PAGE 2

1. Personal Information

1. What is your last name? (Optional)

2. What is your first name? (Optional)

3. What is the name of your organisation? (Optional)

4. What type of organisation do you work for? (Optional)

- National Non-Governmental Organisation / Civil Society (in 'global South')
- International Non-Governmental Organisation
- Bilateral donor agency
- Multilateral agency
- Foundation
- Private sector donor
- Consultancy
- Independent Consultant
- Other (please specify)

5. What answers best describe your job description / title? (Optional, tick as many as are appropriate)

- Monitoring and Evaluation Officer
- Learning Officer
- Programme or Project Officer

- Grants Officer
- Accountability Officer
- Technical Advisor
- Senior Management
- Other (please specify)

6. Please provide your email address (optional):

7. Are you happy for us to ask you for more information using that email address? (Default option is no)

8. Are you happy for us to use your information as a specific example? (Default option is no)

9. Are you an existing subscriber to the Big Push Forward?

PAGE 3

10. Rate the impact of the results agenda on your organisation's ability to achieve its mission:

1. Highly positive
2. Somewhat positive
3. Equally positive and negative
4. Somewhat negative
5. Highly negative
6. Neither negative nor positive – no change

11. Rate the consequences of the results agenda to your daily work, include your relationships with partners and your ability to deliver on projects:

1. no changes at all
2. mildly significant changes
3. some changes
4. considerable changes
5. extensive changes
6. fundamental changes

12. Rate the consequences of the results agenda for your organisation's ability and willingness to learn

1. Great improvement
2. Some improvement
3. Both improvement and reduction
4. Some reduction
5. Great reduction
6. No changes at all

13. Rate the consequences of the results agenda to interactions within the organisation in which you work:

1. no change
2. some internal discussion
3. some internal discussion / changes in organisational ways of working
4. significant internal discussions / changes in organisational ways of working
5. extensive internal discussions / changes in organisational ways of working
6. fundamental change in organisational ways of working

14. What has changed most? (Pick one from each of the two columns.)

Most important area of change Second most important area of change

- What you spend your time on
- What you need to report on
- How you design or plan funding proposals
- What you are able to implement
- How you implement your work

- Your motivation or level of stress
- What you can approve for funding
- Relationships with other organisations
- How you procure and contract
- Other, please specify

PAGE 4

3. EXPERIENCES FROM THE RESULTS AGENDA

We are looking for descriptions of your experiences of the results agenda: the monitoring and evaluation processes which focus on measurement of the impact, results and/or value for money of development projects.

We are interested in small fragments of information as much as extensive experiences. We have given space for three experiences/fragments, but you can use as many as you want. It is fine if you prefer to disguise the identity of the people/organisations involved, or if you want to write it like a fictional story. As noted, experiences will be kept confidential and will not be used without permission.

If you would prefer to tell us your experiences with the results agenda in another way, please email Brendan whose address is: brendanwhitty at yahoo dot co dot uk

15. Experience 1. What happened? Describe what happened due to the results agenda, and what the outcomes were. You may want to include details about whether the example was positive or negative, who was affected and how.

16. What procedure, policy or practice does the experience focus on?

- Business case
- Strategy paper
- Logframe or other proposal
- Value for money
- M&E system
- Impact assessment
- Systematic review#
- Procurement
- Contracting
- Ad hoc data requests
- Recruitment
- Other (please specify)

17. Experience 2. What happened? Describe what happened due to the results agenda, and what the outcomes were. You may want to include details about whether the example was positive or negative, who was affected and how.

18. Experience 2. What procedure, policy or practice does the experience focus on?

- Business case
- Strategy paper
- Logframe or other proposal
- Value for money
- M&E system
- Impact assessment
- Systematic review
- Procurement
- Contracting
- Ad hoc data requests

- Recruitment
- Other (please specify)

19. Experience 3. What happened? Describe what happened due to the results agenda, and what the outcomes were. You may want to include details about whether the example was positive or negative, who was affected and how.

20. Experience 3: What procedure, policy or practice does the experience focus on?

- Business case
- Strategy paper
- Logframe or other proposal
- Value for money
- M&E system
- Impact assessment
- Systematic review
- Procurement
- Contracting
- Ad hoc data requests
- Recruitment
- Other (please specify).