CULTURES OF EVALUATION:

TALES FROM THE END OF THE LINE

ABSTRACT

Experiences in and around aid agencies suggest that the results agenda militates against a culture of learning and improvement to which evaluations should usefully contribute. In this paper I argue that this is an issue with both ethical and operational dimensions. The reductionist and simplifying effects of quantitative indicators of achievement as instruments of performance management, which I see as characteristic of audit culture, are having a pernicious effect in many aid environments. The technical and instrumental ways M&E is understood within the everyday culture of aid agencies preclude the space and time necessary for an ethically premised culture of learning. These social effects of audit culture detract from our capacity as a sector to deliver positive change, and should be of concern to evaluators, commissioners and consumers of evaluations.

Introduction: *from results based management to sausage numbers*

For a number of years, anthropologists and others have been concerned with the ways cultures of management and audit are affecting development (Mosse 2005, Eyben 2015, Shore and Wright 2015, Whitty 2015, Green 2011, Fukuda-Parr *et al* 2014). In this paper, based on experiences as an 'observant participant' (Green 2011) in the aid industry[1], I discuss some of the ways in which the trends towards results-based management (RBM) and other features of the new managerialism (Turner, Hulme and McCourt 2015) that have come to dominate the NGO sector over the last twenty years are having a profound impact on the way development results are conceived of and managed. These trends in management by audit have ethical implications, because of the ways in which they reduce and constrain the scope and kinds of thinking and research within aid agencies.

I argue that there are perverse incentives in RBM systems, which has led the 'results agenda' to produce its own "Frankensteinian offspring". This includes a culture of results that prioritises quantitative data over understanding, and which generates operational cultures in which doing is prioritised over learning. The widespread use of assessments that can be quantified and measured has led to the production of numbers that have a life of their own, and whose methods of production can be coercive, and hide awkward truths, rather than reveal them (e.g. Shore and Wright 2015, Camfield 2015). This results culture is bound up with performance management systems that produce self-regulating subjects, whose work is often

---

[1] In line with this positioning, the observations made here are from time under contract to which conditions of corporate confidentiality apply, hence the non-specific nature of the remarks here.

1

circumscribed by bureaucratic functions and structures rather than enabling holistic or learning focused thinking. This has significant implications for ethics, circumscribing the ways aid practitioners can

understand and engage with reflexive thinking about their work.

Sausage Fruit from the Results Tree

The wider culture of performance management via metrics can result in the exacerbation of the 'numbers

games', that are played periodically within development from crises around numbers of street children (Ennew 2008) to refugees and migrants entering Europe today, or whenever a particular advocacy or

fundraising mission is identified. In development today these are exemplified in the "sausage numbers" that

NGO workers report amongst their own work (Whitty 2015). They are sausages because no-one is quite sure what is in them, on account of the methodologically flawed processes by which they are formulated[2]. One

encounter I had with these involved the annual count of beneficiary numbers growing from the hundreds to

a few million. This had a positive effect on intra-agency perceptions of the programme to which they corresponded, but more worryingly for internal managers, the numbers had also made it through multiple

'rigour checks' before the error of a local partner's interpretation of the counting guidance was identified. No extra millions were in fact reached, but there they were, slipped through auditing checks and into the

annual report.

Such problems may be the exception rather than the rule, but the systems that produce them are what are in question. Indeed despite investment in M&E across the NGO sector, this increased attention has to a

large extent been focused on producing numbers of dubious accuracy rather than learning (Woodhill 2007). Many years ago I was part of a tiny team of two members of staff with responsibility for managing global

results reporting in an INGO. After some restructures and in line with the new number games around

proving results and numbers of people reached by our work, our small team grew exponentially to five posts. The change stemmed to promise new opportunities to get systems right and, at long last, the changes

which we did or didn't achieve for beneficiaries would become paramount. M&E would be brought in from

end-of-project and technical obscurity to be a key driver for improving the effectiveness of future work. M&E was even rebranded, with phrases and acronyms that often included learning, such as MEL. But our

optimism was brief, tempered by the fact this growth was accompanied by the announcement of a new kind of metrics: counting beneficiaries. Protests that this would almost inevitably be methodologically unsound

were futile, and some years down the line, had led to the kinds of sausage fruit referred to above.

