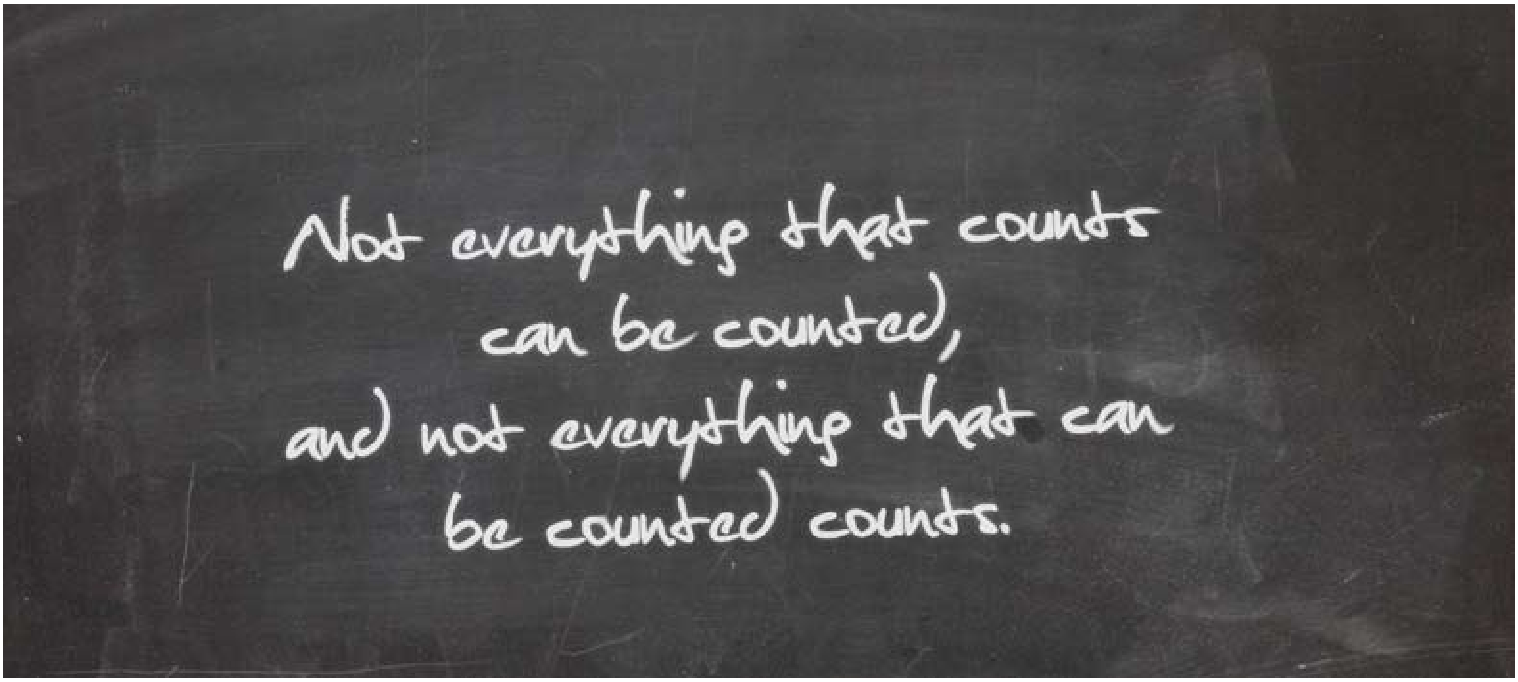
VIEWPOINTSERIES

ISSUE 3:DECEMBER 2008

MAKING THE MOST OF EVALUATION



BY JONATHAN COOK, SCOTT DICKINSON, DANIEL HEUMAN AND GEOFF WHITE

SQWconsulting

MAKING THE MOST OF EVALUATION

FOREWORD

Welcome to the third Viewpoint from the SQW Group. Each Viewpoint draws on our consulting experience to offer observations on a current public policy issue.

SQW Consulting, part of the SQW Group, has over 20 years of experience of evaluating government interventions in the UK for a wide range of national departments, devolved administrations and regional agencies.

This Viewpoint considers the lessons from policy evaluations and the importance of making more of them to improve public sector resource allocation and use. It considers:

- The motivation for evaluation
- Understanding what evaluations do
- Evaluation challenges
- Getting more out of evaluation

We hope you find this Viewpoint useful and welcome your feedback.

