

Monitoring and evaluation

Robin Wilson

‘Evidence-based policy-making’

After New Labour came to power, the slogan ‘what matters is what works’ entered common political parlance (Davies *et al*, 2000a: 1). The more rounded claim was that policy-making should be evidence-based, though in practice this was honoured more in the breach than the observance. As Davies *et al* (2000b: 359) complain, ‘it remains a recurring disappointment that evidence so often fails to have much impact on policy making’.

There was a downside to this new thinking too. Policy is not just an issue of what works, but what works *to what end*. As a result, as Davies (2004: 5) argues, ‘Political ideology is a major driving force of policy making and is in no way made redundant by a commitment to evidence-based policy.’ Policy is never thus merely pragmatic, and policy presented as if it were often sneaks in highly problematic moral assumptions which fill the gap where explicit political values should be. Policy-making is better seen as a ‘communicative’ process, based on dialogue and argument, rather than a mere ‘technical’ product of the evidence (Sanderson, 2004: 8).

Evidence on which to base policy can, however, emerge through advocacy (Nutley and Webb, 2000: 33), including advocacy of a particular end. What was previously indulged (such as passive smoking) can over time be rendered unacceptable and the appropriate policy conclusion (a ban on smoking in public places) drawn. Thus while the segregated nature of society in Northern Ireland used to be treated as a fact of life, Boyle and Hadden (1994) showed both the extent of the hitherto neglected problem and the fact that this was a subject for policy choice rather than fatalistic acceptance. Their definition of that choice, as between separating and sharing, came to set the terms of *A Shared Future* when it was published more than a decade later (OFMDFM, 2005).

It did not, however, set the terms of the Belfast Agreement in the interim, whose ‘consociational’ design (Wilford and Wilson, 2006) was premised on the assumption that high fences make good neighbours. This may explain why

it was impossible to bring the policy to fruition under devolution, and would underscore the need for the policy to be monitored and evaluated, including independently, if devolution were to be restored on a similar basis. (There is, of course, a credible argument that it will be impossible to restore devolution in the absence of a prior commitment to *A Shared Future* on behalf of the parties.)

Evidence, though, is not just important to inform policy design. As Nutley and Webb (2000: 26) argue, rather than being linear the policy process is a cycle of activities. And so: ‘The role of evaluation in the post-decision phase provides an important source of evidence for subsequent rounds of policy analysis.’ In other words, there should be a feedback loop from the monitoring and evaluation of policy in practice to the review and any redesign of the policy itself. This is particularly important where, as is increasingly the case, the policy arena is complex and uncertain and so repeated tours around the policy cycle will be required to arrive at an effective version.

In a linear world, if an outcome B is generated by a mechanism A in a defined and stable context C, then as long as the policy lever A is pulled, the desired result B should be seen to flow and the policy should be unproblematic. So, for example, in a closed economy (C) a lower inflation rate (B) can be achieved by raising the interest rate (A) to deflate demand, and the effect is easy to detect. In today’s globalised economy this is not so simple—a symptom of the fact that one of the features of social science is that it can’t reproduce the closed systems on which experiments in the natural sciences rely. So then the question becomes one of what works *in what context* (Davies *et al*, 2000b: 355; Sanderson, 2000: 219)—and the answer may differ in each case.

This means that any policy outcome becomes the synthesis of many determinations. Suppose the desired outcome B is now the composite of a range of effects, B_{1-x} , in a range of contexts, C_{1-y} , in which a range of interventions, A_{1-z} , is operative. Achieving that outcome is then not only a more complex process but whether one will achieve it to a satisfactory degree becomes much more difficult to know. Only through careful monitoring and evaluation will it be possible to ‘tease out the effectual elements from the ineffectual in a highly contextualised manner’ (Davies *et al*, 2000b: 356)—and, moreover, ‘Many public sector interventions involve the interaction of a number of agencies, and these are often difficult to capture within a conventional evaluation methodology.’