---

[2] These echo the 'poor numbers' ascription Jerven has given to statistics such as GDP in some African states, where the

errors in formulation mean they are not fit for purpose (2013).

The scope of our work as M&E staff included managing both what is probably most readily understood as evaluation in the aid community - externally researched and written end of project/ programme assessments of the workability and effectiveness of an approach, policy or programme – as well as other processes more internal to aid agencies. This second type included products like annual reviews, project impact assessments, and country annual report components, done in-house according to internal policy on reporting requirements, and managed within the kinds information management systems and cultures described here.

These latter kinds of work may be seen as a peculiarity or even anathema to external specialist professionals that such processes are also categorised in the same set of practices as evaluation, given that they do not often meet the strict standards of scientific rigour and validity as those worked on to an external audience (even if they aspire to, while functioning within different and often severe resource constraints). So it is in relation to both types of roles and products that I refer in this paper[3].

An example of the internal practices within which evaluation and evaluative cultures interact is a study a colleague and I commissioned of what had happened to emergency response evaluations in an INGO over the preceding years. The findings confirmed our instincts, and fears[4]. A significant number of them were received via the practice of actively burying bad news, or simply shelved. In one example, a damning evaluation was kept under a for-your-eyes only policy, shuffled between a few select desks on conditions of secrecy. Failure is not an option in a results-driven culture, where success and delivery are all, carrying with them reputations of both organisations and individuals that must be managed. Burying under the carpet lessons that challenge the logframe or strategic plan becomes a tactical response under such conditions. A corollary to this is the fact that learning is considered an abstract exercise rather than a foundational ethical commitment, something that can be done with sufficient expertise but is tangibly separate from the day to day work of an NGO, as exemplified in the ways it's compartmentalised within many INGOs. This includes, in one NGO, an annual internal and lightly competitive process of sharing lessons that were small enough not to risk damaging reputations, with wider parts of the organisation, such as how a project had incorrectly targeted the wrong group of vulnerable people and needed to be reformed by learning from talking to beneficiaries. While such learning endeavours were undoubtedly important, and indeed the competition was an initiative to promote learning, but it was within a culture in which the very idea of learning had been so squeezed out the everyday busy-ness of doing that it needed reinstating.

---

[3] Indeed more work could be done to document these, expanding on the helpful comments of many such staff to illuminate this sometimes murky area of practice in Whitty (2015), and consider both the synergies and differences between such practices and the specialist professional work of the first category of work outlined above

[4] An assessment of DFID's learning processes had some similar conclusions about the use of learning from evaluative processes there (ICAI 2014).

How did we get here?

The cultures inside aid agencies have played their part. Indeed the cultures of audit argument initiated by Power (1997) has prompted further reflections on the systems (e.g. by Shore and Wright 2015) which produce a picture of a results–driven audit culture. This culture is complete with Power's 'rituals of verification' in which the gaze of power narrowly reflects on what Eyben has termed "artefacts of evidence" (2015), evaluative products, such as business cases or social returns on investment, that are ultimately focused on value for money.

This unfolding of culture occurs because "processes for assessment have social consequences, locking up time, personnel and resources" (Strathern 2000: 2). These social consequences of audit include the de-prioritisation of learning, with both a lack of space and time dedicated to it in the busy world of the modern INGO. The detail of how staff time and bureaucratic processes are locked up is mundanely woven into the everyday fabric of people's work.

A key effect is that time is systematically dedicated to generating the numbers required at the regular rituals of verification (Power 1997) that are used for managing both people and programmes such as six monthly performance reviews and annual quantitative indicator data collection exercises.