Chris Green
Chief Executive Officer, SQW Group

December 2008

VIEWPOINT SERIES - FEEDBACK AND MAILING LIST

To send us your views on this publication, please contact Geoff White, Director, SQW Consulting
E: gwhite@sqw.co.uk

To sign up for further publications in the Viewpoint series, please contact Karen Laurence, SQW Consulting
E: viewpoint@sqw.co.uk

Printed on paper from a sustainable source by a waterless printer

EXECUTIVE SUMMARY

Evaluations should inform policy decisions by providing evidence on the outcomes of previous interventions and their value for money. But often evaluation reports just sit on shelves gathering dust. What can be done to improve their policy contribution? This Viewpoint addresses this question by drawing on the body of evaluation evidence that has accumulated over the past twenty years.

Official guidance – such as the Green and Magenta Books – urge that evaluations should be an integral part of the policy formation and delivery cycle. We agree. Our experience is that good evaluations can help make good policy – as was the case with regional and business support policies. However we have also found that:

- The guidance can be interpreted in a mechanical way - to tick the box in an audit form to satisfy performance management requirements - or evaluations can be commissioned to defend political decisions
- The policy lessons from evaluations may be ignored because they are carried out and presented in a variety of ways that makes it difficult to compare the value for money of different interventions even where the objectives are the same.

Evaluators and their commissioners must confront some serious challenges if evaluations are going to inform policy development. We think this really matters just now as the world faces up to the big policy choices and consequences linked to the simultaneous impacts of the credit crunch and climate change. Whether this is achieved depends not only on whether evaluators are up to the challenge, but also on whether public sector commissioners are prepared to commit the resources to, and engage in, the process of evaluation as an integral part of the iterative policy cycle.

Current evaluation frameworks and methods need to be developed further if they are to help efficient allocation and use of public sector resource (such as accounting for the varying persistence of benefits and assessing value for money). There also needs to be increased investment in evaluation capacity within government departments, devolved administrations, regional agencies and local government. And a more harmonised approach to conducting evaluations and presenting their results is required if they are to become more comparable and begin to constitute an accessible repository of shared evaluation evidence.

THE MOTIVATION FOR EVALUATION

Evaluations are commissioned for a number of reasons and these often determine the nature of the commission, the budget available and the use made of the evaluation evidence. Broadly, two types of evaluation can be distinguished. Public sector commissioners may wish to learn about how well the project or a programme they sponsor is being delivered while it is being delivered. In the jargon, these are formative or process evaluations. In other cases, they may want to look at the achievements of the project or programme activities against its objectives. These are summative or impact evaluations. Both can assist public bodies by helping to improve programme or project design/delivery and contributing to future decisions about the allocation and use of public resources.

There are, of course, other motivations for commissioning evaluations – the desire to expose a competing view to public criticism via a scrutiny process or to comply with the accounting requirements of the programme or project. The former motivation often means evaluations are used primarily to provide political ammunition and/or to get media headlines. The audit motivation often produces evaluations that are ‘tick box exercises’ which, having served their bureaucratic purpose, are then put to one side.

These motivations are short-sighted and one-dimensional. They do not make the most of what good evaluations can offer – insights into effective policy formulation, resource allocation and programme/project implementation. Commissioners of evaluations and evaluators themselves need to know the full potential of evaluations – what they can do – but also what the challenges are.

UNDERSTANDING WHAT EVALUATIONS DO

Evaluation is about assessing what can be significant in terms of policy impact but cannot simply be monitored, counted or even directly observed. It recognises the truth of the aphorism attributed to Einstein that “everything that can be counted does not necessarily count; everything that counts cannot necessarily be counted”. Evaluation is about considering the wider and longer-term effects on the behaviour and performance of those directly and indirectly involved in a publicly-funded intervention. At its simplest, for example, it is about admitting the possibility that assisting one business may have adverse effects on other, non-assisted businesses and finding the evidence that suggests whether the net effect is a benefit or a cost.

“Not everything that counts can be counted, and not everything that can be counted counts”

There are two current and much more significant policy issues where considering these wider, longer term and systemic effects is, and will continue to be, absolutely essential – what Johann Hari of *The Independent* called the “two great meltdowns – the credit crunch, and the climate crunch” (October 2008). The nature of these effects was revealed by Lords King and Stern.

- The Governor of the Bank of England expressed his concern about the moral hazard that might be associated with policy responses to the first effects of the credit crunch. “In essence”, he said in his evidence a year ago to the Treasury Select Committee (September 2007), “this moral hazard argument is that, should the central bank act and effectively provide extra liquidity at different maturities against weaker collateral, markets would, especially if the liquidity were provided at little or no penalty, take it as a signal that the central bank would always rescue them should

"Policy makers and evaluators need to face up to the challenge of understanding the longer-term and wider effects of their past policies and those they may be planning"

they take excessive risk and get into difficulties". As it turned out, there was an even bigger set of potential wider effects that was to put this moral hazard concern into the shade – namely, the possible, hugely adverse externality for the economy as a whole arising from information, other market and regulatory failures in its financial sector – all at a time of looming recession.