It is important to recognise in this regard the distinction between the individuals and groups whose agency will bring about tangible change at the micro-level (including both citizens and non-governmental organisations) and the macro-environment of institutional structures and rules in which this will happen (Sanderson, 2000: 219). The latter can not substitute for the former but must be supportive for maximum effect to be achieved.

Practitioners and those with whom they work will glean tacit knowledge from their experience, which can itself offer a great deal of critical evidence (Nutley and Webb, 2000: 23), but this is likely to remain in their heads unless deliberately distilled (Wilson, 2005). Research on which I have embarked, seeking to distil ‘what works’ for reconciliation via interviews with practitioners, is already highlighting the importance of a supportive policy and political context.

From all this, on the one hand, Davies (2004a: 24) concludes: ‘Most governments require sound evidence on both the effectiveness of outcomes and the effectiveness of implementation and delivery of policies, programmes and projects.’ But, on the other, he goes on: ‘The availability of both types of evidence is often in short supply.’

It is tempting in this regard to reduce the evidence one seeks for monitoring and evaluation to *indicators* which can be readily measured. As Sanderson (2000: 216) complains, ‘concern with understanding, explanation and learning is subordinated in discussions of evaluation in official circles to issues of measurement and accountability’.

There are three difficulties with this. First, there may be ‘deadweight’ effects arising because (in any open system) it is impossible to know whether changes in the indicators would have happened regardless of the policy, and there is a problem of adequacy: the indicators may well exclude important aspects which can only be detected by qualitative methods (Sanderson, 2000: 226, 223). As the Cabinet Office recognises (Davies, 2004b: 5), qualitative research ‘can bring real depth to the understanding of the contexts in which policies operate and their implementation, processes and outcomes’.

Secondly, we are not only interested in what works but also *how and how well* it does (Sanderson, 2000: 228). We need to know if the overall ‘intervention logic’ that framed the policy is right and if the causal mechanisms identified do indeed produce particular effects in particular contexts; only then can the feedback loop come into effect. Interestingly,

Sanderson (2000: 223), taking as a negative example in this regard earlier prior and unclear policies on ‘community relations’ in Northern Ireland, contends: ‘The basis for identifying and measuring outcomes is the set of objectives specified for a programme, because these represent the desired or intended outcomes and effects which should be made explicit in the intervention logic.’

Thirdly, reducing evaluation to the collection of indicators, perceived as a technical exercise, fails to take account of the tacit knowledge of practitioners and to address the perspectival differences among stakeholders. An opportunity is thereby lost not only to enrich the process of evaluation but to address it in its properly normative context.

A Shared Future

As Williams and Thomas (2004) have said of sustainable development (in a Welsh devolved context), ‘community relations’ in Northern Ireland is a typical ‘wicked issue’. That is to say, it is a complex and non-linear policy challenge, which is perceived differently by different stakeholders, which requires systemic change rather than tinkering and for which the results may not be clear for decades. It is not amenable to a ‘one-club’ approach, it requires ‘joined-up’ government to effectuate it, and there can be no assumption of a normative consensus. The only thing we can be really sure about is that the first Triennial Action Plan will not get everything – or even most things – right and that not everybody will be satisfied.

The critical question is whether Northern Ireland is set on a trajectory towards the overall *ASF* aim, of ‘a shared society defined by a culture of tolerance’ (OFMDFM, 2005: 10), or whether it continues, willy-nilly, along the default route of ‘Balkanisation’ into two separate ‘communities’ and ‘cultures’, looking askance at each other in uneasy mistrust. In other words, are we seeing a vicious circle of the proliferation of interfaces, stereotyping and enemy images, control of dissent and occasional eruptions of violence, or a virtuous circle of easing tensions, cross-communal dialogue, a more pluralist society and the embedding of peace? The ‘intervention logic’ is that this change from vicious to virtuous circles can, albeit with great difficulty and very slowly, be secured.

