
3 Institutional and process issues

For evidence and evaluation to contribute materially to the selection of policies, it must be supported by institutional frameworks that incorporate evidence into the decision making process and encourage, disseminate and defend good evaluation (Banks 2009). The institutional framework should also ensure that the resources allocated to the evaluation are commensurate with the benefits it produces.

The experience of the United Kingdom is informative. Its evidence-based policy movement, which began in 1997, has seen greater investment in research, in the evaluation capabilities of the public service, and in the use of pilots and randomised policy trials (Mulgan 2003). Yet, over twelve years later, a House of Commons report on health initiatives suggests there is still some way to go to embed evidence in policy development:

The most damning criticisms of Government policies we have heard in this inquiry have not been of the policies themselves, but rather of the Government's approach to designing and introducing new policies which make meaningful evaluation impossible. Even where evaluation is carried out, it is usually "soft", amounting to little more than examining processes and asking those involved what they thought about them. All too often Governments rush in with insufficient thought, do not collect adequate data at the beginning about the health of the population which will be affected by the policies, do not have clear objectives, make numerous changes to the policies and its objectives and do not maintain the policy long enough to know whether it has worked. As a result, in the words of one witness, 'we have wasted huge opportunities to learn'. (2009, p. 5)

Table 3.1 sets out some of the features of institutions and processes that may assist in incorporating evidence into policymaking.

Table 3.1 A matrix of suggested principles for evidence-based policy
Institutional and process issues

<i>Suggested principle</i>	<i>Why?</i>
1. Design the most appropriate evaluation arrangements.	There are trade-offs between independent, ad hoc review, 'in-house' review, or a standing external review process.
2. Maximise transparency — make data and evaluation public and provide for peer review and public consultation.	Provides quality assurance, improves credibility, aids government accountability, and facilitates improved evaluation over time.
3. Establish a monitoring & evaluation program, including resourcing, at policy commencement.	Ensures data and evidence are available for evaluation. Will address the paucity of data in some policy areas.
4. Consider sequential policy roll-out, pilot studies, or randomised trials as appropriate.	Useful for policymaking under uncertainty, where there is little settled evidence or where costs of failure are high.
5. Disseminate evaluation and pool results across jurisdictions.	Improves evaluation practices and increases links between researchers and government. Assists in translating vast amounts of 'research' into policy 'evidence'.
6. Ensure evidence is linked to the decision making process.	Provides an incentive for, and discipline on, government agencies to provide rigorous evidence to support policy proposals.

Principle 1: Design the most appropriate evaluation arrangements

Evaluation is likely to be more effective if there is some level of detachment, or independence, between the evaluator and the subject being evaluated. Options for providing evaluation are wide ranging and can vary from departmental (internal) evaluation units, evaluation institutions (such as the Auditor General), parliamentary committees, independent review panels, private consultancies and academics. The choice of evaluator will depend on the purposes of the evaluation task (for example, internal organisational learning, or broader policy review) and what might be at stake.

Internal evaluators have the advantage of intimate knowledge of the policy and easy access to information (both formal and informal knowledge). Working with those who designed and implemented programs, internal review can be assimilated into the organisation's operations and can disseminate improved approaches to staff in a non-threatening way. However, it can be difficult for in-house evaluation to achieve the desirable independence of perspective, where broader issues of policy are involved or if the issue cuts across agencies. External evaluators can overcome this problem, having the advantage of independence, but they may have less understanding of operational detail of the policy framework and may need long lead

times to master a new topic and marshal evidence through consultation and research.

The international Evaluation Cooperation Group¹ has drawn on the good practice standards of official audit and evaluation agencies that span across government and the corporate sector, to identify four dimensions of evaluation independence:

- *Organisational independence* ensures that the evaluation unit and its staff are not under the control or influence of decision-makers who have responsibility for the activities being evaluated and that they have full access to the information they need to fulfil their mandate.
- *Behavioural independence* enables the evaluation unit to set its work program, produce high quality and uncompromising reports and to disclose its findings without management-imposed restrictions.
- *Protection from outside interference* keeps the evaluation function free to set its priorities, design its processes and products, reach its judgments and administer its human and budget resources without intrusion by management.
- *Conflict of interest safeguards* guarantee that current, immediate future or prior professional and personal relationships and considerations are not allowed to influence evaluators' judgments or create the appearance of a lack of objectivity (Picciotto 2008, p 6).

An example of how institutions can design evaluation units to combine the advantages of internal insight to benefit organisational learning for the institution, with independence of analysis comes from the World Bank Group. Its Independent Evaluation Group is headed by a Director-General of vice-presidential rank, but instead of reporting like other Vice-Presidents to the President of the Bank, the Director-General reports directly to the Board. The position has only administrative links with the President. The Director-General is appointed by recommendation of the Board for a fixed five year term. While the occupant may come from the ranks of Bank management, the contractual appointment is not renewable and the occupant may not return into Bank employment — provisions which, together with the other arrangements noted, reduce the risk of conflicts of interest and bolsters behavioural and organisational independence (World Bank 2003).