- The Stern report on the Economics of Climate Change (October 2006) was clear that "climate change presents a unique challenge for economics: it is the greatest and widest-ranging market failure ever seen. The economic analysis must therefore be global, deal with long time horizons, have the economics of risk and uncertainty at centre stage, and examine the possibility of major, non-marginal change". It concluded that acting now to stabilise carbon emissions would be costly – maybe as much as 5% of GDP – but that this paled against the costs of a business as usual case where 20% could be slashed off consumption per head.

This is the very stuff of appraisal (what might happen) and evaluation (what did happen). You might think evaluation would benefit from hindsight but it still requires an assessment of 'what might have happened' and 'what didn't happen as expected'. And this needs investigation of things that cannot easily be observed. Unlike film-makers of the natural world who can use infra-red to witness what is invisible to the naked eye, evaluators do not have that kind of technology despite the developments in economic modelling methods. Instead, they have to work with theory and partial evidence suggested by surveys, case studies and anecdotes, memories and records of policy stories and experience. They then have to draw conclusions through the prism of common sense and judgement.

It is perhaps this that makes policy makers and their officials uneasy about the results of evaluation. It is certainly what makes good evaluations controversial. They seek to lay bare what can be said on the basis of evidence and ring-fence where theory, assumptions and judgement have to be exercised. Policy makers may be right to be wary about evaluation given the current state of the art but they – and evaluators – cannot deny the need to face up to the challenge of understanding the longer-term and wider effects of their past policies and those they may be planning.

We have identified the key evaluation challenges by drawing on our own evaluations and reviews of other evaluations. Many of these are of relatively small-scale programmes and projects – nothing like the blockbuster interventions proposed to address the climate and credit crunches. Nevertheless, we think that most of the challenges are relevant to larger-scale evaluations.

EVALUATION CHALLENGES

Embedding evaluations in the policy cycle

Despite the official guidance in the 2003 Treasury Green Book[1] and the Government Social Research Unit's Magenta book,[2] it is our experience that evaluation has not yet been robustly embedded in the policy cycle. Three reasons might explain this.

- **First**, the demand for evaluation has so far largely been focused on the spending departments and their agencies. It has been notably absent from fiscal and regulatory interventions. Evaluations have thus been seen primarily as an instrument of public expenditure control and, as such, treated by their commissioners with due caution if not out-right resistance.
- **Second**, evaluation is often regarded as part of a linear process of accountability – bolted on at the end of a programme or project – rather than something to be used to reshape and refine thinking around key policy issues. This is reinforced where there is a mismatch between the timing of the policy cycle and the evaluation process.
- **Third**, there is very little by way of systematic recording and comparing of evaluation results to provide the basis for benchmarking or development of best practice. Thus, to date, evaluations have not provided evidence in a form that can readily help with policy decisions.

Using logic models

The role of evaluation is to test the robustness of the rationale for an intervention. This is best done by use of a logic model which provides a framework for describing the theory, assumptions and evidence underlying an intervention and "links outcomes (both short and long term) with programme activities/processes and the theoretical assumptions/principles of the programme."[3]

Various formulations are available for such logic models. **Figure 1** has been designed by SQW to combine the logic-model structure used by the Innovation Network[4] with a basic performance management framework that runs from strategic appraisal to impact assessment and learning.

From the evidence of our evaluations and evaluation reviews, it seems that policies and programmes often have poorly articulated rationales with little attention given to expectations with respect to outcomes and impacts. In other words, big chunks of the logic model are often missing or only faintly outlined. This may be because the interventions were responses to rapidly changing threats or opportunities. It may have seemed better for the public authorities to respond quickly and only as thoughtfully as circumstances allowed.