It is important to recognise in this regard that segregation is not the fundamental problem. Rather, a state of segregation is one manifestation of a culture of *intolerance*, which itself is a process which will reproduce itself, through this vicious circle, unless government and non-governmental organisations intervene to generate a more powerful, opposing logic. For this reason, degrees of segregation in various spheres should only be seen as proxies in assessing *ASF*. The underlying aim should be to monitor and evaluate the outcome of the effort, under the banner of *ASF*, to shift the balance between intolerance and tolerance.

What is meant by tolerance? *ASF* does not define it but Habermas (2004: 17-18) puts it this way:

The coexistence of different life forms as equals must not be allowed to prompt segmentation. Instead, it requires the integration of all citizens – and their mutual recognition across cultural divisions as citizens – within the framework of a shared political culture. Citizens are equally empowered to develop what is for them their cultural identity and might appear to others as cultural idiosyncrasies, but only under the condition that all of them (across boundaries) understand themselves to be citizens of one and the same political community.

The aim of *ASF* becomes, in our language, the desired outcome B , and the 13 objectives identified (OFMDFM, 2005: 12-13) government hopes approximate to B_{1-x} in terms of their adequacy and completeness. The interventions to be identified in the TAP are then A_{1-z} , in the settings C_{1-y} . But it is implausible that all the elements of the outcome, B_{1-x} , or all the causal mechanisms, A_{1-z} , have been, in each case, correctly or, *in toto*, comprehensively adumbrated first time out.

So what we need to monitor and evaluate is the degree to which the objectives have been realised and whether this does indeed add up to progressive achievement of the aim. If they are not being realised, then the programmes and projects itemised in the TAP to achieve them need to be reviewed, and if the aim is not being achieved then the objectives are in some cases wrong or are together incomplete. And we need to ensure that the process of evaluation is one in which practitioners and stakeholders play an integral part.

Evaluating the policy

Set in this context, it is clear the section of *A Shared Future* on monitoring and evaluation (OFMDFM, 2005: 75-8) is inadequate. First, it does indeed reduce the evidence to a set of indicators to be generated. Secondly, it contains no link back to the objectives of the policy. Thirdly, it does not engage the wider and continuing conversation that needs to take place. And thus, overall, it fails to complete the policy feedback loop.

As a result of the indicators that are being devised being asked to bear far more weight than they should, they have inevitably grown like Topsy, with more than 140 identified so far. This should in itself have been sufficient to set alarm bells ringing and pointed to the need for a more pluralistic research methodology, including qualitative aspects such as case studies and focus groups, especially when one takes account of the micro- as well as the macro-level (Wilson, 2005). As Einstein warned, what gets counted counts, and policy evaluation is a much more profound and complex process than counting beans.

The error has been compounded by the decision to generate indicators which would also embrace the Race Equality Strategy. This has led to the identification of new ‘priority outcomes’ which are the basis for the indicators but are detached from the objectives of *A Shared Future*. Only four of the 13 objectives of the latter appear among the 11 ‘priority outcomes’ to be assessed. The indicators need to be ruthlessly pruned to those which will be key benchmarks of progress.

A way out of this messy situation, without going back to the drawing board, would be to focus on the ‘culture of tolerance’ which the two strategies are seeking to realise (recognising also the egalitarian thrust of the race strategy)—and there is a basis for important synergies to be recognised and assessed. Here are ten very simple and widely understood indicators which could be employed.

They would themselves have incentivising effects on governmental behaviour: it would, for example, not make a lot of sense to turn down four integrated-school plans – as happened, of all times, during the 2006 Integrated Education week – if the Department of Education knew one of the indicators was the number of kids in integrated schools. Moreover, they allow of a recognition that the policy has both short- and long-term perspectives: they

start, for instance, from the immediate reality of murals and marches while being clear that the long-term goal is to reinvent these ‘cultural traditions’ so that they no longer carry sectarian, or even paramilitary, baggage.