¹ The Evaluation Cooperation Group is composed of the heads of evaluation of the multilateral development banks and observers such as the OECD's Development Assistance Committee's Evaluation Network head and the Director of Independent Evaluation at the International Monetary Fund.

Some suggest that for evaluation to be successful, it has to be performed in several places, which serves to cross-check outcomes. There are a number of examples in Australia of such arrangements, with one of the more prominent operating under national competition policy (box 3.1).

Overall, evidence-based policy cannot rest on any one evaluator or institution. Evaluation should be decentralised and embedded in the way every policy agency works, but with the evaluator chosen depending on the policy review task at hand. The OECD (1999) recommends:

- internal evaluation when the main objectives are organisational learning and improved process;
- external evaluation by researchers and consultants when seeking new perspectives on policy or where there is a need for specialist evaluation skills;
- independent evaluation when assessing the effectiveness and efficiency of a policy.

Box 3.1 National Competition Policy: A hybrid model

Under the national competition policy (NCP), State and Territory governments were responsible for implementing NCP reforms, with the National Competition Council (NCC) acting as an independent monitor and assessor of performance. For example, State and Territory governments were responsible for reviewing legislation to assess whether regulatory restrictions on competition were in the public interest and the NCC would assess the rigour of the review and report to the Australian Government on compliance. State and Territory performance was linked to competition payments.

This accountability framework established pressures, via inter-jurisdictional demonstration effects, to maintain a commitment to reform over time and to adhere to agreed review processes.

As a by-product of the assessment/monitoring process, the collection and dissemination of information on reform processes and outcomes promoted 'learning by doing' experiences across jurisdictions. It has also aided the process of fine-tuning implementation processes — such as the development of better guidelines for the legislation review process — as well as identifying problem areas in need of follow-up work to progress reform satisfactorily, such as for road and water reform.

Source: PC (2005, p. 144).

Principle 2: Maximise transparency

Transparency can help raise public awareness of policy problems, lead to better-informed analysis of policy options and build support for reform. In a sense, transparency is a ‘safety net’ for evidence-based policy — a form of quality control that provides opportunities for correction or refinement when the evidence is not complete. It can elicit new sources of information and alternative analysis, expose weaknesses in prevailing analysis, and shed light on how the positions of sectoral interests relate to overall community impact, thereby helping achieve better policies and outcomes.

Transparency is often thought of in terms of public access to relevant information, such as freedom of information laws. But ideally, transparency goes further and provides for contestable policy review with:

- the opportunity for stakeholders to comment;
- access to data, assumptions and methods; and
- peer review and accountability provisions that, for example, require policymakers to publish their decisions and the evidence which supports them.

Transparency in this broader sense creates an important opportunity to replicate results and test alternative assumptions. This is particularly important in areas of complex economic modelling.

For decision makers, however, transparency can be a double-edged sword. Even effective policies and programs can have flaws that present communication challenges for government. And unfettered transparency is simply not possible in certain areas like defence or when commercial-in-confidence information is supplied to government. It is not surprising, therefore, that many jurisdictions struggle to balance the need for transparency with the time and risk associated with transparent policy development and evaluation. Some international initiatives aimed at increasing transparency are outlined in box 3.2. An historical Australian example is provided in box 3.3.

Box 3.2 Transparency: Some international examples

United Kingdom

The UK Government has established a central website where all impact assessments (similar to RISs) are publicly available. This enables full access to the evidence base used to justify the need to regulate, including details of the options that were considered and discarded as well as assessment of the impacts on equity and the environment.

New Zealand

New Zealand allows access to the full text of Cabinet documents and Cabinet decisions, unless the benefits of doing so are outweighed by a public interest in keeping the information confidential. (In Australia, Cabinet information is automatically confidential and only released under the '30 year rule').

United States

The Office of Management and Budget, together with federal agencies, are required to assess every budget program and publicly report the results, demonstrating if programs are successful or if they fall short. To date, 1015 programs have been assessed, finding 193 effective, 326 moderately effective, 297 adequate, 26 ineffective and 173 programs did not demonstrate results.

International

In 2005, the World Bank commissioned an independent, public evaluation of its research activities between 1998 and 2005. Chaired by Princeton economist Angus Deaton, the evaluators were asked to examine Bank research paying particular attention to its reliability, rigour, completeness and relevance and, in addition, to give their overall assessment of the strengths and weaknesses of Bank research.