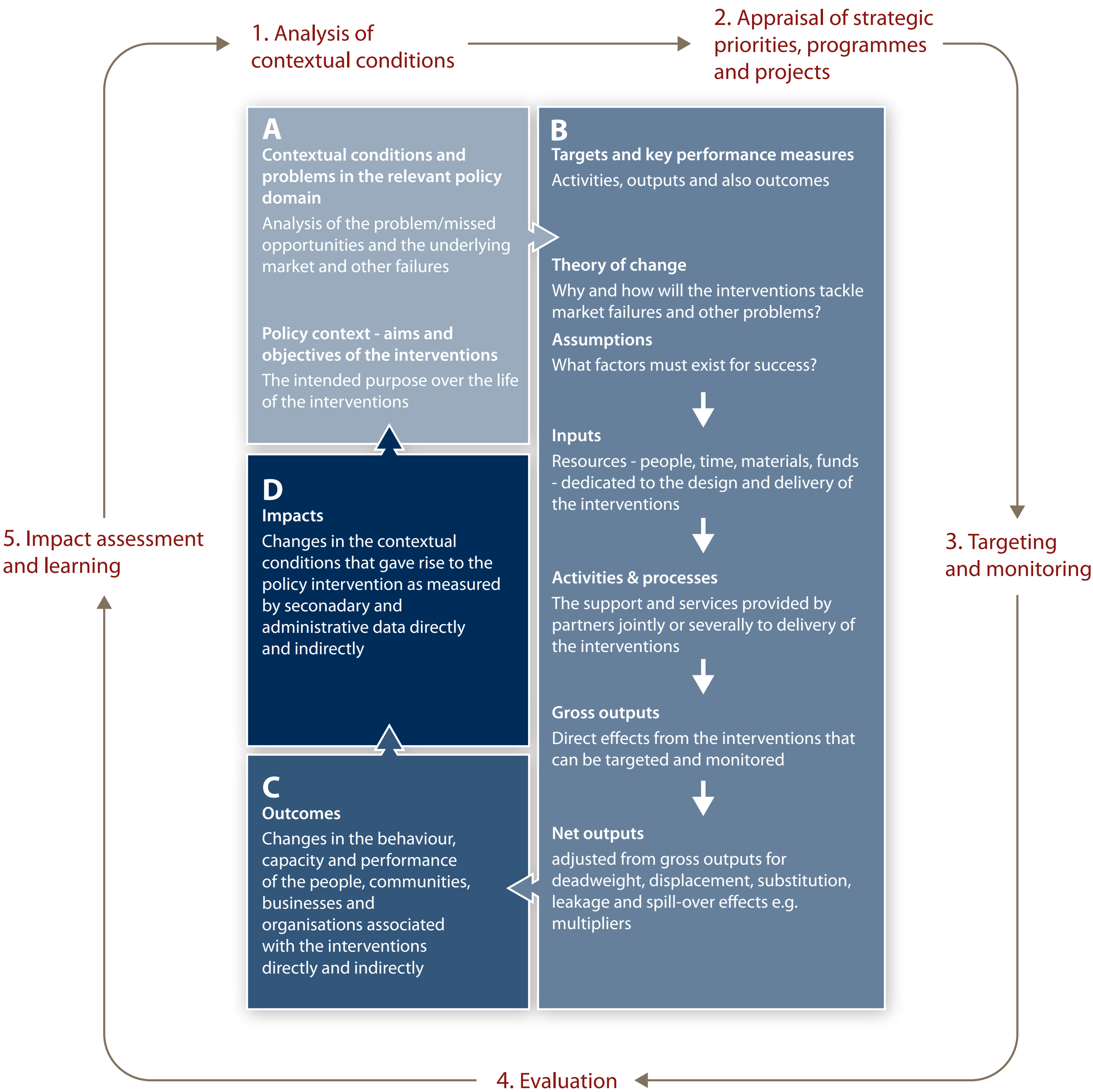
"Policies and programmes often have poorly articulated rationales with little attention given to expectations with respect to outcomes and impacts"

In such cases, the evaluation will involve a reconstruction of the rationale based on intentions stated at the time and within the context defined by prevailing data and evidence. Not only do we recommend the use of logic models for this purpose but we also suggest that policy makers might consider using and populating them when policies and programmes are first designed and appraised.

Improving monitoring data and attribution of outputs

Evaluations can be diverted from their purpose of assessing wider and longer term effects by having to make good the inadequacies of monitoring data. This has happened to such an extent in our

Figure 1: Logic model and performance measurement framework for public sector interventions



Source: SQW Consulting

experience that it suggests that improvements are needed in the design and use of monitoring and management information systems. Often the problem is to do with 'garbage in, garbage out' – reflecting a lack of incentive, an unwillingness and/or inability of project/programme managers to provide accurate monitoring data and to update and upgrade it as the intervention unfolds.

A specific area for improvement is in the attribution of outputs within monitoring systems and procedures. By this we mean the method by which outputs are allocated between two or more public sector agencies who have contributed funding to a project or programme. This will become an increasingly important issue as policy drivers continue to prompt joined-up and partnership working, yet the ground-rules by which this allocation should be done are not clearly set out in guidance. A variety of pragmatic solutions tends to be adopted – often with lack of transparency as to how the attribution was done or how it was up-dated in different and changing circumstances. We suggest the basic operating principle should be that, whatever attribution method is used, total project/ programme outputs should always be recorded within monitoring systems so that it is clear what proportion of the total has been attributed to the agency in question.

