

Evaluation: a booming business but is it adding value?¹

Frans L Leeuw



Frans L Leeuw is Professor of Law, Public Policy and Social Science Research at the University of Maastricht, the Netherlands, and Director of the Netherlands Justice Research Center (WODC), Den Haag, and is also President of the Dutch Evaluation Society.
Email: <frans.leeuw@maastrichtuniversity.nl> or <fleeuw@cuci.nl>

Evaluation, monitoring, inspection and performance auditing have become a booming business. The numbers of professional societies have strongly increased and there are so many studies that authors recently referred to them as ‘streams’ and ‘fleets’. Evaluation has also been called a ‘positive social epidemic’. This article reveals that the booming business not only has positive consequences and ‘added value’, but also unwanted downsides. ‘Evaluation machines’, ‘death by evaluation’, ‘evaluitis’ and the ‘performance paradox’ are examples.

This article presents three possible explanations for the situation in which a booming business goes hand in hand with the beginnings of diminishing returns of evaluation for society. The first explanation is that evaluations are becoming part of management support systems that produce information that often only reinforces management decisions. The second looks into the unintended and negative consequences of ‘paradigm wars’. And the third suggestion deals with the limited explanatory power of evaluations.

The problem

Evaluation, monitoring, inspection and performance auditing are no longer ‘baby’ or ‘infant’ industries’. In fact, the number of studies and reports has increased so much over the last two decades that Rist & Stame (2006) gave their latest book the title: *From Studies to Streams* [of studies]. Some years earlier, Weiss (1998) used the metaphor ‘fleets of studies’. The number of jobs for evaluators has also increased (Leeuw, Toulemonde & Brouwer 1999; Preskill 2008), as has the numbers of (national) evaluation societies (Furubo, Rist & Sandahl 2002) and their memberships. Also, since 2001, membership of the American Evaluation Association, for example, has grown more than 79 per cent (from 3055 members in 2000 to 5479 by December 2007). Therefore, Preskill (2008) coined the idea of evaluation becoming a (positive) *social epidemic*. Part of this ‘epidemic’ are the ‘systems of evaluations’ that have their own organisations, handbooks, methodology, money and staff (Leeuw & Furubo 2008). Examples are: *performance auditing and monitoring*, *M(onitoring and) E(valuation)*, *experimental evaluations* and *evaluation and accreditation*.

For those believing that these booming businesses only lead to positive consequences and ‘added value’, this article brings *bad news*. There are rather unpleasant *side effects*. In his speech for the 2008 Congress of the American

Evaluation Association, Dahler-Larsen criticised the existence of systems of evaluations and called them 'evaluation machines'. A little earlier Ryan (2003, p. 7) used the expression 'death by evaluation', to describe a situation with:

too much evaluation ... [where] ... evaluation requirements ... dominate everything else, and [where] ... a compliance culture in evaluation emerged as a result.

Swiss economists Frey and Osterloh (2006; see also Frey 2006) coined the concept of 'evaluitis' to describe current developments in the field of (academic) research evaluations and point at the 'hidden costs' of these activities, at 'questionable benefits' and at 'superior alternatives'.

The proliferation of performance auditing has also not been profitable for society because perverse effects such as 'ossification', 'dramaturgical compliance' and the 'performance paradox' have been detected. Instead of making organisations more effective, the opposite is occurring, where too much auditing is 'around' (Feller 2002; Power 1999; van Thiel & Leeuw 2002).

These and similar comments lead me to conclude that the 'added value' of evaluative information is still a problem. Despite the positive aspects of evaluation work, summarised in textbooks (such as transparency, accountability and organisational learning), there also appear to be a number of *downsides* that have to be subtracted from the overall 'added value'.

Part of the problem is also that it is not clear what exactly is meant by 'added value', despite decades of debate. *Conceptual use, instrumental use, symbolic use, process use, knowledge transfer, social improvement, betterment, enlightenment and influence* have been topics on the agenda and they have continued to stay on. This is also true for the other side of the coin: *under-utilisation, misuse and abuse* (Shulha & Cousins 1997; Walters 2008). Many papers have been published describing the actual level of 'evaluation (ab)use', but the overall results are rather confusing. Sometimes it is found that 'evaluation has usually had only modest influence on policy and practice, [with] most studies ... [seeming to be] used in selective bits, reinterpreted to fit existing preferences, or ignored' (Weiss et al. 2008).² At other times, it has been found that *some* evaluations have been rather influential and have caused social improvement (Henry & Mark 2003; World Bank 2004).³ And it has been documented that evaluations, if and when used, almost never lead to the termination of policies but more often to adjustments of existing policies (Dahler-Larsen 2000). This might stimulate interventions to 'breed' further interventions, thereby leading to inefficiencies and budget maximisation by civil servants.⁴

