Outcomes, Accountability and Community & Voluntary Organisations: Holy Grail, Black Hole or Wholly Possible?[1]

Garth Nowland-Foreman, LEAD Centre for Not for Profit Leadership
(garth@lead.org.nz)

Context: The Winds of Change

Since 1984 the New Zealand Government was an early and bold adopter of neoliberal approaches to public administration that collectively came to be known as *new public management*. Specific elements included an increased reliance on market-oriented strategies such as deregulation, privatisation, outsourcing, the structural separation of purchasers and providers, an enhanced emphasis on performance measurement, a shift from input to output-based funding, and the delivery of public services by third parties under contract (Boston, et al. 1996). However, the extent of adoption of these principles to actually *become* the very system of public administration was unprecedented; no other country has demanded accountability through contract-like arrangements to the extent that New Zealand has (Schick, 1996:1).

As part and parcel of these reforms, community and voluntary organisations, "once ancillary to state-provided public services, [were] enlisted as state service delivery agents." (Butcher, 2015:38). As one large voluntary social service provider recounted at the time: "Then, in 1981 came a 'letter of understanding' setting out 'mutual' expectations - [at first it just seemed like] putting in more legalistic terms what had always been agreed by both [the government agency and the voluntary organisation]. By 1985, the gentle winds of change turned in a hurricane that blew through an unsuspecting voluntary sector. Across the country, voluntary organisations [previously] receiving government grants were shaken and disheveled by mini-tornadoes of contestability, accountability, the funder-provider split, performance indicators and the overwhelming and icy winds of managerialism. Where previously a close partnership and joint interests had shaped the relationship…, the emphasis was now on independence and separate identity; on controlling and containing the relationship through formal contracts which initially specified inputs and then focused on outputs. The basis of the former partnership was irrevocably changed." (quoted in Nowland-Foreman, 1998:109).

Iconoclastic British accountant, Michael Power (1994), has identified what he calls the "audit explosion". This emphasis derives, he argues from two powerful but contradictory

---

trends for government in OECD countries: less spending, but with more controls. Increased pressure to contract out, down-size, and decentralise, while at the same time exerting greater control over the very functions that have just been made autonomous. Power believes this is at the heart of the drift towards what he disparagingly calls "managing by numbers" - a trend which ultimately made performance management a core public administration policy.

Even without the availability of consistent definitions and measures, Government funding in New Zealand for voluntary and community organisations clearly increased rapidly from a relatively steady level up to the 1960s (estimated at around $4 million for *voluntary social services* in 1967), to a conservative estimate of government funding for *voluntary welfare organisations* of around $1 billion in 2002, and an estimated $2-3 billion government funding for *all community and voluntary organisations* by 2004 (various sources quoted in O'Brien, et al. 2009:13). Although there were some notable exceptions (increasingly so in recent times), this was less a case of 'contracting out' (the classic privitisation of previously state-provided programmes), and more a case of 'contracting in' community and voluntary organisations to exercise greater specification and greater control over them by government (albeit more at arms-length and using the tool of the contract) (Nowland-Foreman, 1995:4). Performance management as governing thus came to involve both centralisation, utilising new and intense forms of control, and decentralisation of responsibility for programmes (Hoggett, 1996, quoted in Paton, (2013:21).

In the decades since these reforms were first instituted there have been perceptible shifts in rhetoric (away from a more combative *agency theory* which stressed voluntary organisations' self interests, and more towards a *partnership* approach). Disenchantment with the excesses of '*contractualism*' grew beyond disaffected voluntary organisations, spreading even to one of *new public management's* principal international proponents when he reviewed the practical impact of its thorough-going implementation in New Zealand (Schick, 1996 & 2001). The department mainly responsible for implementing many of these reforms with voluntary organisations was soon itself warning an incoming government that "current arrangements with the sector based on purchase-of-service contracting may not, in themselves, be sufficient to maintain a healthy not-for-profit sector (Department of Social Welfare, 1996: 22). Despite some (largely symbolic)

changes to machinery of government arrangements[2], even under a change of government the momentum to undo the mechanics of principal-agent contracting effectively stalled. And more recently, even these largely symbolic revisions were unpicked and the momentum reversed, after a subsequent change back in government.

There may be many reasons why the (widely acknowledged) excesses of 'Contractualism' appear not to be able to effectively moderated. Perhaps it is the seductive simplicity and remarkable consistency of the underlying Agency Theory. Perhaps its because, in the three decades since the reforms, the public service may have been largely purged of anyone in senior leadership roles who is not an advocate of the reforms, or who, as a new younger public servant, has only ever had experience of this approach to public sector management. Schick (2001:3) also suggests that "in contrast to other countries in which reform meant adding peripheral elements to the pre-existing managerial system, in New Zealand, the reforms *are* the system. There is no other managerial system. This means that dismantling the reforms would require the government to divest itself of the ways in which it prepares and administers the budget, runs departments, links ministers and managers, and decides what to do." Regardless of the reasons it has stuck, and despite widespread criticism, the New Zealand version of Contractualism has also set up the country as a natural adopter of Outcome measures.

Meanwhile in the 1990s on the other side of the world, the call for government to be more 'results-oriented' by funding Outcomes rather than Inputs (Osborne & Gaebler, 1992) and for a greater focus on evidence and Outcomes in philanthropic giving (United Way of America, 1996) was gaining increased prominence and adherents. By the beginning of 1997, 22,000 copies of the United Way manual, *Measuring Program Outcomes: A Practical Approach* (1996), had been sold to agencies who are being asked by United Way and others to provide measures of effectiveness in their grant applications (Walker & Grossman 1999:1).

In New Zealand, Contractualism had originally focussed on holding both government agencies and the organisations they contracted with, to account for more readily specified and countable Outputs. While there were some earlier attempts by a previous government to pilot *Funding for Outcomes* (2003), and to promote (a version of) Friedman's *Results*

---

[2] These changes included two hard-hitting inquiries conducted jointly with community & voluntary sector representatives, establishment of a Minister with the portfolio of the Community and Voluntary Sector, establishing an Office for the Community & Voluntary Sector, a Prime Ministerial-endorsed Statement of Government Intentions for Improved Community-Government Relationships (known affectionately as the SOGI), and various pilot initiatives to promote 'joined up' funding, reduce compliance costs, establish more long-term and stable funding, and progressively address part-funded contracts, etc. (Butcher, 2015).

Based Accountability framework [3] (Pomeroy, 2007), these were both halted as Ministers and governments changed. However, more recently there has been a sudden and thorough flurry of activity under the current government which directly or indirectly picks up the growing international pre-occupation with Outcomes and Results:

- In 2012, the new government established its own *Investing in Services for Outcomes* initiative within the Ministry for Social Development: proposing funding decisions driven by the government's priorities; establishing a 'Capability Investment Resource' to improve quality of funded social services and better achieve outcomes; and promising more streamlined, results-based contracts. This has since been proposed to be even more tightly aligned with four specific government priorities to be reflected in 'results measures' in 'providers' contracts under the Ministry for Social Development's new *Community Investment Strategy* (June 2015). It also promises an increased emphasis on 'building the evidence base', 'improving quality of data collection', 'further simplifying compliance requirements' and 'continuing to build provider capability' to implement this approach under an overarching Results Measurement Framework (https://www.msd.govt.nz/documents/about-msd-and-our-work/work-programmes/community-investment-strategy/community-investment-strategy.pdf).

- In March 2013 Cabinet directed the Ministry of Business, Innovation & Employment, through its Government Procurement Branch, to lead a 3-year *Streamlined Contracting with NGOs* project, to reduce duplication and compliance costs. This included developing standard Outcome-Based Agreements, and *Results Based Accountability* (Friedman, 2005) is once again incorporated within the contracting framework as the preferred mechanism to support an increased focus on Outcomes in government contracting (http://www.business.govt.nz/procurement/procurement-reform/streamlined-contracting-with-ngos).

