Trickle Down
how narrative control shapes programs and their evaluations

Saville Kushner
University of Auckland

> - *We should do more to work from small samples and, especially, case studies;*
> - *We need to intensify work on the validation of narrative approaches and case study;*
> - *Evaluation should do more to elicit and exchange multiple views of any program;*
> - *We should resist the temptation to reduce the evaluation experience to a single narrative – go for properly explicated options rather than recommendations;*
> - *We ought to forget Qual vs Quant – focus on Stake's (2003) 'criterial/experiential' approaches;*
> - *We should document contestation over results;*
> - *We should do more to emphasise context;*
> - *Evaluation associations should shift focus from delineating evaluation competencies to broker conversation with evaluation sponsors on the validity of methods and the publication of evaluation reports.*

The evaluation field is in something of a tangle. Equipped with methods, purposes, leadership and relationships that were designed for economically and morally expansive times, it wanders like a migrant artisan in this age of austerity. Out of time, nurtured on the teat of burgeoning welfare states and public sectors, this still-liberal, humanistic discipline meets the new age with discomfort. It is made incompetent by the unfamiliar, sometimes illiberal demands of technical compliance, forced back into that narrow part of the evaluation spectrum that is occupied by regulation, inspection and the policing of results. No wonder evaluation associations feel bound to state and restate their competencies.

In this inauspicious world progressive evaluation is forced back on softening the angular edges of performance management, cuts and 'outsourcing'. Logic modellers, results managers, outcomes measurers and impact assessors, methodological gold standardeers and 'what workers', auditors, performance managers, inspectors and regulators do our honourable best to continue to advance social aims, crowding around 'strong' designs for the security they provide. Hoping for fixed co-ordinates that confirm much-needed gains, we narrow tolerances and control variables. We take a fix.

Elsewhere, we shape evaluation in other images – we lean towards action research, affirmative action, collaboration, participation, 'personalisation' – seeking to force evaluation into a humanistic embrace. We notice that people are sometimes fearful of evaluation and so we discipline it with indulgence and courtesy – make it 'live', go for nurture over scrutiny. We broaden the tolerances, proliferate variables so as to capture real experience and worry less about the co-ordinates than about the people being co-ordinated.

Neither are quite to the point. Still the foundations of welfare states and public sectors erode, wealth inequalities inexorably widen, corrosive relationships between economic status and educational achievement endure, the social safety net thins out. There is not much evidence that evaluation prevents the corrosion. Here is the context for our work in the 'developed' world.

OECD statistics for Gini Coefficient	mid-1970s	late 2000s
USA		
After taxes etc.	0.316	0.378
Before taxes etc.	0.406	0.486
UK		
After	0.268	0.342
Before	0.338	0.506
NZ		
After	0.271	0.331
Before	0.408	0.455
OECD ave.		
After	n/a	0.314
Before	n/a	0.457

Note: The higher the rating the higher the inequality in income distribution. All movement is towards growing wealth and growing inequality. There is some slight evidence of a redistributive effect of taxation and transfers (e.g. pensions), but overwhelmed by growing disparities. In each case, GDP rose over the period by around 15%. As countries became wealthier, so that wealth was redistributed with growing inequity.

Source: http://stats.oecd.org/Index.aspx?QueryId=26068

-o-o-o-o-

Of course, it is not the evaluator's job to change the world – that is for program people. We document their attempts to do so. But so persuaded of the worth of programs we evaluate that we sometimes succumb to enthusiasm for getting involved, and the evaluation pegs itself to results, joining the clamour for evidence of impact.

The impact focus has allowed us to take our eyes off the ball of programme quality. Program quality is the real accomplishment. Results can be manufactured, short-term, contextual, unattributable, indeterminate in their merits and scale, contested, distracting, irrelevant, politicised, partisan, miss the point – or signify real gains. But they are not *pedagogical*. We don't *learn from* results, we appreciate them – in fact, results too often displace learning. Program quality refers to the *curriculum* of the program, the learning site, an understanding of how a program pulls together intellectual and material resources and how it deploys them – the nature of the contraption that produces results – the culture that is the program. Here is a definition of program culture (taken from an early evaluation) – it is the source of program quality:

Programme Culture

We may talk of a programme as a culture – by which we refer to the texture of its social, intellectual and political life. Programmes exhibit characteristics of cultures – rules, roles, relationships, rites and rituals – but also the archetypal cultural condition of a tension between individual and collective. All cultures define conditions of collective association; but culture can only be realised through individual thought and action. It is important for the evaluator – even those unconcerned with portraying experience – to be aware of how and when a programme exhibits a robust culture. Successful programme culture refers to social and organisational conditions which help to create shared understandings about what words

and actions mean and within which interactions can take place with the minimum of negotiation but a tolerance of argument. They are conditions which encourage people to orientate their individual actions to the aims of the programme. Such conditions would be made up of a common vocabulary, sustained personal contact and a core (not a totality) of common values and interests together with a tolerance of where those values and interests diverge. A programme culture is an achievement rather than a design; it is recognised through a feeling of community more than through statements of allegiance to common goals - it is, that is to say, experiential rather than rational. In a successful programme culture individuals can find meaning in their work from the achievements of their colleagues.

We seem to have set aside some hard-won lessons that arise from this view of a program: that no innovatory program has a singular logic or aims; that no program rests on a consensus – internal or external - over how it should act; and that no program can claim unity or certainty over its results. Or, put another way…..all innovatory programs embody a diversity of logics, some political and organisational, some individual or collective…all programs are sites for constructive argumentation…all program accomplishments, where they can be identified, can be judged in different ways against different (meaningful) criteria. And a range of validity challenges prevent us working back from results to make inferences of program quality. These have served as foundational premises for program evaluation.

The point being that all programs have multiple explanations – diverse narratives.

It follows that affirmation, action research and some participatory approaches position us within the values frame of *some* program constituents and therefore erode our independence and our critical warrant. There should be those who nurture and tend for programme people and their ideals – but there should also be those who maintain a critical distance and who can entertain the paradox of the right people doing the wrong thing, who can confront the petit deaths and failures others have the license to tolerate or the courtesy to ignore. Programmes are political sites, competitive markets for ideas and priorities – they have 'winners' and 'losers'. If evaluators do not measure a program's productivity with impartiality, who will? We have to identify results, gains and losses, but we also have to understand their provenance, and how to give them a sense of scale and meaning. There is rarely a good case for hiring a substantive expert to conduct the evaluation.

-o-o-o-o-o-

Some of our clustering has been around supposed differences between quantitative and qualitative approaches. Though not exclusively, impact-focused people tend towards number and scale, whereas process/relationship focused people lean towards portrayal and narrative. This, too, has been a distraction (Cook, 1997) – methodology displacing the democratic focus of evaluation. But Stake (2003) argues that whatever differences there are among us they are not well represented by Qual/Quant. Too many of one fulfil the promises of the other. He proposes a dynamic relationship between 'criterial' and 'experiential' approaches. Little recognised, as yet, he has redefined the field (again).

If we are to understand an educational event (says Stake) we need criteria, and so we face a choice between discovering criteria 'elsewhere' and importing them (e.g. national standards, performance indicators), or else we generate the criteria within the case. The former denies context; the latter emphasises context. The former speaks of rights outside of the case (e.g. citizen rights) whereas the latter speaks of rights (to be represented) *within* the case.

So we can ask our 'what workers' and our 'impact assessors' where they derive criteria – from transferable or *sui generis* sources. And we can try both to achieve what Stake calls 'binocular vision' – gaining depth of field by using each to moderate the other, making

transparent and allowing for methodological bias. We can do the same with our action researchers and 'affirmatistes'.

We need to explore Stake's schema more in empirical ways, test it out. It sounds important. It promises an approach to evaluation more suited to the times we live in, critical but not dismissive of alternatives.

But what I want to do is to focus on another – somewhat related – dimension of difference and alternative. This comes down to a broader disarray of political discourse which can reproduce itself in evaluation practice. It goes back to programs having multiple and competing logics. The implication of this is that evaluation should be honouring all logics, eschewing favour and privilege, constitutionally impartial. The problem arises when all possible narratives are reduced to one, and when that is sustained with techniques of narrative control.