Assessing wider effects

Our review of the evaluation evidence on business and trade support demonstrates that many of the evaluations have not provided assessments of net outputs or outcomes (see the top chart in **Figure 2**).

"Impacts on contextual conditions need to become a standard feature of evaluations – linking intervention outputs and outcomes to relevant secondary data sets"

Instead, they were used to substitute for the absence or inadequacies of monitoring data and to estimate deadweight (i.e. the extent to which beneficiaries would have behaved in exactly the same way without the intervention).

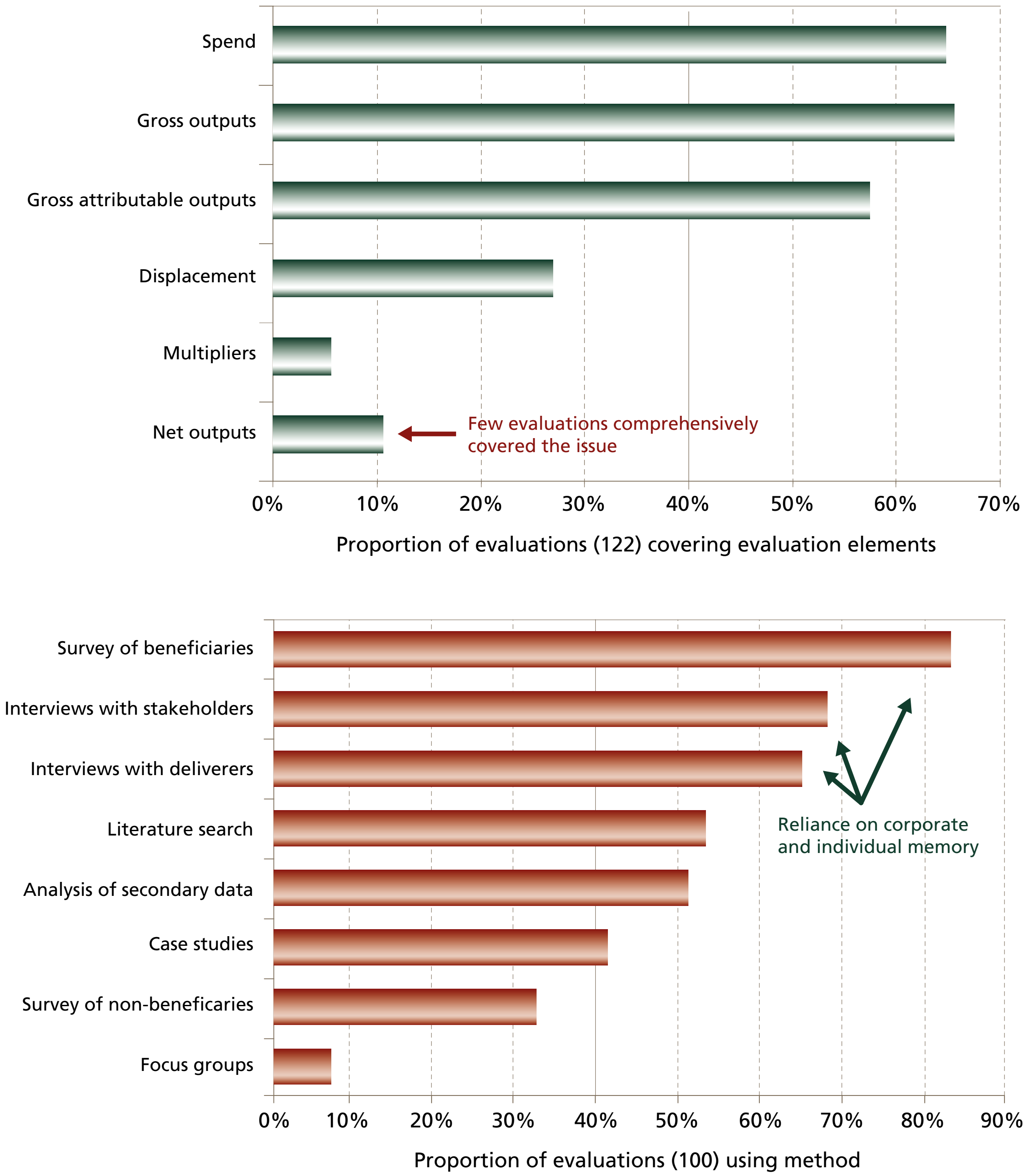
Given the importance that should be attached to the assessment of wider and longer-term effects, we think that very much more emphasis needs to be given in evaluations to:

- **Displacement effects** (e.g. where an assisted firm or person benefits at the expense of someone else) which are critically dependent on the size of the spatial area under consideration and the conditions in the relevant product and factor markets
- **Multiplier and spillover effects** where more work is needed to understand the circumstances in which and the mechanisms by which multiplier, spillover and other wider dynamic effects come about and can be estimated
- **Impacts on contextual conditions** need to become a standard feature of evaluations – linking intervention outputs and outcomes to relevant secondary data sets – at least to give a sense of the relative scale of the outputs.