The suggested indicators follow, with the agents responsible for collecting the data indicated:

1. trends in perceptions of the state of ‘community relations’ (NILTS);
2. ‘out-group’ attitudes, including *vis-à-vis* ethnic minorities (social scientists);
3. proportion of adult relationships which are mixed (NISRA);
4. number of children enrolled in integrated schools (Department of Education);
5. proportion of ‘super-output’ areas with no ethno-religious group comprising more than 80 per cent (NISRA);
6. number of sectarian interfaces, including ‘peace walls’ (NIO);
7. trends in electoral behaviour (political scientists);
8. rates of sectarian, racist or homophobic incidents (PSNI / NI Crime Survey);
9. number of national/paramilitary flags and paramilitary/sectarian murals (local authorities / NIHE); and
10. rates of controversial communal parades (Parades Commission).

The first two are attitudinal. The first refers to two annual Northern Ireland Life and Times Survey questions, as to whether respondents *believe* ‘community relations’ are better or worse than five years ago and whether they believe they will be better or worse in five years time. Illuminatingly, this ‘feel-good’ quotient had been falling in recent years, only rising again after the institutions established by the Belfast Agreement were suspended. The second refers to psychometric work on ‘out-group attitudes’ themselves (Hewstone *et*

al, 2005). This could be replicated on a bigger and more consistent scale through a large-scale survey in advance of each ‘State of the Region’ report or the funding of specific questions in the NILTS.

The next four indicators are behavioural. The third and fourth are mindful of the danger of survey respondents rehearsing what they believe to be a normatively endorsed attitude, rather than one they necessarily share, when it is of no personal consequence to them. While recognising the limited choices individuals have – we know that about half of those who would send their kids to integrated schools if they could say there isn’t one nearby – these two indicators would demonstrate the willingness of citizens, if you like, to put their money where their mouth is.

The fifth and sixth indicators would reflect the degree of intercommunal security. We know from the history of Belfast in particular that segregation has been a ratchet process, sharply enhanced during periods of polarisation, before being spontaneously – but slowly and never fully – reversed during periods of peace. Whether minorities feel secure from ‘ethnic cleansing’ is a key indicator. Meanwhile, the prevalence of ‘peace walls’ and other physical barriers – expressions of *insecurity*, as these are, which explains why they only get longer and higher, once in place – is a converse measure.

Both of these relate to the seventh, and final behavioural, indicator—electoral behaviour. Bizarrely, this is not included among the 140+ considered so far, though it is an obvious and critical measure of polarisation (Wilford and Wilson, 2006: 13). This is not a comment upon individual Northern Ireland parties, still less individual voters: it is a comment on the fact that we have, more than ever, two political *systems* rather than one, with elections becoming purely intracommunal contests where the ‘other’ might as well be invisible. It is axiomatic (as well as evident) that no power-sharing system can be sustained in such a context, and nor can any effective political leadership be given towards ‘a shared future’.

The last three are indicators of *intolerance*. As regards the eighth, there are problems about the PSNI data, as sectarian incidents in particular are only now beginning to be counted (an interesting shift in itself), following similar innovations *vis-à-vis* racism and homophobia. Moreover, the worryingly rapid rises in the latter two may well be inflated to a degree by increased readiness to report and record (which are both welcome). The Northern Ireland Crime Survey should ask similar questions in every iteration to indicate the extent of

unreported/unrecorded incidence.

The ninth is the only indicator, apart from the second, where the data are not already collected, and could usefully be allocated to local authorities as part of their ‘good relations’ plans, as well as (thinking particularly of murals) to the Housing Executive. The figure for the first element of this indicator should, of course, eventually be nil, as *ASF* specifies that as lampposts are public property, *no* flags should be flown on them. The second aspect would distinguish ‘chill factor’ murals, rather than seeking to sanitise the public sphere. Similarly, the tenth refers specifically to ‘contentious’ parades, since there is a genuine issue here of the need for tolerance, within defined limits, of manifestations of freedom of expression we may individually find repugnant – a relatively small fraction of the thousands of marches in Northern Ireland every year.