Evaluation foundations were laid on values pluralism (MacDonald), argumentation (House), multiple aims (Weiss), uncertain and emergent goals (Scriven), 'progressive focusing' (Parlett & Hamilton), unstable generalisation (Cronbach) and the idiosyncrasies of context (Stake). These were the ornamentation of expansive times, well-stocked economies, confident government. But evaluation today – from impact assessment to affirmative action - too easily leans towards singular explanations – cause-effect, 'what works', program logic, practitioner value, participatory consensus. These often tell authentic stories – but they are always only one among numerous – and often not the best. Where does this tendency towards singular explanations – single narratives – come from?

-o-o-o-o-

Evaluation thrives or survives in its contexts – financially, obviously – but intellectually, too.

It seems that the political world is increasingly subject to 'single narratives' – singular explanations of social, political and economic events that avoid dispute, are selective in evidential claims, and which seek to reduce uncomfortable complexities to managed simplification. The grand single narratives set a tone and a logic of governance. 'Education is not ideological', announced Tony Blair, bringing an end to the constructive clash of values and elevating technical (single-story) solutions to schooling. It brought an end to curriculum debate, and so an end to the essential intersection of government and ethics. Henceforth, there was only one story in town: *educational achievement*. Again, outcomes displace quality.

The casualty of the single narrative is public understanding – but also a loss of citizen agency. Arid land for the evaluator to graze. Here are some examples you will recognise.

- 'weapons of mass destruction….immediate and present threat….war on terror' (*there were no WMD to speak of – no discernible threat. The lack of evidence was exposed, the narrative was challenged and it fell. But it stood long enough to sustain the 'war on terror' narrative…and to wage terrible war*)
- 'man-made climate change…our lifestyle is unsustainable' (*there is some small evidence for this, but for alternatives, too – with the balance strongly against man-made climate change, and somewhat against any unusual climate change. The narrative still stands, drives international discourse, alternatives are branded 'denials', children are being robbed of hope.*)
- 'debt crisis…public sector unsustainable at current levels…need for austerity' (*little general evidence of debt crisis, stronger evidence of banking crisis, growing evidence-based consensus that public sector cuts lead to debt crisis. The narrative*

stands and drives global fiscal policies and the expansion of poverty, alternatives are branded 'deficit-denials'.)
- 'randomisation and controlled experiment are a methodological gold standard' (*Glaxo-Smith Kline was recently fined US$3bn for, among other things, manipulation of RCT evidence.*)

One of the more successful has been the narrative of anthropogenic climate change and the disputed role of the UN International Panel on Climate Change (IPCC). Numerous studies have challenged the validity of IPCC procedures, membership, reports and leadership (Michaels, 2005; Laframboise, 2011) and the selective use of evidence (and even its distortion – Montford, 2010[1]). If rational argument were to prevail, climate change – anthropogenic or natural – would be a matter for extended public and scientific debate, with the IPCC cast in the role, not of cheer-leader for a single narrative, but as ring-master in the interplay of diverse claims. The role of the IPCC is as evaluator of the evidence and convenor of public debate. Instead, it has succumbed to the petty temptations of influence and simplification. The single narrative prevails.

-o-o-o-o-o-

Some narratives are the detritus of reasoned argument - 'you can control cholesterol with diet' was a [failed] experimental hypothesis that became a social mantra. Others are manufactured with a purpose ('war on terror'). Either way, they lend themselves to narrative control as soon as they intersect with interests or proclivities. Is mammarian screening helpful or harmful? Does school achievement coincide with good education? Is nuclear power a threat or a benefit to the environment? Sometimes the matter is decided by the loudest voice or the deepest pocket – sometimes for good, other times not. Sometimes reasoned deliberation enters on the basis of evidence. That cannot be guaranteed. With a pre-prepared policy in mind, narrative control is an attractive option.

But to exert narrative control is not to win it. That is done more by consent than coercion. The 'age of anxiety' persists and creates the psychosocial conditions that make us disposed to believe – or to suspend disbelief. To my knowledge, four Nobel economists (Stiglitz, Krugman, Pissarides, Sen) have denied the basis of austerity along with Martin Wolf, an editor at the *Financial Times* and other leading economists – but they whistle against a gale, they carry insufficient weight in an argument driven by something other than evidence.