In one large INGO, the annual rounding–up of quantified results data took up at least two months of the working lives of country office M&E staff, and a similar amount of that of staff in head and programme offices to collate and organise data. In addition to these processes, much of the preceding year would have been given over to ensuring everyone understood the correct methods for inputting data for the indicators, and ensuring that every programme and project was collecting the right data to pass on upwards (although this was never perfect, hence the sausages). This data was organised and processed according to key predetermined indicators, which were then passed higher up the ranks of management to those authorised to interpret the data, in a manner not dissimilar to Harper's (2000) observations of IMF data management systems on country 'fact missions'[5]. Here, Harper notes that rank and seniority determine who is allowed to analyse the data 'correctly' and then sanction the relevant interpretations. Such processes privilege the views from specific vantage points within managerial hierarchies. The active exclusion of multiple voices and interpretations, echoes questions central to this volume's discussion of ethics, and to where people's rights fit into the picture (see also Barnett and Munslow 2014). It also illustrates the processes by which data can become increasingly concrete and powerful the further removed they are from their source (Camfield 2015).

---

[5] And echoing processes of calculation discussed in Mitchell's (2002) account of governing Egypt and Rottenburg (2009) on metadata.

Such processes also raise questions about individuals' subjectivity in audited systems, as time is devoted to doing and to being able to record the achievements of effort against individual and programme

performance metrics that constitute keystones of personnel and programme management. These processes

aim to produce individuals "who will calibrate their work and worth against their organisation's performance indicators" (Shore and Wright 2015: 26). Whilst such compliance is not inevitable and NGO

workers may resist, and propose alternatives, changes and improvements to the way things are done, the scope for this and the potential impact of such work depends on both management styles, and the kinds of

top-down strategizing and definitions of what counts as success or a result. The reality is also one in which

many aid staff work extremely long hours, in a competitive industry. They also often fit into a culture where people are so busy doing, so that they can show they delivered, and that the programme achieved its

intended outputs and results, that there is often little time or space to think or reflect (Eyben 2015).

These rituals of verification of quantified artefacts in a culture of audit are aligned to specific power

imbalances that recall the roots of audit in European feudal society, where it was a quintessentially unequal

affair (Shore and Wright 2000). While there are clearly profound differences between medieval hierarchies and those inside aid agencies, the edifices of both the feudal audit and contemporary audit culture and

practices are constructed on foundations of mistrust (Douglas 1992, in Strathern 2000: 4). Such an ethos of suspicion and doubt is in deep epistemological and ethical contrast to the more participatory approaches

that have in recent decades been an important counterpoint to power imbalances inherent in much

development intervention. While some evaluation practice today draws on these shifts (e.g. Cornwall 2014), this push upwards is against the might of hierarchical power and funding accountability, and its

quantified audit systems. We seem a long way from asking what evaluation look like if it was done by the

beneficiaries, for the beneficiaries, so they could choose who 'intervened' in their communities and not. The kinds of ethics of commitment that might prevail in such circumstances would be both highly different

from evaluation as normal, as currently practiced, and provide a very different space for the articulation of the meaning of development outcomes.

Ethical Consequences: bitter fruit

In this atmosphere of what I call 'metric madness', another INGO example includes annual programmatic reports where learning was reduced to the space of a few sentences in a pre-designed Excel spreadsheet

(with macros locked), and whose nuggets of insight into how something had or hadn't worked could only be

accessed by disentangling the text from the surrounding numbers. These meanwhile, sought to measure every possible area of effort, if not result. By focusing attention on that which can be counted, outputs

easily get called outcomes, the result of the expeditious use of M&E vocabulary to fit narrowing management objectives and perspectives, in which simple results chains are key, and complexity too

complicated. As such key results are kept to the simplest levels of outcome, with data coming directly from programmes. In such data-generating systems, the proximity of data to the programme provides the key

basis for its reliability, rather than being subject to external validation or scrutiny.

The kinds of complex and subtle changes for beneficiaries that M&E practitioners like myself had dreamt of

largely slipped back into our dreamscapes, where they had been before the large investment in results.