Box 3.3 Transparency: an Australian example

The Wallis Inquiry into financial sector trends and regulatory reform: 1996 - 1997

The Wallis Inquiry was set up on 30 May 1996 to examine the consequences and implications of the deregulation of the Australian financial system flowing from the Campbell Report of 1981. It still stands as a creditable example of how special-purpose, one-off inquiries can be set up according to the best principles of transparency, even when dealing rapidly with issues of great sensitivity (the prudential supervision of the financial system) and important conflicting commercial interests (the respective roles of banks, insurance companies and the like). The Inquiry was tasked to recommend future regulation in the light of likely financial market developments, and with a view to best serving the interests of consumers, while preserving system stability and fairness.

The Inquiry was comprised of a panel of three senior businesspeople with experience in both the financial sector and the broader economy, and two highly-regarded professors of finance. It was served by a Secretariat led by a senior Commonwealth Treasury official.

The Inquiry published an issues paper to elicit discussion and submissions, and received 268 submissions from financial institutions, academics, State governments, the Reserve Bank of Australia, the Australian Competition and Consumer Commission and private citizens. (All major institutions' submissions were public.)

It also undertook a program of public consultations for which transcripts were made available, further enhancing the transparency with which evidence was shared with the public and interested analysts. Inquiry members tested ideas through further public engagement via 10 public addresses. The Inquiry was one of the first in Australia to use the Internet to disseminate submissions and its work.

In April 1997, the Government received the final Wallis report, and took immediate decisions to somewhat liberalise the restriction that prevented mergers among the six major banks and insurance companies, but affirmed that it would not allow mergers among the four major banks. By September 1997, the Government had completed its consideration of the Inquiry's recommendations, and set the regulatory architecture for the Australian financial system that prevails today: the Reserve Bank of Australia with macro-prudential responsibility for overall financial and payment system stability; the Australian Prudential Regulation Authority to supervise financial institutions regardless of their business specialisations; and the Australian Securities and Investment Commission to supervise market integrity, disclosure and consumer protections issues.

A dozen pieces of legislation were introduced or amended in 1998 to give effect to the new regulatory architecture. The Wallis Inquiry foresaw and successfully addressed the problems of regulatory fragmentation, gaps and overlaps that have been an element of the financial sector crisis in the United States.

Source: <http://fsi.treasury.gov.au/content/default.asp>

Genuine transparency provides opportunities for adequate input and review. But experience suggests that consultation sometimes occurs only after a policy decision has been made. Edwards (2001) recount of the development of the Child Support Scheme in the mid-1980s is a characteristic example:

Having made in-principle decisions by the middle of 1986, ministers did not want these decisions undone. They therefore endorsed a discussion paper which assumed the major planks of the scheme as effectively ‘non-negotiable’ and sought views on ‘second-order, but complex’ issues (2001, p. 77).

Greater upfront consultation through the use of traditional Westminster approaches to broad new areas of policy or major policy review, such as ‘Green Papers’ to scope approaches, and ‘White Papers’ to propose intended Government policy can significantly improve policy development. Recent Australian examples include Green Papers on national aviation policy, financial services and credit reform and the Carbon Pollution Reduction Scheme.

The Productivity Commission and its predecessors have evolved a range of tested working processes that enhance transparency and improve the evidence base in policy consideration. Some of these processes are enshrined in the *Productivity Commission Act 1998*:

- Elicit evidence on key points for deliberation by way of a public issues paper.
- Proceed by seeking submissions and conducting public hearings or roundtables to gather evidence.
- Make all evidence and submissions publicly available and contestable, unless there are powerful reasons to the contrary.
- When using quantitative modelling, use several models if they are in reputable use, or use modelling roundtables to ensure key model characteristics and parameters are chosen in the light of professional best practice.
- Engage external experts as consultants where necessary, or as referees for research.
- Test tentative conclusions and recommendations by publishing a draft report for comment and review.
- Publish the final report for debate and government deliberation.

Principle 3: Establish a monitoring & evaluation program, including resourcing, at policy commencement.

The options for evaluation are restricted, and sometimes severely limited, if there has been no prior attention to monitoring, evaluating and reporting on a policy over

time. Establishing a systematic monitoring and evaluation program for significant policies at commencement, including committing resources, can play a major part in tackling the evaluation problem when the time comes for review (see box 3.4). This is because the ability to measure whether there has been progress (or not) depends on taking an upfront decision on what to measure, gathering baseline data (the measurement of initial conditions) and setting up a system to monitor results:

Government should seek to implement policies in a way that facilitates policy evaluation; it should invest in data collection for program evaluation; it should be willing to release that data externally; and it should support research by funding external researchers and sponsoring in-house research for public release. (Borland et al. 2005, p. 114)

Box 3.4 Evaluating the family law reforms

Evaluating family law has in the past been patchy, piecemeal and complicated by some very difficult counterfactual questions — such as whether the family law system helps maintain relationships with children and whether the services are effective in reducing conflict and helping parents to agree on child-focused, post-separation parenting arrangements.