The limitations of official thinking place additional onus on the Community Relations Council, which will be tasked under *ASF* to produce every three years a ‘State of the Region’ assessment. The CRC, depending on its research budget, needs to support strategically research that will fill in the gaps in the narrative, and to co-operate via its research network with those already working in the field. The CRC can, of course, make the same use of the available data highlighted above as government. Moreover, there is a welter of baseline academic research, well summarised in the contribution by Hargie and Dickson (2003) to their edited volume, on various dimensions of the community-relations ‘problem’.

CRC should also do its own monitoring work. It would be highly desirable to fund a post which would carry out on a larger and more systematic basis the monitoring of sectarian/paramilitary/racist incidents in the media which already is done in house. Particularly following the local weeklies would be important to ensure CRC’s sense of ‘the state of the region’ is well grounded. For example, it would be important, particularly in the context of ‘good relations’ plans, to know how effectively local authorities were taking issues of power-sharing, the neutrality of venues and so on – to what degree, in other words, local civic leadership was emerging.

But commissioned research, especially using qualitative methodologies, should play a key part, too. The report could usefully contain panels deriving from case studies of reconciliation work by NGOs, as well as of statutory agencies (including councils) attempting to implement *ASF* in their own

domain. There is also a role for semi-structured interviews with key ‘stakeholders’ in painting the overall picture. And focus groups would flesh out the evidence from the survey data as to how ‘community relations’ were evolving on the ground. Finally, comparative evidence from the progress, or otherwise, of other divided societies, such as in the Balkans, would provide helpful benchmarks for Northern Ireland.

An important aspect of the CRC’s contribution should be to integrate those public policies which do not have a focus, or at least not a primary focus, on ‘community relations’, but nevertheless have the capacity for significant effects upon its state. The outworking of the Review of Public Administration – given the fear expressed by four out of the five main parties in Northern Ireland that this could lead to a ‘Balkanisation’ scenario – will be an important factor, as will, given the loose language in *ASF* about ‘community’, the Neighbourhood Renewal Strategy. Most obviously, the decision to fund ‘loyalist communities’ goes entirely against the grain of *ASF*.

All this could be put together into an attractive, accessible and authoritative publication, which the CRC should aim to publish towards the end of the three-year action-plan cycle with a view to setting the agenda for the next one. It should also seek to use the report to inspire a broader public debate about where this society is going and what needs to be done to get, and keep, it on track towards realising the aim of *ASF* – stressing the key messages that this is *the* Northern Ireland problem, that living together is an integral part of modern life and that ‘a shared future’ is desirable and ours to shape.

Lastly, a wider forum is needed to bring together government, NGOs, academic experts and practitioners if progress is properly to be assessed. In any event, there is a real danger that the purely interdepartmental Good Relations Panel envisaged would not develop the key relationships with individuals and groups on the ground who are essential to taking *A Shared Future* off the page.

Here, a model is available in the voluntary sector forum, which regularly brings together voluntary and statutory representatives. It has proved an effective body, because it operates as a dialogue between equals (it is co-chaired by a voluntary and a statutory figure) rather than a mere sounding board for government. It has thus played an important role in the evolution of a critical policy – that of government support for the voluntary sector. This has steadily progressed from the work of the taskforce established (with broad

stakeholder representation) to investigate it through the subsequent Department for Social Development documents, *Investing Together* and *Positive Steps*.

Conclusion

What monitoring and evaluation should entail is a process of iteration whereby the evidence of research, the experience of practitioners and the views of stakeholders are used to renew and reform policy as it develops. This should be characterised by:

- a focus on the objectives and their realisation in diverse contexts,
- the use of qualitative as well as quantitative methodologies, and
- an iterative conversation involving researchers, practitioners and stakeholders.