But what is not receiving serious attention is the 'added value' of evaluations to *theory development* and *theory testing*. While the history of evaluation

shows that evaluations had their roots in disciplines such as Sociology and Psychology, nowadays these connections are less common. Vaessen & Leeuw (2009) call this the 'Mind the Gap' problem, describing the distance between explanatory theory, disciplines and evaluations.

The overall situation we face is, therefore, characterised by a *booming business in evaluation and look-alike activities on the one hand* and on the other the *danger, if not the beginnings of diminishing returns of these activities*. The question is: *How come?*

Why booming business but also diminishing returns?

I present three suggestions in response to this question. The first has to do with evaluations becoming more and more part of *management support systems*. They produce information that often only reinforces management decisions. Although one may believe that this always adds value to evaluation, in fact this is not the case. The second suggestion looks into the unintended and negative consequences of 'paradigm wars'. And the third suggestion deals with the limited *explanatory power* of evaluations. Here the argument is that policies and programs are often treated as 'black boxes' with *logic(al frameworks) or program logics* that describe the components, and then depict how they fit together (usually in a sequence of inputs, activities, outputs and outcomes). However, the theoretical, that is, *explanatory power* of this kind of 'analytical evaluation' (Scriven 1998) is limited and does not contribute much to the 'added value' of evaluations.

Evaluation as a management support system

Based on a study of the use of evaluations in Danish municipalities, Dahler-Larsen (2000, p. 85) demonstrates that some well-known metaphors of the relationship between evaluation and decision-making did not apply. The idea of evaluations as 'dusty report[s] on the shelf' was not correct, but neither was 'the stranger [that is, evaluator] with the axe' metaphor (with the implication of evaluations terminating programs). Dahler-Larsen's findings are:

... more consistent with a third image, in which evaluation may be viewed as a management support system. Managers use evaluation to produce knowledge not about whether programs work, but how they can be adjusted to work better. Managers may even sacrifice rigor and comprehensiveness for timeliness and relevance in evaluations, provided that evaluation knowledge can be translated immediately into actions to improve programs.

He links this finding with earlier work by Hellstern (1985) (who more than 20 years ago

wrote about this development) reporting that: 'Today [that is, 1985] evaluation is in the process of being integrated into administrative routine on many levels and in different institutions'.

The Danish study looked into a broad field of policies, but only regarding municipalities. However, a similar development can be seen in policy fields that have *no* direct relationship with municipalities, such as higher education programs and development aid.

In Europe, harmonisation of criteria and guidelines in the field of higher education evaluation and accreditation have made the criteria and evaluation approaches predictable and have turned the evaluative 'system' largely into a *tick-and-flick-system* to be 'used' by education managers with similarities to audit systems. If they contribute to learning, it is largely single loop in nature (Leeuw 2003, 2004).

In the development world the numbers of project evaluations, strategy evaluations, process evaluations and program evaluations easily add up to thousands and thousands (Leeuw & Cooksy 2004). Despite all this, the Center for Global Development (2006) recently referred to an 'evaluation gap', because valid and reliable information on the (*net*) *impacts or effects* of programs and interventions is rarely available.

Getting (process, project and strategy) evaluations going has become popular, in particular in western countries and in donor organisations, sponsoring governments and the World Bank and its look-alikes. *Evaluation capacity building (ECB⁵)* and the pressure to develop and implement *M&E (monitoring and evaluation) systems* have stimulated this development. But empirical and explanatory knowledge about (*net*) effects of programs and interventions is, despite all this, often lacking.⁶

Although there is nothing wrong with the *principles* of ECB and M&E (Beere 2005), on an aggregate level, and with many organisations implementing them, a system is being developed that creates the wrong incentive for evaluators, evaluands, policymakers and politicians. The incentive is not to develop and deliver the *best policies* (or no policies at all, letting markets and common property regimes do the work), but the incentive is to have ECB systems implemented and run in order to reduce risks to their reputations for principals, managers and sponsors (Power 2004). Clements, Chianca & Sasaki (2008) pointed at another downside in their study on 'results-based M&E' within the World Bank:

The disadvantage of results-based M&E is that it does not establish the worth of program results. A program that reaches all its timid targets may be less cost-effective than one that fails to reach ambitious goals. Indeed, a results-based evaluation regime establishes incentives for program planners to select targets that are easier to reach.