- In 2013 Cabinet authorised the Ministry of Health to trial a *Social Bond*. Separate Registrations of Interest were requested in 2013 and 2014 for Outcomes and Service Providers. As at the time of writing a major NGO and a retail bank are in negotiations with the Ministry of Health as potential partners

---

[3] Results Based Accountability is a particular outcomes-based performance management tool developed by Friedman (2004) that has gained ascendency in numerous government funding programmes in Australia, New Zealand, and elsewhere. It is "broadly defined by two underlying ideas: the justification of practice on the basis of outcomes (results); and the demonstration of these outcomes by evidence" (Houlbrook, 2011:46).

in the first Social Bond. As a complicated financial mechanism based on 'payment for results', Social Bonds require well-developed Outcomes and especially rigorous, agreed and unambiguous measures for these Outcomes (http://www.health.govt.nz/our-work/preventative-health-wellness/social-bonds-new-zealand-pilot/social-bonds-progress-date).
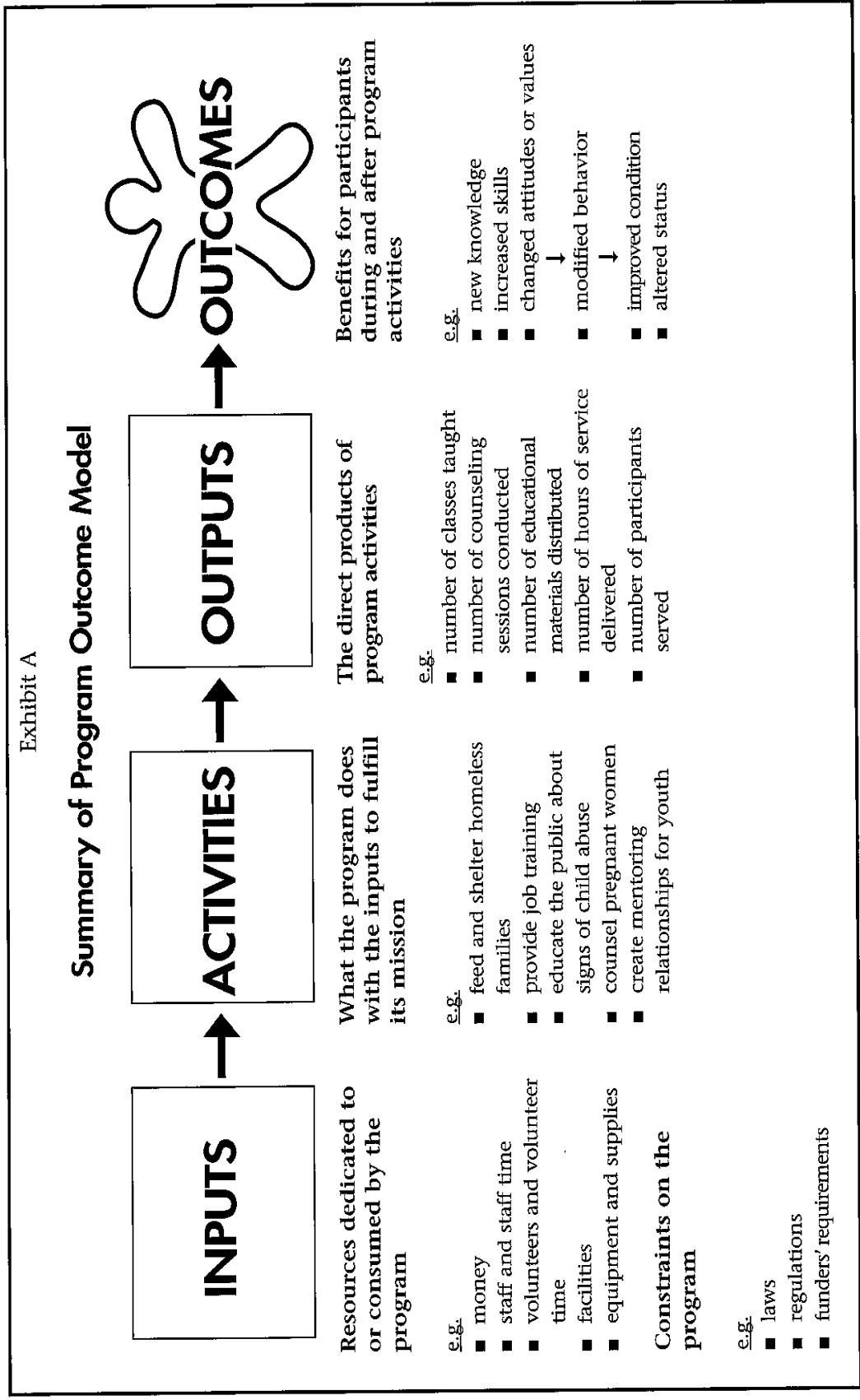
- In June 2014, the Productivity Commission, on reference from the Ministers of Finance and State Services, launched an inquiry into *More Effective Social Services*, "paying particular attention to commissioning arrangements". It released its overdue final Report in August 2015 which favours vouchers (to encourage responsiveness and innovation) 'navigators' to help people with complex needs get the services they need, and intermediary third parties or 'social insurers' to make commissioning decisions. It identifies a range of commissioning options from in-house provision to vouchers ('individualised funding'), proposes more contracting for Outcomes, and highlights the importance of monitoring and evaluating, the collection of standardised 'big data', and a stronger evidence base to identify (and spread) what works and discard what doesn't (http://www.productivity.govt.nz/sites/default/files/social-services-final-report-main.pdf).

What is an Outcomes approach and where did it come from?

Plantz, *et al* (1997:6) outline the history of non-profit performance measurement that led to the current focus on Outcomes by the United Way of America (a early philanthropic leader in both adopting and promoting Outcome measures):

- Initially, performance measurement focused on *Financial Accountability* – documenting how funds were raised and spent, and the financial viability of an organisation. This is expressed in accounting and financial reporting standards. For example, the United Way of America's 1974 *Accounting and Financial Reporting*. However, while an organisation may be financially viable, it may not necessarily be achieving anything.

- So organisations soon began to measure what the funds allowed the organisation to generate, namely it's *Programme Products or Outputs*, such as number of counselling sessions held, brochures distributed, days of care provided, etc. For example, in 1976 United Way of America's programme classification system defined 587 human service categories and suggested product measures for each. While such reports told us how much service or activity was provided, it didn't necessarily tell us anything about the quality of those services or activities.

- Concerns with *Standards of Quality in Service Delivery*, such as staff qualifications, staff-to-client ratios, specific service delivery practices, record keeping, confidentiality, conditions of facilities etc, soon led to the formation of accreditation and certification groups. The Council on Accreditation of Services for Families and Children, formed in 1977, was one of the first independent accrediting bodies in the United States. However, if we knew what quantity of service of an acceptable quality were provided, it still wasn't necessarily apparent *who* was being assisted (and who might be missing out).

- In the 1980s, funders began seeking assurances that organisations provided services to those 'most in need', 'hard to reach' or traditionally 'under-serviced'. This prompted collection and reporting of *Participant Related Measures* or client record systems, which usually included demographic characteristics (age, gender, income, race, marital status, area of residence, etc) and information about the client's presenting problem or status.

- As each of these measures were added onto each other and became increasingly onerous, the prospect of being able to collect and report on just a few telling measures (as purported to be so in the world of business) was increasingly enticing. Several major accounting firms and others created *Key Performance Indicators* for non-profits, often ratios among various categories of inputs, services, outputs and total costs. For example, in 1984, the accounting firm Peat Marwick produced such a seminal resource (Elkin & Molitor, 1984).

- Dissatisfied that generic management indicators still said little about the services they were providing, and in response to an upswing in use of customer satisfaction surveys in commercial organisations, there was a push later in the 1980s from non-profits and their accrediting bodies to measure *Client satisfaction*, as an over-arching summary of quality assurance. Elements of staff courtesy, condition of facilities, etc, as well as overall satisfaction measured also often included feedback from clients, users or participants on physical and cultural accessibility, timeliness of service, etc.