In July 2006, the Australian Government introduced wide-ranging reforms to the family law system, including amendments to the Family Law Act recognising shared parental responsibility and providing alternative dispute resolution mechanisms, and integrated and new services, such as Family Relationship Centres. The reform package cost almost \$400 million over four years.

Prior to the reforms, the Attorney-General's Department and the Department of Families, Housing, Community Services and Indigenous Affairs commissioned the Australian Institute of Family Studies (a statutory authority) to undertake an evaluation of the family law reforms over the next ten years.

The evaluation comprises three studies:

- the families project which will assess how families are faring under the reforms
- the service provision project will assess the extent to which the new system of service delivery is meeting the objectives of reform
- the legislation and courts project will assess the implementation of legislative changes that govern post-separation parenting arrangements and processes in the family law courts.

The evaluation will build on baseline research gathered pre-reform, including a general population of parents survey (5000 families), a caring for children after parental separation survey (500 families) and a survey of family law service providers (400 participants). Other sources of data being collected are program data (throughput, service use etc), service provider feedback and a longitudinal survey of separated parents to improve the understanding of the long-term effects of family law policy.

Source: AIFS (2007).

Almost every existing policy evaluation report stresses the need to obtain better data to improve the quality of evaluation. Insufficient data is particularly a problem for regulatory, social and environmental policy, and where it is collected, administering departments can fail to collect the data most useful for evaluation, or be reluctant to release it for external analysis. For instance, the lack of baseline data for the COAG Indigenous trials meant it was very difficult for subsequent evaluations to assess the impact or outcomes of the trials. And in the field of education policy, much research on student outcomes in Australia has had to rely on the OECD's PISA, a test administered to a sample of 15 year-old students in OECD and non-OECD countries. The use of PISA reflected, at least until recently, a lack of publicly reported, timely data on student outcomes and characteristics in Australia, rather than any qualitative advantages that PISA offers (Leigh and Thompson 2008, p. 65).

One approach to encourage better monitoring and data collection is to place a requirement on agencies proposing any kind of significant policy to identify how it will be monitored and assessed, such as in the RIS process. Built-in requirements for review, like sunset clauses, have similar objectives. However, their effectiveness is sometimes questionable since there is no sanction if agencies fail to monitor the performance of the policy, or monitor it only perfunctorily.

There are also international examples of tying a certain proportion of program funds to monitoring and evaluation, with public reporting obligations to government. For example, the United States *Second Chance Act 2008*, contains a provision to set aside two per cent of program funds for evaluation of programs that facilitate prisoner re-entry into the community. Evaluations must 'include, to the maximum extent feasible, random assignment ... and generate evidence on which re-entry approaches and strategies are most effective'. This type of hypothecation can encourage rigorous evaluation, but does not guarantee it.

The United States Office of Management and Budget has recently outlined a program to build rigorous evidence to drive policy evolution (box 3.5). The program links funding under certain programs to building rigorous data and evaluation processes into the programs from the outset. In other experiments, it allocates funding in some policy areas to programs with proven results, while giving initial funding to other programs with weaker evidence of success, but on condition of generating more robust evidence which will determine their funding in future budgets.

Box 3.5 The Office of Management and Budget's encouragement of evidence-based policy

President Obama's director of the OMB, Peter Orszag, has outlined a program to place more emphasis on evidence-based policy decisions. The program seeks to:

- Design new initiatives to build rigorous data about what works, and then act on evidence that emerges — expanding the approaches that work best, fine-tuning the ones that get mixed results, and shutting down those that are failing.
- Invest more in program evaluation, with early attention to education, labour, community service and youth programs.
- Design 'new initiatives with evaluation built into their DNA', such as the Department of Education's 'Grow what works and innovation fund', and programs of home visitation (to provide support to vulnerable parents) and the prevention of teen pregnancy.

These last initiatives use a two-tier approach, with one tier providing more funds to programs with strong evidence that they deliver good results, and the second, more experimental tier conditionally funding promising approaches with some evidence of success, but with future funding conditional on rigorous evaluation and strong results:

Organisations will know that to be considered for funding, they must provide credible evaluation results that show promise, and be ready to subject their models to analysis. As more models move into the top tier, it will create pressure on all the top-tier models to improve their effectiveness, so they continue to receive support.

Source: Orszag (2009).

An incentives-based approach is developing in Australia under the COAG Reform Agenda, whereby Australian Government funding to the States and Territories is tied to monitoring, evaluation and reporting obligations.