Improving evaluation methods

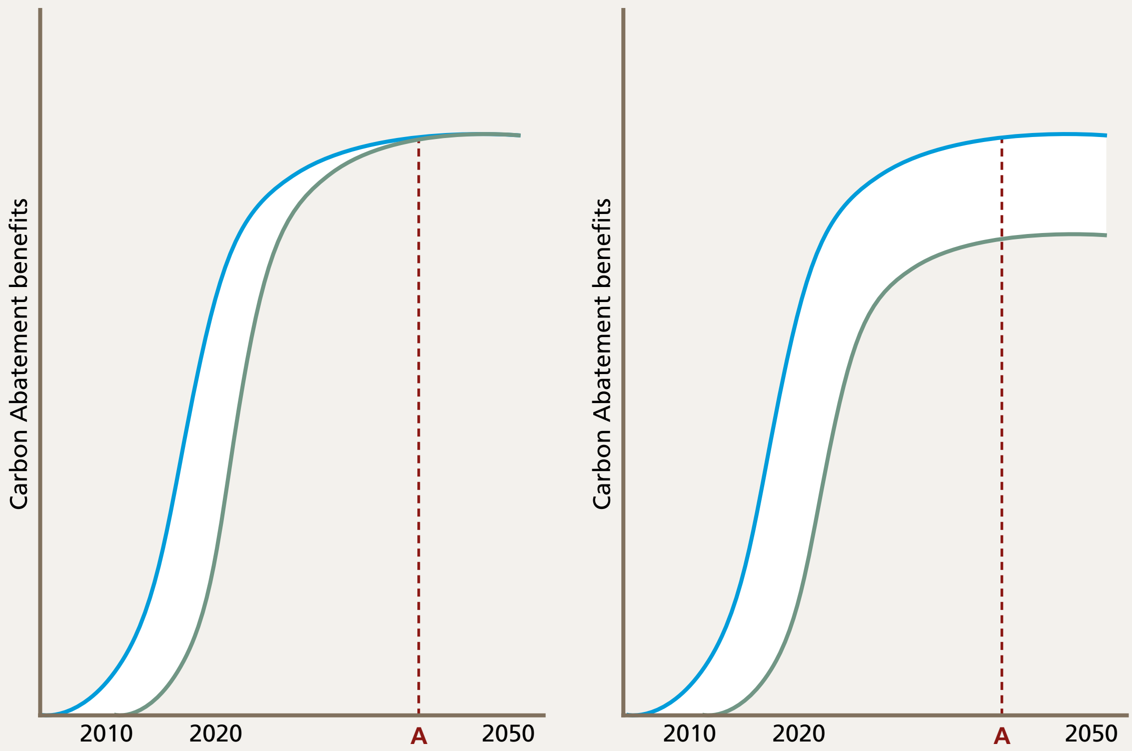
The lower chart in **Figure 2** shows the frequency with which surveys of beneficiaries and stakeholders were used in the evaluations we reviewed to inform evaluation judgements. Such surveys rely on corporate and individual memory within beneficiary and third party organisations. This is not always available where there has been a degree of organisational restructuring and staff turnover. Moreover, these approaches are limited with respect to the insights that they can provide on displacement and multiplier effects.

Figure 2: Evaluation methods and coverage (SQW 2006, 2007)



Source: SQW Consulting, Reviews for the Enterprise Directorate and UKTI in the Department for Business, Enterprise and Regulatory Reform (2007)

Figure 3: Persistence of carbon abatement benefits from low-carbon technology support on different assumptions



Assumptions

1. The end of the technical or average service life of the asset supported by the intervention is at point A.
2. At this point, technological advances mean that the replacement investment offers the same benefits whether or not the intervention had taken place.
3. Therefore, the carbon abatement benefits attributable to the intervention are represented by the difference between the S curves with and without the intervention.

Assumptions

1. The intervention caused some market or production transformation effects - e.g. by prompting a technological change that would not have otherwise occurred.
2. The carbon abatement benefits therefore persist beyond point A and perhaps in perpetuity on an assumption of a radical transformation in the production possibility frontier.
3. This outcome assumes no catch-up - no convergence of the two S curves.