Internally, M&E has in many places gone from obscurity directly to metrics, understood as a form of performance management. Metrics has become the byword for success, and indeed the term is often used

as a short-hand for M&E. What counts is very much what can be counted as Cornwall (2014, echoing

Einstein), wryly observes.

A lack of time and space for learning thus persists amidst a reality in which M&E is still under-resourced,

under-used, and misunderstood. For many esteemed colleagues, the bottom line was that M&E was boring, little more than a technical exercise in collecting and organising predetermined data. If monitoring is often

a formal and ritualistic collecting and aggregating of data for donor reports, filed to collect dust until the evaluator turns up, it is little coincidence that the M&E appears at the back of project proposals, and at the

back of management budgets and priorities. Power's (1997) descriptions of colleagues' dim views of the

interest of accountancy sound very familiar, and reinforce the parallel with technical auditing of figures. Indeed the subcontracting by DFID of some programme management to large management consultancy

firms whose roots are in accountancy underscores the extent to which management by numbers seems a

plausible option within a reductionist view of what development entails, and seems to stem from a position in which contextual understanding and technical expertise are irrelevances[6]. Given the concomitant

emergence of debate within development circles about the need for flexibility, adaptation and complexity of local context for successful interventions (e.g. Ramalingan 2012 and 'Doing Development Differently'),

this seems a counter-intuitive move.

Meanwhile however, the pressure on country office staff to produce the right numbers seems relentless.

That task is also shared with partners, whose capacity to deliver the kinds of robust auditable figures the

results agenda demands is often questionable. An example of the challenges of ensuring rigour in such cascaded responsibilities is given by Pasano[7] (2015), where the number of children reported by a partner is

'sometimes'. Pasano's bewilderment at the array of ways partners record and count the outcomes of their

efforts is indicative of the challenges of cascading down responsibility for reporting in audit focused systems, where the kinds understanding of the methodology and systematic accounting practices required

for accurate audit may simply not exist. Indeed it echoes the issues of translation and brokerage analysed

---

[6] Indeed, Shore and Wright note the power of the 'Big 4' international consultancy firms, who via a range of services segueing with the governance agenda, were operating in 140 countries with revenues of $110 billion in 2011/12 (2015:25).
[7] This is a pseudonym, of an NGO worker writing for the Guardian development professionals website

elsewhere in development practice (Mosse 2011, Olwig 2013) and raises its own set of practical (and ethical) questions. Whilst such approaches evidence the messy contested and negotiated meanings of aid in practice (Lewis and Mosse 2006, Rottenburg 2009), the power of numbers and the finance to discipline and impose meanings across time and space constituted a powerful strand within such potentially disputed terrains. As results culture is extended via the cascading of results production to the partners in developing countries, it is worth also noting the culturally disciplining and time-defining effects of this. The imposition of definitions, metrics, and methods of what counts and how it is to be counted is not only a frontier of governmentality and power, but one whose confines produce dysfunctional consequences for both the audited organisations and the beneficiaries represented.

Of course, one of the key reasons for the rise of the numbers game is the increased rigour demanded in accounting for funds, and to donors, and of donors to their tax payers. In many instances, NGO and CSO partners have to report to multiple donors, from multilaterals to ministries, and a small organisation may find the tasks of collecting the data for the specific indicators and other metrics demanded of each donor NGO a significant task. The likelihood that they will have much flexibility in what data they are allowed to collect, and with what methodologies is often quite low given the top down nature of reporting requirements.

 A key problem produced by management by results, and cultures of audit, is that such processes of negotiation, and of translation, are subject to being funnelled back to the pre-determined numbers that the RBM systems have *a priori* determined as meaningful, correct and relevant. The values that matter, which count, are those of the audit derived system, and the histories and lessons produced by the vast social networks and brokerage processes by which development is really enacted, are written out of the history of development.