These policy reforms are to be implemented in the context of National Agreements/National Partnerships, containing the objectives, outcomes, outputs and performance indicators for policy reform, and clarifying the roles and responsibilities that will guide governments in the delivery of services across the relevant sectors. The performance of all governments in achieving mutually-agreed outcomes and benchmarks is to be monitored and assessed by the independent COAG Reform Council and reported publicly on an annual basis.

Principle 4: Consider sequential roll-out, pilot studies, or randomised trials as appropriate.

In Australia, most policies and programs are implemented on at least a state-wide, if not national scale. Very few policies are subject to sequential roll-out, pilot studies

or randomised trials to test what works, although there have been some exceptions (see box 3.6). Time and resource constraints, the legislative process and political pressures have been seen as obstacles to genuine policy trials, because once a policy proposal has been announced for universal application (for example, a computer for every school child) it is politically difficult to change policy content or offer a program to some but not others. Some policymakers also raise ethical and moral objections to some trials, particularly randomised controlled trials, because they withhold or delay a potentially beneficial policy from some individuals.

Box 3.6 Two Australian randomised policy trials

Between 1999 and 2001, the then Department of Family and Community Services conducted two randomised trials on the Job Network, examining the effect of interviews and follow-up contact from professional staff on workforce participation by the long-term unemployed. They found that the intervention led to a reduction in the number of hours worked but an increase in the number of hours spent in studying or training (Farrelly 2008)

In 1998, the NSW Government established a policy trial to test the concept of a Drug Court. During the previous decade, several countries had experimented with drug courts, but evidence on the effectiveness of these courts was limited. The NSW Drug Court trial studied non-violent offenders who met a series of criteria, including dependence on illicit drugs, and willingness to plead guilty. Participants were then randomly assigned either to the Drug Court, or to a regular court. Over the two years in which the trial was in operation, 514 people participated in the trial. The evaluation found that individuals processed through the Drug Court had lower rates of recidivism for drug offences than those processed through the normal criminal justice system. While the cost of the Drug Court exceeded the cost of regular courts, the cost per offence averted was substantially lower under the Drug Court (Leigh 2003).

Leigh (2003) responds to some of the common objections to policy trials. He notes that governments never provide assistance to all those who would benefit from it. The objection that those in need will miss out is ameliorated by the fact that researchers, while they may believe the intervention will be positive, genuinely do not know whether it is preferable to be assigned to the treatment or control group — otherwise they would not conduct the trial. He cites a review of medical trials showing that the treatment outperformed the control only about half the time and quotes Chalmers (1968), ‘One has only to review the graveyard of discarded therapies to discover how many patients might have benefited from being randomly assigned to a control group’ (Leigh 2003, p. 343).

Trials will not always be suitable (see section 2, box 2.8), particularly if an immediate response is important or where the costs are low, where it may be better to rely on post-implementation review. But, Australia is in little danger of ‘overusing’ trials, only having undertaken a handful of significant policy trials to date.

Greater use of policy trials would enable policy effectiveness to be tested prior to full scale implementation. Trials will be most valuable when the evidence base is thin, but even when the policy prescription is generally well-accepted, trials can be useful to explore the practicality of implementing policy in a particular ways, and discovering what delivery methods work best or are most cost-effective.

Trials can also act as insurance policies, helping to protect governments and taxpayers from expensive failures, and preventing good policies from being discarded because of implementation problems. They can also help build support for unproven policies.

For example, giving universal high school student access to laptop computers gives rise to complex issues of IT support structures and curriculum evolution to gain a benefit from enhanced access.² Resolving these issues takes time. If the policy were conceived as a series of trials that used natural differences in state support and curriculum approaches, it might be possible to initiate the program at least as promptly, learn from continuous feedback from well-designed evaluation, and achieve total coverage at much the same time, with better educational outcomes.

It is commonly thought that well designed roll-out and evaluation of complex programs is a daunting task, with large costs and complex institutional designs. Such views mean trials are often reserved for small, incremental projects rather than fundamental policy redesign. But these views are not supported by the experience of Mexico in undertaking sweeping, progressively evaluated welfare reform in the late 1990s (box 3.7). Since the initial pilot program, there has been a systemic and sustained effort to monitor and evaluate the Progres-Oportunidades program using sequential rollout. Moreover, the program’s continuous, publicly demonstrated success through evaluation gave the program widespread popular support, and helped insure it against partisan political attack in subsequent election campaigns and a change of government.

² Clotfelter et al. (2008) provide some analysis as to why the impact of computer use on educational outcomes is not straightforward.