- Increasingly, it was recognised that 'satisfaction' alone is not always the purpose of a non-profit or its programmes, especially in commonly occurring contexts of information asymmetry. And it didn't necessarily say anything about whether the organisation's activities had made any difference (beyond being 'likeable'). Was anyone actually any better off as a result of the service? This is where *Outcome Measures* were potentially seen as having much to offer – shifting the emphasis from measuring 'busy-ness' or 'worthiness' to what good or change the programme

Exhibit A

Summary of Program Outcome Model

INPUTS → ACTIVITIES → OUTPUTS → OUTCOMES

Resources dedicated to or consumed by the program

e.g.
- money
- staff and staff time
- volunteers and volunteer time
- facilities
- equipment and supplies

Constraints on the program

e.g.
- laws
- regulations
- funders' requirements

What the program does with the inputs to fulfill its mission

e.g.
- feed and shelter homeless families
- provide job training
- educate the public about signs of child abuse
- counsel pregnant women
- create mentoring relationships for youth

The direct products of program activities

e.g.
- number of classes taught
- number of counseling sessions conducted
- number of educational materials distributed
- number of hours of service delivered
- number of participants served

Benefits for participants during and after program activities

e.g.
- new knowledge
- increased skills
- changed attitudes or values →
- modified behavior →
- improved condition
- altered status

An Outcomes focus is based on a logic chain moving from *Inputs* through *Activities* (sometimes called *Throughputs*), and *Outputs* to *Outcomes*, and sometimes with a further step of *Impacts* added. *Exhibit A*, on the previous page, includes the definitions and examples used by United Way of America in its first manual on Outcome Measures (1996:3).

In some respects, the relationships (or 'logical links') between each of the boxes or steps in the process are more important than the individual boxes themselves. This represents the programme logic behind what an organisation does. To check alignment it's possible to ensure each step to right, answers a 'why' question, and each step to left answers a 'how' question. For example 'why do we need these *Inputs*? In order to effectively undertake these *Activities*. How do we achieve these *Outputs*? By undertaking these *Activities,* etc.

The general idea of Outcome measurement is indeed simple: ruthlessly focusing on the *lasting benefits for participants* - and perhaps communities - not just paying attention to the *Inputs* used (producing measures of scope and scale), the *Activities* undertaken (producing measures of effort or busy-ness) or immediate *Outputs* produced (producing measures of efficiency or productivity). As Paton (2003:55) notes, in its most vigorous form, Outcome measurement is tied to a 'social investment' view of funding. Who can be against more careful specification of what grants or programmes are intended to achieve, and more systematic efforts to assess exactly what they do achieve?

The expectations of Outcome measures are certainly ambitious. Muir & Bennett (2014:8-9) have grouped the most commonly expressed expectations under three broad headings:

- *Knowing you really are making a difference*: so that all stakeholders (internal and external) can be assured that an organisation's activities achieve its mission, vision and values, and especially achieving positive change for the people and/or communities served;

- *Building better organisations*: creating a culture and systems for organisational learning and improvement, a better understanding of the people and/or communities served, an understanding what works and what doesn't work (in different contexts), and a basis for refining, shutting down or expanding particular programmes;

- *Accountability and transparency*: meeting compliance requirements and demonstrating accountability to funders and supporters, as well as the people and/or communities served, having a more compelling narrative to sell their organisation to potential supporters and funders, increasing trust and legitimacy, and better able to participate in new funding instruments often based on the capacity to monitor Outcomes (for example, individualised/voucher funding, and Social Impact Bonds).

Paton (2003:37) further summarises the reasons usually offered for supporting measurement-based performance management in particular:
- "Designing performance explicitly, and specifying the level expected, demands greater focus and *clarity* - it reduces the scope for obfuscation, ambiguity and misunderstanding.
- "Measures can *summarise* the important aspects of a complex situation making it easier to spot where and when expectations are not being realised.
- "Measurement expresses and encourages an approach to decision-making (including recognition and reward) based on *facts and analysis*, rather than anecdotes and opinion, not to mention wishful thinking and self-serving claims.
- "Measurement can assist *learning* by allowing greater comparability - over time and between groups - and by helping identify 'what works'."

Walker & Grossman (1999:1-2) argue, "If the alternative is not caring about Outcomes, or not caring about the reliability of how we assess them—not caring about accountability—then the recent emphasis on Outcomes is an unreservedly good thing. But, in fact, a more complex reality underlies the current 'Outcomes movement':

> *"Performance measurement has a chequered history even in the traditional manufacturing and bureaucratic settings ostensibly most conductive to measuring outputs and causally attributing outputs to individuals and organisational sub-groups. There are many more instances of dysfunction - instances where performance measures stimulate less than optimal or even counter productive behaviours - than there are instances of demonstrable success."* (Austin, Larkey et al., 1998 quoted in Paton, 2003:37).

But then it gets complicated…
Perhaps the first lesson is don't take yourself, or your Outcomes, too seriously.

The Outcomes model appears deceptively simple (and was deliberately designed to be so, Hendricks, *et. al.*, 2008), and textbook examples at first glance make its application look clear and straight-forward. However, when applying it to real life situations, things suddenly start to get more complicated. In the 1980s when two different government agencies in Australia, the Federal Industry Commission and the Victorian Department of

Human Services, were both pushing Outcome-based funding as logical, simple and straight-forward, the two agencies made the mistake of using the same example, of a counselling service, to illustrate their argument. Unfortunately, what one called an *Input*, the other saw as obviously an *Activity*, what one described as an *Output*, the other considered an *Activity*, and they both had very different understandings of what comprised the *Outcomes* (reference). There is a false and misleading precision in the drawing of firm lines around separate little boxes on a two-dimensional page and pretending these conceptual artifices actually exist in the real, physical world; when of course the concepts will always remain abstract and often ambiguous.

This is not necessarily a specific problem for Outcomes measurement, only. Though many advocates of Outcomes and their measures often seem to believe these are objective entities that just need to be accurately identified. As Box and Drapner (1987:424) identify "all models are false [because at best they are approximations], but some are useful", and warn that the real problems arise when the model (an approximation) is treated as if it were a concrete reality or absolute reality. Yet, if the very classification of activities is ambiguous and arguable, what might this say about the hoped for unequivocal results, when they do not emerge. As Paton (2003: 28-29) observes,

> *"[Most performance] frameworks will either be seen as not working well or fast enough in some respect... This is because many of the issues that social enterprises exist to address are not amenable to this sort of rationalised solution... [C]ause-effect relations are often poorly understood and in any case the issues are imbued with political and value dimensions that render apparently rational solutions unrealistic.[4] The point is not that the programmes are pointless and unwarranted, but that they only work sometimes, or up to a point, or for a time...*

> *"The problem is that the language of performance takes no prisoners. Through its lenses, the world is straightforward, situations are or should be controlled, the issues are clear, the criteria unambiguous – and results have either been achieved or they have not. Uncertainty, patchiness, ambiguity, riders and qualifications – all these can be read as excuses, signs of weakness... Hence performance management seems to be accentuating the familiar discrepancy between an organisation's public and private faces. Whatever the strains, stresses and*

---

[4] In fact, proponents of the United Way Outcomes model are explicit that in order to make the model achievable, it does not attempt to be an evaluation, and cannot make any claims to support assertions of causality or explain why a certain level of outcome was achieved (Hendricks, 2008: 16).

> *disorder that are occurring 'backstage', the 'frontstage' impression is still upbeat, a tale of rationality and order, progress and achievement."*

This leads to incentives for 'gaming' of Outcome measures, in order to artificially boost measured results, and underplay negative Outcomes - which is discussed further below.

A further aggravating factor is that most human services are complicated by having clients or participants. How do these fit neatly into the little Outcomes boxes? They are conspicuously absent from United Way's model illustrated in Exhibit A. Perhaps they do not figure in the model because, embarrassingly, the only box they could fit in, is as an Input to the programme. However, people might not like to be thought of as your programme's 'raw material'. They are certainly not inanimate objects to be 'manufactured' in the hands of a programme to produce an Output or Outcome. To complicate matters, they interact and (at the very least) co-determine the possible Outputs and Outcomes. The Outcomes model is essentially a manufacturing 'production line' metaphor, and that is why it doesn't always sit neatly with the complex, interactive, ambiguous, messy world of human services.

This why the distinction identified by Kendall and Knapp (1995) in approaches to understanding effectiveness and performance is so important. He distinguishes between those who think and operate within a rationalist positivist approach, and those who take a social constructivist approach.

And then there are the practical measurement considerations...
A second important lesson is that if you are finding Outcome measures hard work and frustrating, you are probably on the right track.