Source: SQW Consulting

"Despite the importance of persistence effects, it is rare for evaluations to take them into account or even discuss them. The consequence is that interventions with very different persistence effects end up being treated much the same. And that makes no sense at all."

Improvements in evaluation methods are required to address these issues. Greater reliance needs to be placed on the use of secondary data analysis, of control sampling of non-beneficiaries, and of longitudinal surveys to get a more reliable fix on what might have happened without the intervention and its wider consequences. These methods are more expensive but do tend to be more reliable and reduce the extent to which conclusions have to be based on the evaluators' judgements.

Accounting for persistence effects

Benefits from interventions last for different periods of time – some benefits occur only while the intervention is in place, others last well beyond its completion. Critical to the assessment of persistence effects are not only the length of life of the assets at which the intervention is directed (e.g. commercial office space or a piece of technology) but also the extent to which the intervention transformed market and production possibilities.

This is illustrated in **Figure 3** with regard to the evaluation of the persistence of the benefits from the support of low-carbon technologies. Central to the extent of persistence in this case is whether the support brought about transformations in technological and production possibilities that just would not otherwise have happened at all or not for a very long time.

Despite the importance of persistence effects, it is rare for evaluations to take them into account or even discuss them. The consequence is that interventions with very different persistence effects end up being treated much the same. And that makes no sense at all.

Evaluating value for money

Public sector organisations are rightly concerned with making value for money judgements about their interventions and comparing them to assist resource allocation decisions. Assessing value for money involves consideration of its three components – economy (the unit costs of delivery), effectiveness (achieved outputs and outcomes against objectives and targets) and efficiency (cost-effectiveness or cost-benefit assessments).

The degree to which value for money varies between interventions is often difficult for public sector organisations to judge because of problems with cost-effectiveness measures (i.e. outputs per unit of public sector spend). Comparing the cost-effectiveness of interventions is not a problem where they generate the same type of output and where the contexts in which they operate are similar. However, where interventions produce different mixes of outputs, comparison of the 'apples and pears' becomes much more problematic. Other variations between interventions, such as the degree of intensity of the assistance, can cause simple cost per output measures to vary considerably (e.g. within business support) and therefore be potentially misleading for resource allocation purposes when comparing even broadly similar interventions.

Whilst value for money assessments need to escape from the cul de sac of cost-effectiveness assessments, some would say that the alternative of cost-benefit analysis (CBA) is like riding a helter-skelter – much more exciting but gets you no further forward. They would be wrong.

Certainly, it is our experience that few attempts have been made to use CBA in the policy area of economic development. But, there is no technical reason for it being any more difficult to apply CBA here as compared with any other policy domain. And it might help to enhance the contribution of evaluations to improved priority ordering and resource allocation by offering comparable cost-benefit ratios. So, we recommend it should be given a go. This will require improved understanding in assigning or imputing monetary values to benefits (such as jobs or attained skills), estimating their persistence and discounting them over time (rarely carried out in evaluations) and in addressing the limitations of single benefit-cost ratios.

GETTING MORE OUT OF EVALUATIONS

If policy makers and their delivery agencies are to get more out of evaluations, and if evaluations are to become an integral part of the policy development/delivery learning cycle, at least five things need to happen.

Greater investment in evaluation capacity

There has to be more investment by policy makers to incorporate evaluation in the policy cycle. Evaluation has tended to be the Cinderella in the iterative sequence that runs from policy appraisal, implementation, monitoring, feedback and learning. Currently, in general, it is not sufficiently integrated into the design of policy and, consequently, there is still limited capacity for effective commissioning and application of evaluation findings.

Bigger commitment to impact assessment

Policy makers need to be more ambitious in the demands they place on the evaluations they commission, but pragmatic in their specifications. That means they must, at the same time as being challenging, also acknowledge that evaluation is not a counting exercise. It requires assumptions and estimation methods to carry out impact and value for money assessments using cost benefit analyses. Policy makers need to be prepared to get more involved in making the judgements that are required.

Clearer guidance on evaluation methods

Existing evaluation guidance is not clear enough on the available methods for estimating outcomes, assessing impacts and reaching conclusions on value for money. Nor does it provide good practice on these matters. Some central government departments are beginning to review the methods used and results generated by their policy evaluations.[5] It would make sense if the key outcomes of these reviews were collated and synthesised centrally, rather than only produced by the departments concerned.[6]

"Some would say that the alternative of cost-benefit analysis is like riding a helter-skelter – much more exciting but gets you no further forward. They would be wrong"

"Policy makers need to use evaluation to inform policy making and delivery rather than as an 'end-of-pipe' counting exercise"

Simpler and shorter evaluation reports

Evaluators are often their own worst enemy. Their reports are often seen as 'other worldly', too technical and not sufficiently geared to forward decision-making. Complex issues may have to be tackled in carrying out a good evaluation and a lot of technical work may need to be completed. But, this doesn't have to be on display in the show-room window which ought to be reserved for the key evaluation findings, implications and recommendations. Presentational innovations might also be helpful to make the results of evaluations more accessible (e.g. in the use of logic models and balanced scorecards).