Box 3.7 Mexico's Progres-Oportunidades program

In March 1995, Finance Ministry officials proposed to then President Ernesto Zedillo and his Cabinet that the government replace in-kind distributions of milk and tortillas, and price subsidies for bread and tortillas with targeted cash transfers for mothers, contingent on household members' regular attendance at health clinics. (A proposal to expand transfers and condition them on school attendance was added in 1996.) The objective was to have a single comprehensive program to transfer income to all poor households, whether urban or rural. At the time, that was a radical proposition in developing countries' welfare design, and it was met with scepticism by some Cabinet members. Three questions arose: Would the substitution of cash for in-kind transfers lead to less spending on food and more on goods like cigarettes and alcohol? Would giving cash to mothers lead to family disruption and potentially to family violence? Was making cash transfers contingent on compliance with requirements operationally feasible in a relatively poor country with weaker administrative capacities than many other OECD economies?

To answer those questions, the Finance Ministry implemented a pilot project in the state of Campeche involving 31,000 households, and commissioned an external evaluation. On one hand, the evaluation allowed the officials to reject the idea that use of cash transfers would lead to inappropriate use of the funds and to family disruption. A large majority of households preferred cash to in-kind subsidies and actually valued the link to health services; furthermore, cash transfers did not diminish households' food consumption. On the other hand, the evaluation pointed out that before program scale-up could be contemplated, targeting and selection methods had to be substantially revised and more and better data collected. In addition, the pilot project operated on an *ad hoc* basis, hiring health personnel under temporary contracts and working in premises that were not entirely suitable. A different administrative set-up would be needed for a large-scale effort.

The pilot project was essential for three reasons. First, it provided assurance to the government that the change being completed was not unduly risky, given its potential benefits. Second, it made clear that none of the existing government agencies had any incentive to coordinate a much larger-scale operation, forcing program designers to deal explicitly with that issue. And third, it established a decision making mechanism whereby the results of external evaluation, although perhaps not decisive, would nevertheless be seriously weighed in conjunction with other considerations. Learning from the pilot project and having time to adapt the program accordingly were instrumental in solving many unexpected operational issues. These factors contributed to the elaboration of a more solid proposal prior to program scale-up.

Source: Levy (2006).

Scope for greater use of trials could come from the core social policy fields of education, welfare, health and justice, where it is possible to construct control and treatment groups. In education, where the educational career of children has been referred to as '15 000 hours of compulsory treatment' (Davies et al 2000, p. 7), trials could help inform policymakers on how teacher quality and merit-based pay

affect student performance. In health, different funding models could be usefully tested. And in justice, trials could help inform the debate over the impact of different forms of incarceration and release programs on subsequent recidivism and employment patterns (Leigh 2003, p. 351).

Principle 5: Disseminate evaluation and pool results across jurisdictions

There is a significant amount of policy evaluation undertaken in Australia. Yet, a considerable amount never goes beyond the sponsoring government agency or academic institution. Other government agencies, governments and researchers are unaware or unable to access the evaluation (box 3.8). And, in cases where the evidence base is thin, lack of technical skills and resources impede the generation of robust evidence.

Box 3.8 Why isn't existing evidence used more by policymakers?

Some of the most commonly cited reasons why policymakers do not use existing research evidence include:

- research is not made available
- research is not accessible i.e. it is difficult to access ('hidden' in a morass of research databases) or not presented in an understandable way
- research does not ask policy relevant questions – academic researchers are rewarded for publishing in highly ranked (mainly theoretical) journals as opposed to engaging in policy relevant research
- research does not provide policy relevant answers i.e. did the program work? what are the costs and benefits of a policy? what level of uncertainty is associated with the estimates?
- research is inconsistent with current or proposed policy
- lack of communication (interaction) between researchers and policymakers.

Sources: Lavis et al. (2005); Innvaer et al. (2002); Edwards (2004).

Various 'club-like' arrangements have emerged to disseminate evaluation and learn from other jurisdictions. Such clubs can be organised around principles that lead members to raise the quantity and quality of data and evaluation, improve independence and transparency, and improve the dissemination of results. They also capitalise on scarce evaluation resources and exploit economies of scale.

Case studies

The OECD's Evaluation Network of the Development Assistance Committee seeks to increase the effectiveness of international development programs by supporting robust, informed and independent evaluation. The Network brings together evaluation managers and specialists from OECD members' development cooperation agencies and multilateral development institutions. It runs an online evaluation resource centre, disseminates influential evaluations, develops guidelines for evaluation good practice, and conducts peer review of evaluation functions in its multilateral member organisations.³

The Evaluation Cooperation Group is a similar club, composed of the heads of evaluation of the multilateral development banks, the Director of Independent Evaluation at the International Monetary Fund, and such observers as the head of the OECD's Development Assistance Committee's Evaluation Network. As well as developing evaluation practices among its own members, it aims to develop evaluation capacity in the borrowing country members of the multilateral development banks.⁴ Another recent development in the field of aid evaluation is the International Initiative for Impact Evaluation (box 3.9).