In part, this is because of the considerable practical difficulties associated with a rationalist measurement framework. Paton (2003:39) summarises eight broadly accepted attributes of performance measures if they are to effectively promote learning, feedback and motivation. Namely, they should be:
- *valid and reliable* (non-manipulable),
- *parsimonious* (relatively few, so stakeholders' cognitive limits are not exceeded),
- *comprehensive* (covering all significant dimensions of performance),
- *acceptable, meaningful and credible* to a wide range of stakeholders (have wide face-value),
- *pervasive and integrative* (allowing aggregation and comparisons, internally and externally),
- relatively *stable* (so performance can be tracked over time),
- designed with *explanatory power* (to aid diagnosis of performance and support

learning, not just enable counting), and
* *practicable* (able to be promptly and relatively easily reported).

However, Paton (2003:39), with the support of Meyer (1998:xv) asserts that in practice no measurement system ever manages to combine all these attributes for any sector, and the barriers are likely to be even more acute in nonprofit contexts (finding support for this proposition also from Herman & Renz, 1997, and Forbes, 1998).

Frequently this leads to an overload (Paton, 2003:40 quotes an example where one programme had 169 separate performance measures), and 'indicator oscillation'. The use of a limited set of indicators encourages staff to neglect unmeasured aspects, so new measures are progressively added, ultimately exceeding cognitive limits, resulting in pressure for a tighter focus and greater prioritising. This cull leads to a more limited set of measures, and so on. Paton's case studies (2003:75) and the author's own experience find similar patterns of: oscillation in number of measures used; a gradual increase in the number of measures in use, and turnover (or churn) of the measures in use (see also Nowland-Foreman, 2000). "Overall, the prospect of a stable, coherent and focussed measurement set seems as far off as ever" (Paton, 2003:76).

Schick (2001:13) similarly concludes that in practice, "...measuring outcomes is a frustrating, fruitless task [even for governments themselves. They] invest considerable resources and years later have little to show for it. Managers and policy analysts fight over what is to be measured, over the causes and effects of change, over who should be accountable for what, and so on. They fight over whether a particular result is an immediate outcome or an end outcome, as if these terms had scriptural meaning."

Paton (2003:55) also questions whether the sorts of goals pursued by community and voluntary organisations "including the quality of their work, and the experiences of participants, can sensibly be reduced to uni-directional service delivery outcomes. In dementia and hospice care, for example, the process is all, and the 'outcome' a forgone conclusion."

The reality is, as one moves along the logic chain from Inputs to Outcomes (and Impacts), progressively it:
* Becomes *harder* to actually measure anything
* Becomes more *ambiguous* and less clear what is actually happening
* Means you have *less control* over what happens
* Means you need to *wait longer* to see what actually happens, and
* Is increasingly effected by many more confounding, and often unanticipated *external factors*.

All in all this means the closer we get to measuring Outcomes or Impacts, the more expensive, time-consuming, uncertain and unreliable the measures become. These technical measurement difficulties entrench a tension between measuring what is most important, and what is more readily measurable. This tension is discussed further below, as it can also have important mission-distorting effects.

Regardless, the technical measurement challenges remain substantial. Even among a cohort of early-adopters and enthusiasts of the United Way of America Outcomes Framework, a 1999 survey of 390 organisations (quoted in Hendricks et. al., 2008:31), found about 50 per cent reported that measuring programme Outcomes overloaded their record-keeping capacity, diverted resources from existing activities, and led to a focus on measurable outcomes at the expense of other important results. More than 50 per cent found it difficult to identify appropriate Outcomes, indicators, methods, and data collection tools. About 60 per cent reported insufficient staff time available and inadequate computer hardware and software capacity to store and manipulate Outcome data. More than 60 per cent were concerned about the cost of measuring Outcomes. Despite what is often claimed, Outcome measurement is neither simple nor straight-forward, but a sophisticated and specialised skill, and inherently difficult.

If we are at all serious about attributing any cause-and-effect relationship between a programme and particular Outcomes, we will need some form of (preferably matched or random) control group, and have a reasonably large sample, over a reasonably long time period. This is because a measured improvement may have happened anyway or actually been greater without the intervention; and on the other hand no change may be a better result than a deterioration that may have occurred without the intervention.

But as Walker & Grossman (1999:5) and Hehenberger et. al. (2013:10) point out, sound comparison group methodology is not usually available to most organisations, the number of participants is often not large enough, and the comparison group's activities are often not distinct enough to permit a sound conclusion about a programme's impact - often because of the expense and the specialist skill sets involved. None of the organisations the author has worked with, for example, have been able to create control groups (even if they had no ethical qualms about doing so), though a handful have had recourse to general population statistics, with which to make some comparisons.

As Hehenberger, et. al. (2013:13-14) observe, calculating social impact or Outcomes requires one to adjust raw measures of Outcomes for (i) what would happen anyway ("deadweight"), (ii) the action of others or external factors ("attribution"), (iii) how far the Outcome is likely to reduce over time ("drop off"), and (iv) the extent to which the

original situation was displaced elsewhere, or Outcomes displaced other potential positive Outcomes elsewhere (which could be either positive or negative).

It is widely recognised that some programmes are more amenable to defining Outcomes and their measurement than others. In particular, conventional hard indicator-based approaches do not lend themselves to the assessment of Outcomes that cannot either be predetermined or readily quantified.   The literature and the author's experience with various organisations suggest that frequently this includes capacity building, community development, policy advocacy, and many preventative programmes.

The short time scales of most approaches, and the need to be able to report impact soon after a project is completed, rarely allow for the assessment of *long-term impact.* Also, the more time that is allowed to pass between an activity and the assessment of its impacts, the more the impact of the activity will be diluted by external factors.

Unanticipated impacts, if not central to an organisation's mission or objectives, are also unlikely to be top priority for staff or funders, hence the lack of tools to adequately address these issues. Unanticipated Outcomes also include the negative aspects of an organisation's impact, which are rarely captured using any existing tools.

Finally, 'collective impact' (Outcomes achieved by a cluster of organisations) is difficult to measure using any of the existing tools as the measurement process involves the co-operation of more than one organisation or project (Collis, et. al., 2003:13-16).

And then there are the mission-distorting effects...
A third important lesson is to be careful what you wish for (or promise).

> *"During the time of colonial rule in India, the British government became concerned about the number of cobras on the loose in Delhi. To address this issue they offered a bounty for every dead cobra the residents delivered. To begin with this approach worked as planned, with a large number of snakes being killed for the reward. But enterprising locals soon realised that they could breed cobras and kill them to claim the reward too. When the government realised that they had motivated this perverse behaviour they scrapped the bounty programme. But with their snakes now worthless, the breeders simply set their snakes loose in town, which made the cobra problem in Delhi worse than ever."* (Davidson, 2015:A13).

Kearns (1996) describes several common features of imposed accountability system, all of which apply well to Outcome measures:
   • Assume the worst in people and they may live down to your expectations,
   • Excessive project focus undermines purpose,

- · 'Accountability' can undermine responsibility,
- · Risks can be minimised, but so can opportunities, and
- · Imposed evaluations handicap learning.

Assume the worst in people, and they may live down to your expectations: There is evidence in the accountability literature that, ironically, the more people are treated as untrustworthy, the more they are likely to behave in just such a way. At one level this is just another application of McGregor's (1960) Theory X and Theory Y framework for understanding motivation. But even at a sub-conscious level, there is evidence that people do more of that to which more attention is paid. One organisation the author was working with, a Community Law Centre, originally was only measuring (and reporting) on the number and characteristics of clients. Each year the numbers being seen kept climbing. The 'results' looked spectacular, but in fact with relatively stable numbers of volunteers, increasing numbers of clients meant less time available for each, and as a consequence more clients' issues could not be resolved and needed to be referred on to private solicitors (which could have been counter-productive to their mission of 'improving access to justice'). As a consequence they changed their record keeping system to count number of clients whose issue was satisfactorily resolved (along with some information about test cases and law reform activities) and suddenly the exponential growth in client numbers slowed, and over time more and more clients' issues were resolved by the Centre. (Interestingly you'll note that the organisation chose Outputs for on-going monitoring and reporting, rather than Outcomes - though it did develop some narrative annual reporting on selected Outcomes.)