The New Sociology would have it that this consent – this 'something other' - is based on the psychosocial conditions of 'fear' – that we live in a world whose excessive benefits come with constant anxiety over their potential loss. What exacerbates this is the disconnection of the individual from collectivities, the erosion of groups and a growing sense of facing threats alone (Durodie, 2007) – the sum of all this is a loss of effective agency. This in turn leads us to look to authority – especially political authority – to alleviate the fear (unjustifiably, says Zygmunt Bauman, since the political classes, too, are isolated from effective agency and equally subject to fear). Hence, our vulerability to simplification and narrative control.

"An exaggerated perception of risk lends itself to growing demands for greater regulation and social control. Accordingly, people increasingly look to those in authority to enhance their sense of security by mitigating the worst effects of the natural world and human society, as well as the actions of those who seek to change them." [Durodie, 2007]

-o-o-o-o-o-

---

[1] – but see, also, the refutation – Mann, 2012

These are unreasonable burdens for evaluators to carry. With our diminishing resource base we will not change the world – and nor is it our job to do that. And nor will we assail the formidable walls of the United Nations or the IMF nor reduce social panics. But neither can we ignore these things. At the most local level, the power of single narratives threatens to distort relationships. To reiterate, evaluation is touched by its contexts: methodologically expansive in economically expansive times; methodologically ascetic in times of austerity. But so are program actors. They, too, are cue-conscious. Narrative control trickles down. How do our evaluation contracts play *in micro* in relation to the diminishing of the public sector and the imposition of narrative control?

In all evaluations – from $1,000 to the $100,000 and beyond - we may confront exaggerated risk, the leaning towards simplicity, unwarranted deference to authority, fear of instability. In any evaluation we face choices in the framing of evaluation questions and analyses which enhance or displace agency; in any evaluation our courage may be tested in insisting upon open reporting. All evaluations demand that we are wily in smuggling in sometimes inconvenient ethics. But it is true, too, that the trickle-down of narrative control touches evaluation method.

-o-o-o-o-

Fear and anxiety extend to evaluative enquiry and possibly account for the internalisation of evaluation into the administrative system with the inevitable erosion of independence; the assertion of 'methodological gold-standards'; and the common practice of withholding evaluation reports from the public domain. I speak only of the USA and the UK – pending a firmer grasp of ANZ contexts – but I offer the following view from the UK.

The laceration of public sectors in the UK (80% cuts to university funding; 30% to probation and social work; 20% to the police; 20% to schools) comes with one of the more immediately damaging single narratives: that public sectors are non-productive, over-extended and unaffordable. This narrative holds sway, persuades fearful publics. It might be challenged by the plethora of evaluation reports on public sector work that have been commissioned over the years. But it is not. Too many evaluation reports have been shelved out of anxiety and caution, made 'confidential' by a culture of risk aversion and internal accountability. Others are methodologically proscribed, fashioned from their design into a given narrative form and content, limited to impact assessment. The *'exaggerated perception of risk'* runs unchallenged by multiple narratives – there is no prominent narrative of the worth of public sectors. I have lacked courage in my own $1,000 - $100,000 evaluations – or else self-censored my evaluation out of sympathy with beleaguered sponsors.

No-one has benefited from this, much less the citizen whose quality of life and security depends on public works and who remains too-frequently unaware of what is contained in many unread evaluation reports: public accomplishments, cherished qualities which lie behind performance shortcomings, and alternative stories which speak of complexity rather than the simplistic 'worked/did not work'. Certainly there is little benefit to the administrative system that, over time loses its grip on the complexities of the programs it administers.

Eleanor Chelimsky was head of the Government Accounting Office of the USA and a sponsor to the early emergence of program evaluation. She has recently written (2012) of 'the politicisation of the evaluation process' through the control of method.