The consequences of the diffusion of ECB and M&E is the production of 'fleets' of reports, fact

sheets, memoranda, leaflets and handbooks that play their role as part of the management support system and that produce an institutional atmosphere of 'being-in-control' and 'assurance'. However, the downside is that this type of information is almost becoming an economic good that needs industry-type production, which is at best suited to produce information for single-loop learning. This type of production is probably also increasing what Pawson & Tilley (1997) have described as 'short-termism' in evaluation. From a politician's or administrator's position, 'short-termism' is of course relevant, but from the perspective of producing evaluative information that 'speaks truth to power' and that promotes understanding of what is going on and why, it is not.

Stern (2008, p. 254) sends a somewhat similar message regarding developments within the European Union (EU):

Evaluation in the member states of the EU ... has come uncomfortably close to the world of political power. The very success of evaluation, which is institutionalized in order to be close to policy-making, brings with it risks. We like to think that we can 'speak truth to power' but often we risk getting burned by getting too close to the flame!

Paradigm wars

Paradigm wars can have negative effects on the relevance of disciplines for society. Sociology during the 1970s and 1980s is an example⁷, as is educational research.⁸ Paradigm wars are different from *substantive, theoretical or methodological debates* between researchers with the goal of stimulating growth of knowledge. In a paradigm war a 'war' is going on with winners and losers. Labels and frames are used to put the other down, as are ostracism and other social 'techniques', such as nicknaming. The more a discipline or field of research is characterised by 'wars', the less it will add value to society.

How is the situation with regard to evaluation? Evaluation has had, and still has, its paradigm wars. Mertens (2005) wrote about it, as did Tashakkori & Teddlie (1998) in their book *Mixed Methodology: Combining Qualitative and Quantitative Approaches*.⁹ A classical example has been the fight between quantitative and qualitative methods that took place in the social sciences during the late 1960s and 1970s and had consequences for evaluation. A more recent one is the fight about the importance of randomised controlled trials (RCTs) versus other designs, when undertaking impact evaluations. Here an interesting example of the nickname strategy can be found; the not-too-positive term 'randomistas' started to come into use. The word even made it into *The Economist*.¹⁰ It designates those evaluators, often assumed to be economists, who are accused of believing that RCTs are the 'Only Golden Standard', when carrying out impact evaluations and that any other design or method choice is inferior. Nowadays,

also a nickname for the ‘opposition’ can be heard (‘relativistas’).

Empowerment evaluation is another example. Smith (2007) criticised Fetterman and Wandersman (2005) because their book provided:

... not merely a description of an evaluation approach but, more importantly, a description of what evaluation as an enterprise should be, what values and principles should dominate and direct its practice. Empowerment evaluation, then, in my judgment is less a method, less a theory, and primarily an ideology.

In a society where evaluation is an industry in its infancy, hardly anybody cares about these fights. Those sorts of fights have long gone; wars are now taking place within a *Big E(valuation) world* with streams of studies, large numbers of organisations, higher visibility, more money and power involved, and a greater need for bridging and binding between domain and approaches (Stern 2008). Thus paradigm wars are not contributing very much to the added value of evaluation to society. On the contrary, they even might give evaluation a negative reputation, as can be seen from a statement from industrial psychologist Perloff (2003): ‘Evaluation is a helter-skelter mishmash [and] a stew of hit-or-miss procedures’. In 2001, Mel Mark, a former president of the American Evaluation Association, suggested that we should try to ‘avoid the next generation of paradigm debates’.

Lack of explanatory power of evaluations: the Mind the Gap problem

What do the following findings in two fundamentally different domains have in common?