How much more so, might be the pressures where an organisation is funded and relies for its existence on achievement of certain Outcome measures. Elsewhere (Nowland-Foreman, 2009) I describe how leaders of a number of community and voluntary organisations reliant on Outcome funding in Australia progressively felt backed into 'cherry picking' or 'creaming' clients on the basis of those most likely to achieve the desired Outcome measures for the organisation - rather than prioritising, and often positively screening out, those conventionally seen at disadvantage or in greatest need. So common did this approach become that the sector even developed its own euphemism to describe this phenomena in less offensive terms - they selected participants on their "capacity to benefit"[5]. Stolte (2004) found a similar phenomena in similar circumstances in New Zealand. It was also one of the key risks Treasury officials warned the government, when the government wanted to pilot Social Bonds (New Zealand Treasury,

---

[5] Lest anyone believe this problem of 'creaming' is able to be simply dealt with by a two or three tier payment systems, depending on a 'degree of difficulty' of the client, this phenomena was occurring in a programme with exactly such a system. The difficulty is no weighted funding system can adequately capture the range and complexity of all real world clients.

2014).

As soon as 'league tables' of schools' National Certificate of Educational Achievement results began being produced, some Principals were accused of holding back some students from participating in the exams in order to improve their average results (Moir & Downes, 2015:A4).

An excessive focus on projects can undermine purpose and *'Accountability' can undermine responsibility*: This is a feature of 'contractualism' made even more explicit by Outcomes measurement: "The focus on ex ante specification has sometimes led to a checklist mentality which is positive from the perspective of having managers take accountability seriously for tasks they are expected to complete. It is less desirable if it narrows responsibility to simple compliance with what is on the list…" (Schick, 2001:12).

The difference between a 'check list' accountability and a broader sense of responsibility is tragically demonstrated in the field of individual performance bonuses, when in 1995, the year of the Cave Creek tragedy (where 14 Polytechnic students and one of their tutors died when a Department of Conservation viewing platform suddenly collapsed as a result of improper construction), the Department Chief Executive received his full performance bonus, as he achieved all his performance targets for that year.

> *"As soon as performance measures are used as a means of control, the people being measured begin to manage the measures rather than the performance. Incidents are reported selectively. Data are manipulated and presented in ways that make them look favourable. Individuals seek to undermine the measurement system."* (Neely, 1998:1 quoted in Paton, 2003:41.)

If unsubsidised employment on a certain day is the measure of success, then a job, any job on that day, it will be - even if, as the author has seen, that means the organisation has to employ the person itself. Even if that person would actually be better off in the long term, going back to school.

Gray (1997) observed it in local regeneration programmes in the UK, and in responses to the former Soviet Union's planning targets, Schuster (1997:257) describes the same phenomena in the use of performance indicators in the arts, and Smith (1993) found that the use of performance targets in maternity hospitals - where one might suppose that Outcomes might be more readily established - in terms of "tunnel vision, sub-optimisation, myopia, convergence to a norm, ossification, gaming and misrepresentation" (all quoted in Paton, 2003:41).

In New Zealand, a commercial private prison operator has been accused of under-reporting 'adverse events', fudging results, and using the transfer of prisoners (back into the public system) in an attempt to satisfy contracted performance requirements (Davidson, 2015:A5).

Even community and voluntary organisations can become so keen to meet their targets, that they loose sight of their original goals. In particular, collective, long-term and unanticipated Outcomes are rarely high on the list of priorities for many organisations and funders, and (as we have already noted above) there are also significant practical problems involved in measuring them, which means that these type of Outcomes may - as whole categories of activities - tend to be neglected even more (Collis et.al., 2003:13). The author has also observed this process of rendering invisible (because they are harder to measure and thus not on the target list of Outcomes) whole categories of activities related to capacity building, community development, policy advocacy, and many preventative programmes, and at times even any activity with a longer term time horizon.

The emphasis on individual organisation performance and narrowly defined programme results has been observed in both the UK (James, 2000) and New Zealand (Schick, 2001) as undermining collaborative action to such an extent that whole new public policies and incentives have needed to be developed to promote 'joined up' programmes. At times it seems that some funding programmes (like the Cobra Effect) create more problems than they solve.

Complex concepts like responsiveness and innovation have been observed to be 'crowded out' by efforts to ensure Outcomes measures are simple and do-able. Perversely what is more readily measurable becomes determinative of what is monitored and reported on, rather than significance or relevance. At one stage a New Zealand government department was toying with the idea of measuring as 'Outcomes' of one particularly complex and varied programme, the number of clients successfully completed in under 3 months, the number successfully completed in 3-6 months, and the number successfully completed in more than 6 months. It took one organisation the author was working with little more than a week to come up with much more relevant and useful Outcome measures (even though less precise as measurements), and these were quickly rejected by the funding agency, as they needed to have the same, easy to define measures nationally for this one programme. When asked why, a rather sheepish officer could only suggest, this was in order to be able to add up the programme results to be able to present them to their Minister.

The need for uniformity and add-ability of course will make innovation more difficult in practice, regardless of the frequent rhetoric that an Outcomes focus will free organisations up to explore and experiment whatever is the best way to achieve those

Outcomes. While convincing in theory, Paton (2003:31) found in practice:

> *"[T]he controlling tendencies of performance measurement will restrict the scope for experimentation. For example, in social housing many of the most important developments of the last 10 years concerning community-based economic development, training and neighbourhood improvement ('housing plus' is the usual term) had no place in the performance indicators used to assess housing associations during that period. These focused entirely on financial and managerial matters – voids, response times, etc – while ignoring the more important and difficult factors that affected the quality of life of tenants."*

Of particular concern, is that it is often the very features that make community and voluntary organisations distinctive, where they add value that are most at risk of being rendered invisible and slowly eliminated - particularly where they are harder to measure (Nowland-Foreman, 1998:121 and *Locality*, 2014:25 quoted in Neilson et. al. 2015:45). Recent research (Neilson, et. al. 2015:4) was commissioned in New Zealand to help identify and better understand how church- and community-based organisations make a difference, adding value above and beyond the funded Outcomes, that any effective organisation (large or small, commercial or non-profit, local or trans-national) could be expected to perform.

This research identified eight key attributes[6] of community and voluntary organisations which create their embedded 'organisation-specific capital' which collectively represents the 'community value' that enables community and voluntary organisations to: build connections among people (*social and cultural capital*); contribute to *social inclusion and cohesion*; contribute to the *empowerment* of individuals and communities; and assist with the development of stronger communities (*community development*). It concludes:

> *"When a government ignores or undermines the special characteristics, strengths and infrastructures of the community and voluntary sector it is likely to generate a service provision system that is highly individualised, disjointed from real needs, and is unable to build the overall strength of the community. This does not serve either government or taxpayer well, but more importantly for New Zealanders it will mean the community and voluntary sector will be unable to operate to its full potential."* (Neilson, et. al. 2015:8-9, 31-36).

_____

[6] Neilson et. al. (2015:16ff) identify the eight key attributes uniquely combined and prioritised by community and voluntary organisations as: the organisation's kaupapa and mission supporting its integrity; geographic, emotional, cultural and 'pro-active' accessibility; being embedded in the community; institutional and community knowledge from an independent perspective; networking and collaborative relationships; flexibility and choice around time; flexibility to respond innovatively and appropriately; and Manaakitanga.

Neilson et. al. (2015:37) also finds that the increased focus on quantifiable short-term Outcomes and 'rational' procurement/purchasing processes is often at odds with many of the aspects of the community and voluntary sector's comparative advantage - and, like Milbourne (2009), see it potentially resulting in "narrowing work, limiting targets to those that can be achieved, discouraging volunteerism, inhibiting developmental or higher risk work, and promoting competition over collaboration."

The four 'community value' Outcomes described by Neilson et. al. (2015:31ff) (building social and cultural capital; building social inclusion and cohesion; enabling community development; and empowering communities) show a remarkable similarity both to Nowland-Foreman's (1998:116) four civil society 'golden eggs' (encouraging active citizen participation; mobilising internal and external resources; promoting and developing leadership; and enhancing cooperation and trust), and Collis' et. al. (2003:55) four sector impact functions of: resourcing; service provision; voice/advocacy; and membership & representation.