In theory, NGOs could be in a position to challenge the logic, when faulty, of donors, and to provide a fix to the broken feedback loop in the relationship between donors and recipients, whereby donors "impose decision on recipients without the latter having a say in the process" (Martens 2006: 694). This pivotal role for NGOs would be fine, if it worked; and the suggestions evoked from the results agenda and audit culture are that it often does not. NGOs role is also subject to wider debate within development (e.g. Banks Hulme and Edwards 2015), with some arguing that within the broader politics of development NGOs have become too in thrall to donors, and lost their critical voice and perspective[8]. This means they may fail to play the key communicative role between donors and recipients that Martens envisions, or in the views of those with a more radical agenda, act to have the transformative social effects that others aspire of them (Banks Hulme and Edwards 2015). Arguably, if processes such as participation have been manipulated into forms of co-

_____

[8] This is of course a much wider debate which here I can only hint at.

optation for beneficiaries, so too they have for NGOs talking their seats at the tables of power in New York and Geneva, and in taking money become complicit in the political machinations of their paymasters

There are fundamental questions about power here, and who is left out of these processes of knowing (and sometimes learning) is important. Access to knowledge is also about access to power, and although many organisations now publish evaluations on their websites, this act of openness is not the same as honest internal discussion about the findings. Furthermore, evaluations can be for and about donor priorities rather than programmatic learning agendas. Evaluations may carry the robust qualities of scientific rigour and external validation, but often in so doing can neglect the more procedural 'how' questions from which better practices be drawn[9]. Good practice needs to be about ensuring staff can be involved in at least some evaluation processes so they too can contribute to and learn from them, and discuss their findings.

However such approaches require a kind of flexibility that is in many instances at odds with the bureaucracy that is seen to characterise many aid organisations[10]. Within the busy everyday practise of agencies, and given the internal political weight evaluations can carry, both time and hierarchy may conspire to prevent all those who might find evaluations useful from being able to access them. Evolutions are also rarely translated into the languages of project staff, let alone beneficiaries.   The result is that many people are left out of discussion about learning, and most certainly the intended beneficiaries.

Conclusions

The ethical premises of audit culture stand in stark contrast to an ethos of learning and openness that are at the foundations of an ethical practise in development. The implications for learning as sketched out above are about the ways in which organisational cultures and daily working practices in aid may be malignly affected by the pressures of management or governance by numbers.

The independent and objective nature of evaluative analysis means it can easily be managed according to hierarchical and exclusive management structures. In the current culture of results inside many aid organisations, knowledge generated by external evaluations risks being neglected.

Audit culture's relentless focus on quantifiable results sits at odds with thinking and working in ways where learning by mistakes is valued. Regular, participatory and active monitoring are undeniably a logistical challenge in many contexts, but also a challenge to the status quo in many areas of practice, which is likely to be the more significant challenge in pressured management by results environments. Instilling an

---

[9] Noting of course that TORs and the quality their execution are amongst the many variables affecting the quality and utility of evaluative reports.

[10] As per in an interesting account of the use of Political and Economic Analysis (PEA) in DFID and the World Bank by Yanguas Gil and Hulme (2015)

ethically premised culture of learning requires tackling the questions of what kinds of knowledge are valued within organisations. It means making time and space within work programmes and scheduled to think reflexively, openly and with a view to learning. Organisations need to value what staff learn about how mistakes are made, and how they can be fixed. Changes to the way evaluations are understood and used requires a cultural shift away from neo-liberal management techniques and cultural practices that promote doing, to ones that promote learning, collaboration and understanding.

References

Barnett, Chris and Tamlyn Munslow (2014) *Framing Ethics in Impact Evaluation: where ae we? Which route should we take? Workshop Report,* Evidence Report #98, October, Brighton: IDS

Banks, Nicola, David Hulme and Michael Edwards (2015) NGOs, States, and Donors Revisited: Still Too Close for Comfort?, *World Development* Vol. 66, pp. 707–718,

Cornwall A 2014 *Using participatory process evaluation to Understand the Dynamics of Change in an Nutrition Education Programme*, IDS Working Paper No. 437, February

Ennew, Judith (2008) *Has research improved the human rights of children?,* presentation at University of Swansea