Better harmonisation across policy domains

Policy makers need to use evaluation to inform policy making and delivery rather than as an 'end-of-pipe' counting exercise. To this end, there is a need to learn more about what works in evaluation across different policy domains through the development of evaluation networks. This will become particularly important where policy encourages greater delegation and devolution to more local levels where clear guidance on evaluation good practice and benchmarks will be required.

A more harmonised approach to the conduct of evaluations and presentation of their results would enable more robust comparisons of their findings and could provide for the development of an easily accessed repository of evaluation evidence as the basis for methodological advancement in the future.

REFERENCES

1 http://greenbook.treasury.gov.uk/

2 http://www.nationalschool.gov.uk/policyhub/magenta_book/ - '...effective policy making must be a learning process which involves finding out from experience what works and what does not and making sure that others can learn from it too. This means that new policies must have evaluation of their effectiveness built into them from the start ...'

3 WK Kellogg Foundation, Logic Model Development Guide (January 2004).

4 Innovation Network, Logic Model Workbook (2005) https://www.innonet.org/client_docs/File/logic_model_workbook.pdf

5 Greenberg D and Knight G for DWP, Working Paper No 40, Review of the DWP Cost Benefit Framework and how it has been applied (2007), and SQW Consulting for BERR, Review of evaluation evidence on business development support, (forthcoming).

6 A DWP (2007) review made recommendations about the treatment of multipliers and general equilibrium effects which are likely to have as much relevance for BERR and the Department for Innovation, Universities and Skills (DIUS). This, and the results of other work (e.g. by BERR on estimating additionality), could form further supplementary guidance to the Green Book to bring consistency to the treatment of impact assessment and cost benefit analysis in different policy areas.

About us

SQW Consulting, SQW Energy, Oxford Innovation and SQW Asia are part of SQW Group.
For more information: www.sqwgroup.com

SQW Consulting is a leading independent consultancy in the area of sustainable economic and social development. The company works with government departments in the United Kingdom and overseas, devolved administrations, regional development agencies, sector skills councils, funding bodies, local authorities, partnerships and international agencies. We provide advice on economic development and competitiveness, environment and sustainable development, higher education, innovation and knowledge transfer, regeneration and neighbourhood renewal, public services reform, science and technology, skills and employment, social inclusion, spatial development, and young people and children.
For more information: www.sqw.co.uk

SQW Energy has a specialist focus on energy and carbon management and works with the energy industry, public and private sector organizations and NGOs, providing tailored consultancy services in energy market design and regulation, energy technology strategy and policy, and carbon assessment and management.
For more information: www.sqwenergy.com

Oxford Innovation is the UK's leading operator of Innovation Centres, providing office, laboratory and workshop premises to over 350 innovative start-up companies in 14 Centres. Oxford Innovation also manages the most successful technology investment networks in Europe with members including business angels, investment funds and corporate venturers. The company provides further innovation support services to achieve economic development objectives for public sector clients.
For more information: www.oxin.co.uk

SQW Asia is headquartered in Hong Kong and provides economic and management consultancy services, including: policy analysis and review; market, feasibility and planning studies for economic and physical development and investment; and advice to support project implementation. Services are provided for Government organizations in Hong Kong and China as well as foreign firms investing in China and Chinese organisations investing overseas.
For more information, contact Chris Green. E: cgreen@sqwgroup.com

For more information about this Viewpoint, please contact

Geoff White, Director, SQW Consulting
T: +44 (0)207 307 7154 E: gwhite@sqw.co.uk

To sign up for further publications in the Viewpoint Series, please contact

Karen Laurence, SQW Consulting
E: viewpoint@sqw.co.uk

THE VIEWPOINT SERIES

The Viewpoint Series is a series of 'thought piece' publications produced by SQW Consulting, SQW Energy, Oxford Innovation and SQW Asia, the operating divisions of SQW Group.

The aim of the Viewpoint Series is to share our thoughts on key topical issues in the arena of sustainable economic development, public policy, innovation and enterprise with our clients and others with an interest in the particular subject area of each paper. In each Viewpoint, we will draw on our policy research and implementation experience to consider key topical issues, and provide suggestions for strategic and practical solutions.

More copies of this publication are available at: http://www.sqw.co.uk/publications

Printed on paper from a sustainable source by a waterless printer