Separate researchers in different research in different countries have identified a similar cluster of ways in which community and voluntary organisations can and do 'add social value' through who they are and how they operate. These are some of the very features that attracted governments to want to work with community and voluntary organisations in the first place. A narrow approach to Outcome measures risks, at best, ignoring these crucial features, and at worst undermining or even eradicating them. While a number of approaches are attempting to capture these 'civic' Outcomes, the field certainly needs further development (see, for example, Collis et. al. 2003:12).

Risks can be minimised, but so can opportunities, and *Imposed evaluations handicap learning:* While it is theoretically possible for Outcomes and their measures to be developed and defined in a 'bottom-up' process by each individual community and voluntary organisation, it is more common - at least in Australia and New Zealand - for government funders to pre-determine what the 'correct' Outcomes will be, and how they will be 'correctly' measured - for example, NSW Department of Community Services for its Community Services Grants Programme (Houlbrook, 2011 and Keeves, et.al., 2012), for the Job Services Australia funded by the Commonwealth Department of Employment Education and Training (Considine, et. al., 2014), for Training Opportunities administered by the then Skill New Zealand (Stolte, 2004), the New Zealand Ministry for Social Development for numerous community funding programmes (Pomeroy, 2007), including Family Budgeting Services, etc. Sometimes there may be nominal 'consultation' with community and voluntary organisations on proposed Outcomes and measures, but frequently even this minimal step is bi-passed.

This is perhaps no coincidence that Outcomes frameworks in their current form did not emerge from organisations concerned with how to govern or manage their own operation better, but rather came from a funder, United Way of America, with substantial moral and financial support from other funders such as J K Kellogg Foundation. Even though these bodies may strongly emphasise that improving programme effectiveness and organisational learning as the primary reason for Outcome measures, with external accountability merely as a secondary purpose.

And tell me again, what actually are the Outcomes of measuring Outcomes?
A further important lesson is that, weighing the pig doesn't necessarily make it any fatter (no matter how accurate or sophisticated the weight measuring technology is). Ambitious claims are made for Outcomes measurement, but ironically there are few Outcome studies on the actual Outcomes of measuring Outcomes.

Nevertheless, in a 1999 survey of 390 organisations, albeit early-adopters and enthusiasts of the United Way of America Outcomes Framework (quoted in Hendricks et. al., 2008:31), more than three-quarters reported that measuring programme outcomes helped them clarify programme purpose (86 per cent), focus staff on shared goals (88 per cent), identify effective practices (84 per cent), improve service delivery (76 per cent), enhance record-keeping (80 per cent), communicate results to stakeholders (88 per cent), and compete for resources (83 per cent). Overall, 74 per cent of agencies believed that "on balance, implementing outcome measurement has had a positive impact on this program's ability to serve clients effectively," and 89 per cent would recommend that a director of similar programme that s/he consider implementing Outcome measurement.

This sounds like positive findings, and it is, even taking into account that the voluntary early-adopters and enthusiasts are more likely to be favourably predisposed and to self-justify the time and expense of pioneering a new system. Interestingly though, this assessment does not attempt to use Outcome measures to judge success, but instead falls back to the earlier 'customer satisfaction' measures, and mostly focusses on self-assessments of *process* improvements, with only one self-assessment of Outcomes or impact in arguably the vaguest possible terms.

Considine et. al. (2014) attempt to assess the impact of the introduction of various aggressive incentives and techniques to 'be more business-like' (including notably the Outcomes-based funding system) on community and voluntary organisations providing training and other job-seeking assistance to unemployed people, funded under the Jobs Services Australia programme. They found that indeed the non-profit organisations did become and certainly their staff (front-line as well as managers) saw themselves as 'more

business-like'. They also found some evidence of 'mission drift', which supports concerns outlined above. However, they found *no* evidence that this led to any improvement in outcomes for their clients or communities.

An analysis of various studies on the impact of an increased emphasis on measurement of performance (Paton, 2003:77) in general found no evidence that this increased attention drove actual performance improvement in any direct way: "Performance improvement efforts were generally pursued independently of concern about one's publicly reported performance. When these were linked it was because particular organisations saw a strategic opportunity... On the other hand the direct and indirect costs of measurement can be considerable."

More specifically, Walker & Grossman (1999:8) conclude that "neither a greater emphasis on Outcomes (the notion that articulating Outcomes will increase their chances of occurrence) nor an improved technical approach to measuring Outcomes is likely to increase the probability of generating those Outcomes." According to their analysis, over the past 25 years, formal Outcome and Impact assessments have been instituted on three basic categories of activity: new programme initiatives, modified programme initiatives and longstanding programme initiatives, with the following overall pattern emerging:

- "evaluations of new programme initiatives are dominated by no or negative outcome and impact findings;

- "evaluations of modified programme initiatives are mixed in their findings of no, negative, modest or good outcomes and impacts; and

- "evaluations of longstanding programme initiatives are more likely to have modest or good outcomes and impacts".

This suggests, they argue, that "issues of operational capacity and implementation quality may indeed affect the Outcomes and impacts our studies are disclosing" and they note the massive review of employment and training evaluations done by the U.S. Department of Labor (1995:63 quoted in Walker & Grossman, 1999:8) came to a similar conclusion:

> *"It often takes time for programmes to begin to work. Many of the success stories in training for the disadvantaged have come from programmes which were operating for five years or more before they were evaluated."*

Organisations are still learning how to deliver the desired services and iron out programme kinks. Formal impact assessment and judgements about effectiveness should only come from assessment of mature or established programmes, they conclude. Walyer & Grossman don't argue against *any* monitoring or evaluation of new or demonstration

programmes.   Much can be learned operationally they suggest, for example, from the negative income tax experiments since the late 1970s ("What are the major hurdles for the programme? Does the programme model make sense? Is it operationally feasible? What refinements would enable the programme to operate better? What types of individuals are attracted to such a programme?")

The Policy Innovation Research Unit, funded by the UK Department of Health, published a substantial review of the literature this year on the empirical benefits to date from Social Impact Bonds (SIBs) and, indeed, payment-for-results contracts more generally, making the sobering observation was that "there remains a dearth of evidence for the assertion that SIBs will lead to improved Outcomes." It also found "no consensus in the literature on whether SIBs are best suited to innovation or 'scaling up'", and concludes:

"*The empirical studies highlight the high transaction costs of SIBs, the complexity of data monitoring and measurement as well as the ability for SIB delivery to offer space for innovation to service providers and improved Outcomes as exemplified by the first Outcome findings from the Peterborough SIB.*" (Tan, et, al. 2015:25).

Paton (2003:47) also reminds us that "...in some circumstances - especially those where measurement is difficult and cause-effect relations are obscure - the contribution of performance measurement may be more symbolic than instrumental, which doesn't mean it is unimportant. The existence of the system may satisfy institutional level requirements or evidence of purposeful and proper management, thereby sustaining legitimacy and ensuring continued access to resources (Meyer and Scott, 1983)."

Holy Grail, Black Hole or Wholly Possible: some possible pragmatic steps forward:

Despite all these difficulties, don't we still really want to focus on what results we are, or are not achieving? Isn't this what motivates most community and voluntary organisations to exist - making a difference? Are we left with nothing more than an idealistic faith and hope in charities?

In community and voluntary organisations the author has worked with there is no discernible reluctance to 'be accountable' (including accountability to a wider of stakeholders than usually referred to), nor is there indifference to what change they are making, what results are being achieved. If anything, there is more of a fascination with how to simply and easily know an organisation is 'on track' and 'making a difference'. Houlbrook (2011:53) reports similar phenomena among New South Wales community services.

Inputs alone will not satisfy, as they are just indicators of scope and scale (or success in

'empire building'). Activities are little more than indicators of 'busy-ness' or how fast you can run. And Outputs are ultimately little more than indicators of productivity - how efficiently we can get our 'products' off the production line.