Eyben, Rosalind, Irene Gujit, Chris Roche and Cathy Shutt, eds (2015) *The Politics of Evidence and Results in International Development: Playing the game to change the rules?*, Rugby: Practical Action Publishing

Eyben, R. (2015) Uncovering the Politics of Evidence and Results: A Framing Paper for Development

Practitioners   http://bigpushforward.net/wp-content/uploads/2011/01/The-politics-of-evidence-11-April-20133.pdf

Fukuda-Parr Sakiko, Alicia Ely Yamin & Joshua Greenstein (2014) The Power of Numbers: A Critical Review of Millennium Development Goal Targets for Human Development and Human Rights, *Journal of Human Development and Capabilities*

Green, Maia (2011) 'Calculating Compassion: accounting for some categorical practices in international development', in D Mosse ed, *Adventures in Aidland*, Berghahn Books: Oxford

Harper, Richard (2000) 'The social organisation of the IMF's mission work', in Strathern ed *Audit Cultures*, London: Routledge

Independent Commission for Aid Impact (2014) How DFID Learns, http://icai.independent.gov.uk/wp-content/uploads/How-DFID-Learns-FINAL.pdf accessed 22 June 2016

Jerven, Morten (2013) *Poor Numbers: How We Are Misled by African Development Statistics and What to Do about It,* Ithaca NY: Cornell University Press

Lewis, David and David Mosse (2006) 'Theoretical Approaches to Brokerage and Translation in Development', In D Lewis and D Mosse eds, *Development Brokers and Translators: the ethnography of aid and agencies*, London: Kumarian Press

Martens, Bertin (2005) 'Why Do Aid Agencies Exist?, *Development Policy Review*, , 23 (6): 643-663

Mitchell, Timothy (2002) *Rule of Experts: Egypt, Techno-Politics, Modernity*, London: University of California Press

Mosse, David (2011) Introduction: The anthropology of expertise and Professionals in International Development, in Mosse ed *Adventures in Aidland*, Oxford: Berghahn Books

Mosse, David (2005) *Cultivating Development,* London: Pluto

Olwig, Mette F. (2013) Beyond Translation: Reconceptualising the role of Local Practitioners and the development 'Interfaced', *European Journal of Development Research*, Vol 25, 3, 428-444

Pasano, Dana (2015) Confessions of a humanitarian: The number of children we reached this month is 'sometimes' http://www.theguardian.com/global-development-professionals-network/2015/dec/04/confessions-of-a-humanitarian-the-number-of-children-we-reached-this-month-is-sometimes?CMP=share_btn_link

Power, Michael (1997) *The Audit Society*: Rituals of Verification, Oxford University Press

Strathern, M ed. (2000) Introduction, in Strathern, M., ed., *Audit Cultures*, London: Routledge

Shore, Chris and Susan Wright (2015) 'Governing by Numbers: audit culture, rankings and the new world order', *Social Anthropology* 23, 1, pp. 22-28

Shore, Chris and Susan Wright (2000) 'Coercive Accountability', in Strathern ed. *Audit Cultures*, London: Routledge

Turner, Mark, David Hulme and Willy McCourt (2015) *Governance, Management and Development: making the state work*, 2nd edition, London: Palgrave

Whitty, Brendan (2015) 'Mapping the results Landscape: insights from a crowdsourcing survey', in R Eyben, Gujkit, Roche and Shutt, eds *The Politics of Evidence and Results in International Development: Playing the game to change the rules?*, Rugby: Practical Action Publishing

Woodhill, Jim (2007) M&E as Learning: rethinking the dominant paradigm, in Graaft, J. de; J. Cameron, S. Sombatpanit, C. Pieri, and J Woodhill, eds, *Monitoring and evaluation of soil conservation and watershed development projects*, CRC Press: Boca Raton, Florida

Yanguas Gil, P., & Hulme, D. (2015). Barriers to Political Analysis in Aid Bureaucracies: From Principle to Practice in DFID and the World Bank. *World Development*, *74*, 209-219