"the organisational context is…an important determinant of what methods are used…these decisions have become increasingly influenced by a single narrative – a narrative that sees increasing numbers of government programs and policies embodying a single idea, or positing a simple, one-to-one cause-and-effect relationship, both of which are established, not

by evidence, but rather by suppressing existing evidence that is inconvenient to the particular idea or relationship being advanced." [p.77]

Of course, this conceals complexities – culpability dissolves in the detail. Those who peddle single narratives do so, often, unwittingly or with honourable intent. My experience in recent years has been of working extensively with official agencies who commission evaluation out of that sense of beleaguering, caught between a perplexingly complex practitioner system below, and a fearful and demanding political system above. They almost always want to know 'what works' – it is almost always the wrong question. What they need to be able to know is where quality lies in that which they oversee – are their programs durable? Do they make sense in the knowledge world of the practitioner? Do they change people's lives for the better? Do they get the point? In 2004 a group of foundational evaluation theorists met for a sixth time (since 1972) in Cambridge University to produce a new manifesto for evaluation in troubled democracies. The first item in the manifesto read:

Evaluators should provide – and be free to provide – information that government needs *as well as that which it* wants

[Elliott & Kushner, 2007]

In societies built on the sophistication of ideas and confidence in confronting complexity, we are all losers in the game of simplification.

-o-o-o-o-o-

Whether or not public sectors are any longer affordable is not a matter to be ruled on by economists, and nor is it subject to an algorithm – the issue is multi-faceted. The same goes for climate change, for mammarian screening, the size of school classes and urban regeneration, policing, and child protection. These are matters for rational deliberation and consensus. Again, I speak from the UK and the USA. Had I lived and worked longer in Aotearoa New Zealand I might speak with less anxiety and caution, more with hope and experience. The Treaty of Waitangi process – with its flaws – and the prominence given to kaupapa Maori with values grounded in collective action, trust and collaboration offer a different kind of 'trickle-down'. At first sight (for me) these are potentially firmer foundations for democratic evaluation methodology than the scientism of 'gold standards'. The Waitangi modelling of evidence-based argumentation leading to consensus seems to me to be saturated with social contractarian values that should underpin all evaluation action. As Thomas Paine, almost alone, would have understood implicitly, all government should be bound by a Treaty of Waitangi-type settlement covering ownership of our public sectors. That settlement would provide the space for evaluators to provide information that is needed, as well as that which is wanted. And that space would have to be filled with the urgent chatter of contested narratives.

How much of this can be ameliorated by evaluators and evaluation commissioners? How might we return to evaluative cultures that acknowledge the inherent instability of program assumptions and the democratic imperative to proliferate narratives?

Go back to the beginning.

- *We should do more to work from small samples and, especially, case studies;*
- *We need to intensify work on the validation of narrative approaches and case study;*
- *Evaluation should do more to elicit and exchange multiple views of any program;*
- *We should resist the temptation to reduce the evaluation experience to a single narrative – go for properly explicated options rather than recommendations;*

> - *We ought to forget Qual vs Quant – focus on Stake's (2003) 'criterial/experiential' approaches;*
> - *We should document contestation over results;*
> - *We should do more to emphasise context;*
> - *Evaluation associations should shift focus from delineating evaluation competencies to broker conversation with evaluation sponsors on the validity of methods and the publication of evaluation reports.*

Contact: s.kushner@auckland.ac.nz

References

Chelimsky, E (2012) 'Valuing, evaluation methods and the politicization of the evaluation process' in *New Directions in Evaluation*, #133, American Evaluation Association, NJ: Wiley

Cook, T. (1997) Lessons Learned in Evaluation over the Past 25 Years, in E. Chelimsky (Ed.) *Evaluation for the 21st Century,*, Thousand Oaks: Sage, pp. 30-52

Durodie, W. (2007) Fear and Terror in a Post-Political Age, *Government and Opposition*, (42:3), pp. 427–450

Elliott, J. & Kushner, S.I. (2007) 'The need for a manifesto for educational programme evaluation' in *Cambridge Journal of Education*, 37(3), pp. 321-336

Laframboise, D. (2011) *The Delinquent Teenager Who was Mistaken for the World's Top Climate Expert*, Toronto: Ivy Ave. Press

Mann, M. (2012) *The Hockey Stick and the Climate Wars: Dispatches from the Front Lines*, NY: Columbia University Press

Michaels, P.J. (Ed.) (2005) *Shattered Consensus: The True State of Global Warming*, Oxford: Rowman & Littlefield (Esp. see Introduction by Michaels)

Montford, A. (2010) *The Hockey Stick Illusion*: *Climategate and the Corruption of Science* London: Stacey International