I believe this brief review shows us that funders have an important role to play in doing things differently if we are to minimise the limitations and harvest useful contributions from Outcomes and their measures:

- *Hang loose and don't impose particular, standardised Outcomes or measures on organisations*. Schick (1997:61), a global New Public Management guru has some sensible advice on keeping Outcomes and their measures in perspective -

  "Outcomes should be seen not as measures of impact but as indicators of direction. They should be employed more for formulating policy than for maintaining accountability. They are powerful directional signals: are things getting better or worse? Have the initiatives taken by the government moved it closer or further from achieving stated objectives? Is the government progressing toward the conditions it favours? Particular outcomes may or may not be the product of outputs, but even when they are not, the government should take notice of them, analyse their significance, seek to explain what has (or has not) happened, and develop appropriate policy responses. Even if it is not accountable, the government should take outcomes into account."

  So, for example, if a significant indicator for us like birth weights are going down, or water impurity levels are increasing, it should be of concern to us regardless of whether we caused it or not. But funders should refrain from externally imposing particular Outcomes, measures or evaluative systems - as this reinforces an organisation's passive role, will result in at least some ill-fitting and inappropriate requirements, will handicap learning, and subtracts the potentially most useful part of the process - organisational conversations about 'What's the difference we want to make' and 'What does success for us looks like'.

  It is important for any requirements to be modest, realistic and patient.

- *Give more emphasis to Theories of Change and Logic models*. Next to conversations about 'What's the difference we want to make' and 'What does success for us look like', one of the most useful things that an organisation be encouraged to get clear and agreed on, is to work backwards from these 'ends' (their view of success) to develop a Theory of Change or Logic Model for how the organisation understands its day-to-day activities can reasonably be expected to work towards those desired 'ends' (contribute to that success). This also helps the organisation:
  - determine how reasonable the assumptions underlying their work are
  - identify intermediate outcomes that may be more readily monitored within a reasonable time-frame
  - justify their work to other stakeholders, including potential funders or supporters

- *Rely more on the 'accountability of a thousand eyes'.* Rather than construct cumbersome, costly and frequently less effective paper-based accountability system, it is usually cheaper (for both funder and organisation) and more effective to just require funded organisation to undertake some form of monitoring and evaluation of their choosing (proportional to the programme activity) and to publicly disclose the results, for example, on the organisation's and the funder's websites. This transparency will also ensure peer quality review, even if only of the most informal kind.

- *Offer support and capacity building for organisations to explore and develop their own learning systems.* On the same principle, developing its own approach to learning systems and being transparent about that is one of the cheapest and most effective ways of encouraging real learning. And given the purpose of any monitoring system is at least ostensibly to encourage organisational learning and improvement, it is worth giving some emphasis and support here directly, rather than just hope it might ultimately emerge from an Outcomes measurement framework.

- *Find stable exemplars and focus impact evaluation on a particular substantive idea.* Rather than expect every small organisation to reinvent the wheel, and evaluate the impact of each programme it operates, often prematurely when programmes are still 'finding their feet', it would be more strategic for funders to collaborate and focus some specific additional resources to enable higher quality impact evaluations among a smaller selection of stable exemplar organisations, and then share lessons widely with any organisation using similar approaches (Walker & Grossman, 1999:8).

As well as gently pushing funders to move on some of the suggested strategies above, community and voluntary organisations can also do some important themselves to seize their own destiny in the great rush to Outcomes and their measures:

- *Hang loose, just start and stop with Purpose.* Whether imposed or not, using whatever Outcome system, it is important for organisations to keep these in perspective, as metaphors rather than reality, as indicators rather than determinative. What is even more crucial is that the organisation is clear and agreed about its own mission, vision and values, and that these are the criteria driving all key organisational decisions, choices and prioritising (including budgeting). You may not be able to say how far towards the horizon you have swum, but you sure know you are not heading back to the shore.

- *Give more emphasis to Theories of Change and Logic Models.* As noted above, being clear on 'what difference we want to make' and developing a Theory of Change or

Logic Model for how the organisation understands its day-to-day activities can reasonably be expected to work towards those desired 'ends' (contribute to that success) is one of the most useful activities an organisation can undertake. It also helps:

- determine how reasonable the assumptions underlying their work are
- identify intermediate outcomes that may be more readily monitored within a reasonable time-frame
- justify their work to other stakeholders, including potential funders or supporters
- and potentially most significantly, identify which ways of implementing programme(s) are more likely to lead to success, and make on-going improvements.

- *Explore and develop your own learning systems*. Also as noted above, developing one's own approach to learning systems and being transparent about that is one of the most effective ways of encouraging real learning. Not only is it more likely to lead to real lessons, but they are more likely to be implemented, if they come from our own systems and our own (collective) reflection. This is also energising and motivating for the people involved. Imposed systems just handicap learning, and encourage defensiveness.

- *Keep a double set of books*. If funders or others still insist on imposing inappropriate Outcomes and measures on an organisation, it is important to be careful, as organisations tend (even when they don't mean to) to do more of whatever is measured and reported on. Some time ago, Bernstein (1991) found in a small but important ethnographic study of non-profits that (as well as having a relentless organise-wide focus on their mission, vision and values) those most successful in coping with a contracting regime, not only monitored and reported what funders wanted but also kept their own (duplicate) record-keeping system for monitoring and adjusting what they knew was important for achievement of their own mission, visions and values - what was also relevant and of value to them. Most organisations stop when they have fulfilled any external requirements, especially when these are inappropriate, but the disciplined, learning organisation will continue and still develop the systems it needs for its own effective on-going development.

References:

Aspen Institute, 1996. *Measuring Community Capacity Building,* Rural Economic Development Program, Washington, DC.

Austin, R. 1996. *Measuring and Managing Performance in Organisations*, Dorset House, New York.

Bernstein, S. 1991. *Managing Contracted Services in the Nonprofit Agency: Administrative , Ethical and Practical Issues*, Temple University Press, Philadelphia.

Boston, J., Martin, J., Pallot, J. & Walsh, P. 1996, *Public Management: The New Zealand Model*, Oxford University Press, Auckland.

Buckmaster, N. 1999. "Associations Between Outcome Measurement, Accountability, and Learning for Non-Profit Organisations. *The International Journal of Public Sector Management* 12: 186-197.

Butcher, J. 2015. "New Zealand's Relationship Accord: A case study in the politics of cross-sector rapprochement". *Cosmopolitan Civil Societies Journal* 2015, 7(2).

Campbell, D. 2002. "Outcome Assessment and the Paradox of Nonprofit Accountability", *Nonprofit Management and Leadership,* Vol.12, Issue 3:243-259.

Christian, R., Quinn, L., Andrei, K., & Pope, E. 2014. *The Reality of Measuring Human Service Programmes: Results of a Survey*, IdealWare, Portland, ME.

Cohen, D. & Zelnick, J. 2015. "What We Learned from the Failure of the Rikers Island Social Impact Bond", *Nonprofit Quarterly*, Aug 7 2015.

Collis, B., Lacey, M., O'Hagan, S., Shah, R., Wainwright, S., & Wilding, K. 2003. *Measuring Impact: Case Studies of Impact Assessment in Small and Medium-Sized Voluntary Organisations*, NCVO, London.

Considine, M., O'Sullivan, S., & Nguyen, P. 2014. "Mission Drift? The Third Sector and the Pressure to be Bussinesslike: Evidence from Job Services Australia", *Third Sector Review,* Vol 20, No1: 87-107.

Davidson, C. 2015. "That didn't go too well, did it?", *The Press* (Christchurch), 10 August 2015:A13.

Department of Social Welfare. 1996. *Strategic Directions: Post-Election Briefing Paper (1996)*, Department of Social Welfare, Wellington.

Ebrahim, A. 2003. *NGOs and Organizational Change: Discourse, Reporting and Learning*, Cambridge University Press, Cambridge.

Ebrahim, A. 2005. "Accountability Myopia: Losing Sight of Organisational Learning", *Nonprofit and Voluntary Sector Quarterly,* Vol 34, No.1: 56-87.

Ebrahim, A. & Rangan, V. K. 2010. *The Limits of Nonprofit Impact: A Contingency Framework for Measuring Social Performance* (Working Paper 10-099), Harvard Business School, Baltimore.

Edwards, M. & Hulme, D. (Eds.) 1995. *Non-Governmental Organisations - Performance and Accountability: Beyond the Magic Bullet*, Earthscan, London.