Moving to the realm of developed countries, the United States Coalition for Evidence-Based Policy represents a useful example in a federal setting of the propagation of funding and the principles of robust evaluation to state and central government programs. The Coalition was founded in 2001 and has a small staff sponsored by the Council for Excellence in Government. (The Council, in turn, is made up of former senior government officials, private sector companies and philanthropic foundations.)

The Coalition seeks to:

- Fund rigorous studies, particularly randomised controlled trials, to build the number of social interventions proven to produce sizeable, sustained benefits to participants and/or society.
- Provide strong incentives and assistance for federal funding recipients to adopt such proven interventions, and put them into widespread use.

³ For evaluation guidelines, see <http://www.oecd.org/dataoecd/12/56/41612905.pdf> . For examples of recent peer reviews of evaluation in various UN agencies, see http://www.oecd.org/document/57/0,3343,en_21571361_34047972_36056761_1_1_1_1,00.html.

⁴ https://wpqp1.adb.org/QuickPlace/ecg/Main.nsf/h_Toc/73ffb29010478ff348257290000f43a6/?OpenDocumnt

Box 3.9 An evaluation club: the International Initiative for Impact Evaluation

Aid evaluation has traditionally evaluated project outputs – whether a school was built – but not aid impact (education outcomes and how this affected economic and other outcomes). The aid sphere is also frequently bedevilled by complex attribution problems. An African country may have over 20 significant bilateral and multilateral donors, so the question is not just whether aid worked, but whose aid worked, and how.

The International Initiative for Impact Evaluation, ‘3ie’, recognises that evaluating development assistance involves the policies (and policy evaluations) of both aid donors and the aid recipient. (For example, whether one donor’s aid to a recipient’s education sector (say to build a school) has a successful impact on education outcomes, depends in part on other donors’ education aid to the recipient (say to train teachers), and importantly, the recipient’s own education policies (say teachers pay and curricula)).

3ie was founded in 2007 by the Centre for Global Development to channel funds to independent impact evaluations around an agenda of key questions that confront policymakers. The members are government officials from developing countries with a strong interest in issues of effectiveness and accountability (Mexico and Uganda are founding members), as well as representatives of bilateral donor agencies, multilateral agencies, non-governmental organisations, and foundations or corporations such as the Bill and Melinda Gates Foundation and Google. 3ie is also partnered with the Campbell Collaboration, the collaborative devoted to conducting and disseminating systematic reviews in the realm of social policy.

Membership entails a preparedness to commit to collaborating and dedicating significant funds to impact evaluations. 3ie funds proposals for impact evaluations that meet standards specified in the Principles for Impact Evaluation. (See: <http://www.3ieimpact.org/doc/principlesforimpactevaluation.pdf>)

The Coalition manages a website on evidence-based programs, *Social Programs that Work* (www.evidencebasedprograms.org), which provide policymakers and practitioners with concise, actionable information on ‘what works’ in social policy, based on evaluations. It also operates a ‘help desk’ for federal agencies providing practical resources to advance rigorous evaluation.

Through its advocacy, the Coalition has helped implement measures including new guidance by the Office of Management and Budget to federal agencies on *What Constitutes Strong Evidence of Program Effectiveness* (OMB, 2008) and various evaluation commitments in new legislative programs.

The Coalition has been externally evaluated as very effective, albeit lightly staffed, and as having dealt so far mainly with the OMB and the Departments of Education

and Justice, all of whom were already well-disposed to randomised controlled trials and other rigorous evaluation. The evaluation reported the Coalition's real test will be in extending and deepening its influence into less supportive policy areas.⁵

In the United Kingdom, much of the activity surrounding the evidence-based policy movement has focused on establishing a number of organisations to improve evaluation practice and disseminate results. Some focus on particular policy fields, such as the NHS Centre for Reviews and Dissemination (Health), but there are several large initiatives that cover a broad suite of policy areas (see box 3.10).

Opportunities in Australia?

Could models such as those above work in Australia? One promising opportunity to build a better machinery for evidence-based policy is the CoAG reform agenda, where complex policy reform issues affecting Commonwealth, State and Territory governments abound: education and health; early childhood education and care; mental health; education and training measures to meet the challenge of rising unemployment; and binge drinking, to name a few.

This framework of policy and financing provides the opportunity to build a collaborative process to finance and undertake evaluation, learn from any differences in state and territory approaches, and build evaluative skills. The COAG Reform Council, which will evaluate the progress of each jurisdiction on an annual basis, may fill part of this brief. The COAG Working Groups and Ministerial Councils could facilitate data collection and further evaluation, but have not done so to date.