The inadequacies identified in this paper can and should be rectified via the TAP. If not, then the CRC needs to be robust in doing so in its three-year ‘health’ check on the state of ‘community relations’ in Northern Ireland. The CRC can model new approaches in this regard—if it has the budget to do so.

Bibliography

Boyle, Kevin and Tom Hadden (1994), *Northern Ireland: The Choice* (London: Penguin Books)

Davies, Huw, Sandra Nutley and Peter Smith (2000a), ‘Introducing evidence-based policy and practice in public services’, in Davies, Nutley and Smith (eds), *What Works? Evidence-based Policy and Practice in Public Services* (Bristol: The Policy Press), 1-11

——— (2000b), ‘Learning from the past, prospects for the future’, in Davies, Nutley and Smith (eds), *What Works? Evidence-based Policy and Practice in Public Services* (Bristol: The Policy Press), 351-66

- Davies, Philip (2004a), 'Is evidence-based government possible?', lecture at the fourth annual Campbell Collaboration Colloquium, Washington DC, February 19th
- (2004b), *The Magenta Book: Guidance Notes for Policy Evaluation and Analysis—Chapter 8: Qualitative Research and Evaluation—How Do You Know Why (and How) Something Works?* (London: Cabinet Office, www.gsr.gov.uk/downloads/magenta_book/Chap_8_Magenta.pdf [accessed March 6th 2006])
- Habermas, Jürgen (2004), 'Religious tolerance: the pacemaker for cultural rights', *Philosophy* 79: 5-18
- Hargie, Owen and David Dickson (2003), 'Putting it all together: central themes from researching the troubles', in Hargie and Dickson (eds), *Researching the Troubles: Social Science Perspectives on the Northern Ireland Conflict* (Edinburgh: Mainstream Publishing), 289-306
- Hewstone, Myles, Ed Cairns, Alberto Voci, Stefania Paolini, Frances McLernon, Richard J Crisp, Ulrike Niens and Jean Craig (2005), 'Intergroup contact in a divided society: challenging segregation in Northern Ireland', in D Abrams, J M Marques and M A Hogg (eds), *The Social Psychology of Inclusion and Exclusion* (Philadelphia, PA: Psychology Press), 265-92
- Nutley, Sandra and Jeff Web (2000), 'Evidence and the policy process', in Davies, Huw, Sandra Nutley and Peter Smith (eds), *What Works? Evidence-based Policy and Practice in Public Services* (Bristol: The Policy Press), 13-41
- Office of the First Minister and Deputy First Minister (2005), *A Shared Future: Policy and Strategic Framework for Good Relations in Northern Ireland* (Belfast: OFMDFM, www.asharedfutureni.gov.uk)
- Sanderson, Ian (2004), 'Is it "what works" that matters? Evaluation and evidence-based policy making', Scott Policy Seminar, January 27th (Belfast: Economic Research Institute of Northern Ireland, www.erini.ac.uk/sem_uploads/SPSandersonPaper.pdf [accessed March 5th 2006])

- (2000), 'Evaluating initiatives to address social exclusion', in Janie Percy-Smith (ed), *Policy Responses to Social Exclusion: Towards Inclusion?* (Buckingham: Open University Press), 216-239
- Williams, Paul and Alan Thomas (2004), 'Sustainable development in Wales: understanding effective governance' (York: Joseph Rowntree Foundation, www.jrf.org.uk/bookshop/eBooks/1859352162.pdf [accessed May 24th 2005])
- Wilford, Rick and Robin Wilson (2006), *The Trouble with Northern Ireland: The Belfast Agreement and Democratic Governance* (Dublin: TASC / New Island)
- Wilson, Robin (2005), 'Football for All: a baseline survey', research commissioned by the Irish Football Association (Belfast: Democratic Dialogue, www.democraticdialogue.org/documents/FootballForAlldoc_000.pdf [accessed March 6th 2006])