Elkin, R. & Molitor, W. 1984. *Management Indicators in Nonprofit Organisations: Guidelines to Selection and Implementation*, University of Maryland School of Social Work & Community Planning, Baltimore, MD.

Forbes, D. 1997. "Measuring the Unmeasurable: Empirical Studies of Nonprofit Organisational Effectiveness from 1977 to 1997", *Nonprofit & Voluntary Sector Quarterly*, Vol 27, No.2:183-202.

Friedman, M. 2004. *Trying Hard is Not Good Enough: How to Produce a Measurable Improvement for Customers and Communities*, Trafford Publishing, Victoria, BC.

Frumkin, P. 2001. *Balancing Public Accountability and Nonprofit Autonomy: Milestone Contracting in Oklahoma,* (Working Paper No.6), The Hauser Centre for Nonprofit Organisations, The Kennedy School of Government, Harvard University, Baltimore.

Grant Thornton. 2014 *Doing Good and Doing it Well? Grant Thornton Australia and New Zealand Not for Profit Sector Survey 2013/14*, Grant Thornton New Zealand, Wellington.

Hatry, H., Morley, E., and Vinson, E. 2001. *Outcome Measurement in Nonprofit Organisations: Current Practices and Recommendations,* Independent Sector, Washington, DC.

Herman, R., & Renz, D. 1998. "Multiple Constituencies and the Social Construction of Nonprofit Organisational Effectiveness", *Nonprofit & Voluntary Sector Quarterly*, Vol 26, No.2:185-206.

Hehenberger, L. Harling, A., & Scholten, P. 2013. *A Practical Guide to Measuring and Managing Impact*, European Venture Philanthropy Association, Brussels.

Hendricks, M., Plantz, M & Pritchard, K. 2008. "Measuring Outcomes of United Way-funded Programmes: Expectations and Reality", Carman, J & Fredericks, K (Eds.), *Nonprofits and Evaluation: New Directions for Evaluation*, No.119: 13-35.

Houlbrook, M. 2011. "Critical Perspectives on Results-Based Accountability: Practice Tensions in Small Community-Based Organisations", *Third Sector Review*, Vol 17, No.2: 45-68.

James, O. 2000. "Regulation Inside Government: Public Interest Justifications and Regulatory Failures", *Public Administration*, Vol 78, No.2: 327-343.

Jeram, J. & Wilkinson, B. 2015. *Investing for Success: Social Impact Bonds and the Future of Public Services*, The New Zealand Initiative, Wellington.

Keeves, L., Treleaven, L., Sykes, C., & Darcy, M. 2012. "Made to Measure: Taming Practices with Results-Based Accountability", *Organisational Studies,* Vol 33, No.3: 97-120 (Jan. 2012).

Lyons, M. & Nowland-Foreman, G. 2009. "Civil Society and Social Capital in Australia and New Zealand" Anheier, H. & Toepletr, S. (eds). *International Encyclopedia of Civil Society,* Springer Reference, New York, NY.

Madan, R. 2007. *Demystifying Outcome Measurement in Community Development*, NeighborWorks America, Washington, DC.

Manhire, T. 2015. "When words aren't enough: Minister's statements well and good but seiousness of prison claims require harder line", *New Zealand Herald* (Aukland), 24 July 2015:A8.

McGregor, D. 1960. *The Human Side of Enterprise*, McGrawHill, New York.

McKinlay, P. 2000. "The New Zealand Reforms – They Worked in Theory: What about the Practice?" Presentation for McKinlay Douglas, Wellington (Nov 2000).

Moir, J. & Downes, S. 2015. "Parata backs tougher pass-rates measure", *The Press* (Christchurch), 6 July 2015: A4.

Muir, K & Bennett, S. 2014. *The Compass: Your Guide to Social Impact Measurement,* The Centre for Social Impact, Sydney.

National Resource Centre. 2010. *Measuring Outcomes*. (Strengthening Nonprofits: A Capacity Builder's Resource Library No.9), Department of Health and Human Services, Washington, DC.

Neilson, B., with Sedgwick, C & Grey, S. 2015. *Outcomes Plus: The Added Value Provided by Community Social Services*, New Zealand Council of Christian Social Services, Wellington.

Nowland-Foreman, G. 1995. *Neither Mendicants Nor Deal-Makers: Contracting Government Funding and Voluntary Organisations*, New Zealand Council of Christian Social Services, Wellington.

Nowland-Foreman, G. 1998. "Purchase-of-Service Contracting, Voluntary Organisations and Civil Society: Dissecting the Goose that Lays the Golden Eggs?" *American Behavioural Scientist*, Vol.42, No.1, Sept 1998.

Nowland-Foreman, G. 2000. "Report on the Social Audit Pilot, 1998/99", COMMACT, Auckland.

Nowland-Foreman, G. 2009. "The Challenge of Accountability and the Opportunity of Responsibility for Third Sector Organisations", Paper to 6th International Society for Third Sector Research Asia Pacific Regional Conference, Taipei (Nov 2009).

O'Brien, M., Sanders, J. & Tennant, M. 2009. *The New Zealand Non-Profit Sector and Government Policy*, Office for the Community & Voluntary Sector, Wellington.

Osbourne, D. & Gaebler, T. 1992. *Reinventing Government: How the Entrepreneurial Spirit Is Transforming the Public Sector*, Addison-Wesley, Boston, MA

Paton, R. 2003. *Managing and Measuring Social Enterprises*, Sage, London.

Plantz, M., Greenway, M. & Hendricks, M. 1997. "Outcome Measurement: Showing Results in the Nonprofit Sector", *New Directions for Evaluation*, No.75:15-30 (June 1997).

Pomeroy, A. 2007. "Changing the Culture of Contracting: Funding for Outcomes", *Social Policy Journal of New Zealand*: Issue 31, (July 2007).

Power, M. 1994. *The Audit Explosion*. Demos, London

Putman, R. 1993. *Making democracy work: civic traditions in modern Italy*. Princeton University Press, Princeton NJ.

Rivas, D., Quinn, L., & Bernard, C. 2015. *Technologies and Practices for Managing Outcomes: Lessons from Large Nonprofits*, Idealware, Portland, ME.

Rudd, T., Nicoletti, E., Misner, K. & Donsu, J. 2013. *Financing Promising Evidence-Based Programs: Early Lessons from the New York City Social Impact Bond*. MDRC, New York.

Schick, A. 1996. *The Spirit of Reform: Managing the New Zealand State Sector in a Time of Change*, New Zealand State Services Commission and Treasury, Wellington.

Schick, A. 2001. "Reflections on the New Zealand Model" Lecture delivered at the New Zealand Treasury, Wellington (August 2001).

Simpson, S. & Dorling, D. 1999. "Conclusion: Statistics and 'the Truth'" (Chapt 47), Dorling, D., & Simpson, S. (Eds.) *Statistics in Society: The Arithmetic of Politics*, Arnold, London.

Statistics New Zealand. 2007. *Non-Profit Institutions Satellite Account: 2003-2005*, Statistics New Zealand, Wellington.

Stolte, O. 2004. "The Measurement of Training Opportunities Course Outcomes: An Effective Policy Tool?", *Social Policy Journal of New Zealand*, Issue 21 (March 2004).

Tan, S., Fraser, A., Giacomantonio, C., Kruithof, K., Sim, M., Lagarde, M., Disley, E., Rubin, J., & Mays, N. 2015. *An Evaluation of Social Impact Bonds in Health and Social Care Interim Report*, (PIRU Publication 2015–12), Policy Innovation Research Unit, London.

Tassie, B., Murray, V., Cutt, J., & Braggs, D. 1996. "Rationality and Politics: What Really Goes On When Funders Evaluate the Performance of Fundees", *Nonprofit & Voluntary Sector Quarterly*, Vol 25, No.3:347-63.

Torjman, S. 1999. *Are Outcomes the Best Outcome?,* Caledon Institute of Social Policy, Ottawa.

United Way of America. 1996. *Measuring Program Outcomes: A Practical Approach*, UWA, Alexandra, VA.

W. K. Kellogg Foundation Evaluation Unit. 1998. *W. K. Kellogg Foundation Evaluation Handbook*, W K Kellogg Foundation, Battle Creek, MI.