An alternative vehicle that could build on existing arrangements is a COAG 'evaluation club', made up of senior Commonwealth, State and Territory policy officials, the Australian Statistician, and key official and academic evaluators from each jurisdiction. Through such a club, the greatest needs for evaluation could be identified, evaluation plans could be developed, and funding and evaluation feedback could be arranged. The process could be developed to integrate policy development, data collection, evaluation and feedback of lessons learned into a continuous process. The club members could share and develop technical expertise, resolve data deficiencies, ensure appropriate funding for evaluation, and provide peer review and peer support for the quality evaluation.

⁵ The external evaluation is available at http://prod.ceg.rd.net/admin/FormManager/filesuploading/indep_evaln_for_WT_Grant.pdf.

Such a club could go beyond present support groups (such as the Australasian Evaluation Society) and inject a higher level policy and resource commitment to strengthening evidence-based policy.

Box 3.10 Evaluation bodies in the United Kingdom

The *National School of Government* (a government department) promotes practical strategies for evidence based policy making through:

- the development of 'knowledge pools' to promote effective sharing of information
- training officials in how to interpret, use and apply evidence
- maintaining a 'policy hub' website providing access to a wide range of current research and guidance on the use of research and evidence in the evaluation of policy
- seconding academic researchers into government agencies to carry out research projects.

The *UK Centre for Evidence Based Policy and Practice* is an initiative funded by the Economic and Social Research Council (an independent statutory authority). The Centre, together with an associated network of university centres of excellence, aims to foster the exchange of social science research between policy, researchers and practitioners by:

- improving the accessibility, quality and usefulness of research
- developing methods of appraising and summarising research relevant to policy and practice
- informing and advising those in policy making roles, through its dissemination function.

The *Evidence for Policy and Practice Information and Coordinating Centre* conducts systematic reviews of research across a range of different topic areas and provides support for others who are undertaking systematic reviews or using research evidence. Currently it has a large number of systematic reviews in the fields of education, health promotion and public health.

Source: Nutley (2003).

Principle 6: Ensure links with the decision making process.

Evaluation can only successfully influence policy where there are incentives for government agencies to provide rigorous analysis and opportunities for decision makers to consider it. That is, evidence needs to be linked to the decision making process, rather than viewed as an adjunct to it.

Most countries have adopted a mixture of ‘carrots and sticks’ to try and integrate evidence into the policy making process, with many introducing formal requirements to evaluate budget spending programs or prepare impact statements for proposed regulation. Ultimately, however, a predetermined policy agenda can have a far greater influence than any form of requirement, and as such, evaluation requirements can only be fully effective when they are directly tied to decision making.

For instance, the Regulation Taskforce (2006, p. 156) found that although the regulatory impact assessment process should in principle provide a strong evidence base to support policymaking, the requirements were often been circumvented or treated as an afterthought in practice. In its view, the single most important way of strengthening compliance was to link evaluation to the decision making process. It recommended that regulatory proposals that fail to meet RIS requirements should not be permitted to proceed to Cabinet (unless granted an exemption by the Prime Minister).

Linking evidence with decision making will require a range of factors: the right institutional structure, adequate resourcing, consultation and transparency and, most importantly, a genuine commitment to evidence-based policy from governments and their agencies (Banks 2009). Nevertheless, there are some specific accountability provisions that may help (see box 3.11), such as the changes to the regulatory impact process mentioned above or the ministerial declaration requirement in the United Kingdom, whereby Ministers must declare that they are satisfied that a) a policy impact assessment represents a fair and reasonable view of the expected costs, benefits and impact of a policy and b) that the benefits justify the costs. Australia is also in a position to take advantage of the federal system, by using COAG arrangements to tie funding to evaluation of initiatives under the COAG reform agenda.

Box 3.11 Linking evidence to decision-making: OECD suggestions for regulatory impact analysis

The OECD identifies a range of institutional and procedural factors that can help maximise the influence of regulatory impact analysis (RIA) on decision making to improve regulatory quality outcomes:

Governments should require impact analysis to be:

- integrated into policymaking, beginning as early as possible in the process
- released for public consultation
- peer reviewed or undergo independent assessment

Governments should provide high level political support, including formal authority for RIA requirements and quality assurance mechanisms, such as assessment and approval from a central regulatory reform authority (an independent 'gatekeeper')

- In the United Kingdom, regulatory proposals that are likely to impose a significant new burden on business, must first be assessed by the Cabinet Office Better Regulation Executive, and then cleared by the Panel on Regulatory Accountability, which is chaired by the Prime Minister, before proceeding to Cabinet.

Governments should require ex-post review of impact analysis, both in terms of assessing whether the RIA met procedural requirements AND the regulatory impacts (i.e. a net benefit test). For example

- In the United Kingdom, the National Audit Office reviews a sample of impact assessments published in the previous year
- Germany requires an ex-post RIA to be completed once experience is available on the new regulation

Source: OECD (2009).