- 1 Put (2005) found on the basis of an analysis of performance audits reports of the UK National Audit Office and the Netherlands Court of Audit, that these organisations use frames of reference for their studies in which there is no place for insights from the social and behavioural sciences that deal with culture, reputations, social capital, human resources, exchange relationships, leadership, and types of learning within and between organisations. The same is true with insights from organisational economics, focusing on risk factors such as the drive for budget maximisation and rent seeking by bureaucrats and politicians. Instead the audit offices stuck to their ‘traditional’ and largely administrative models and variables.
- 2 In recent years, more and more attention is given to evidence-based interventions with regard to preventing and reducing anti-social behaviour. However, time and again, meta-analyses of evaluations have shown that while some of the current interventions are reasonably effective,

many adolescents and adults receive little or no benefit from these interventions. The impact of interventions on recidivism rates is also limited.

It has also been found that most of the factors looked after by evaluators, social workers and therapists in prisons and similar organisations are social and cognitive in nature. Attention paid to (neuro)biological and neuropsychological factors is limited, although there is evidence that an important part of persistent anti-social behaviour has neurobiological and genetic roots (de Kogel 2008; van Goozen & Fairchild 2006).

What both cases have in common are two aspects of the Mind the Gap problem. The first is that evaluations (and performance audits) often do not pay attention to behavioural mechanisms that underlie programs, strategies and interventions, whether or not they are focused on (changing) individual behaviour, group behaviour or organisational behaviour. Therefore, the number of ‘black boxes’ in evaluations continues to be large. Even the growing interest in what passes as theory-driven evaluation is often nothing more than simply a form of ‘analytic evaluation [that] involves no theory in anything like a proper use of that term’ (Scriven 1998, p. 59). While this might be a ‘normal’ evaluation practice, it does not constitute a theory underlying a program. To put it plainly, program logic is not the same as program theory. The process of unpacking black boxes to build program theory requires careful attention to the *explanatory ‘mechanisms’ of change* (Astbury & Leeuw, in preparation). In distinguishing between program logic (that is, ‘implementation theory’) and ‘program theory’, Weiss (1997, p. 46) appears to support these sentiments. Implementation theory, according to Weiss, provides details about how the program is carried out, while program theory:

... deals with the mechanisms that intervene between the delivery of program service and the occurrence of outcomes of interest. It focuses on participants’ responses to program service. The mechanism of change is not the program service per se but the response that the activities generate.

In a study of Dutch policy interventions (Leeuw 2008), I found that evaluators do not very often use (synthesis of) studies on mechanisms. Examples are Elster’s (1989) *Nuts and Bolts in the Social Sciences* and his revised edition (2007) *Explaining Social Behavior*, but also Hedstrom and Swedberg’s (1998) *Social Mechanisms* and Farnsworth’s (2007) *The Legal Analyst*. While these authors articulate social and behavioural mechanisms, based on studies ranging from neurology and psychology to institutional economics and political science, most evaluators do not refer to this type of work, making evaluations low ‘on explanation’.

The second dimension is that there is a discrepancy between ‘why questions’ formulated by

evaluators on the one hand, and the accumulation of knowledge within the disciplines on the other hand. Although there has been reported substantive accumulation of knowledge within and between the disciplines, its 'impact' on evaluation practice is limited. Speer (2009), for example, presents evidence with regard to Economics Nobel Prize winners in evaluation. She shows that many insights, models, theories and methods produced by Nobel Prize economists are often not used or recognised by evaluators. Mark (2001) also criticises the lack of attention paid to the accumulation of knowledge stating that:

A number of the sages in this issue [of the *American Journal of Evaluation*] indicate that evaluation practice is overly concerned about the single evaluation's findings. From this perspective, we do not pay enough attention to the aggregation of findings across evaluations ... Lipsey tells us that he has 'been bothered for some time about how little cumulative knowledge we are developing in the field about the programs we evaluate—what effects they do and don't have, for whom they work, how they bring about change, and so forth.

The rather limited attention paid to the explanatory power of theory and research regarding behavioural mechanisms, therefore, is a third probable cause of the phenomenon that is central in this article, that is, *booming business* but also *the beginnings of diminishing returns*.

Discussion

Evaluation has come a long way. What was hoped for by some of us decades ago, has almost come true: systems of evaluations, handbooks, organisations, reputation, and targeted (or 'imposed') use are all around. However, as is more often the case with goal-driven human behaviour, one is never sure to what extent *successes start to breed their own failures*.

This article addresses such a situation. Despite decades of debates about the 'added value' of evaluations, and despite the booming business in evaluation, there are clear signs of diminishing returns from our work. Three suggestions have been presented to explain this paradoxical situation. To help 'treat' this paradox, let me share two recommendations.

The first is to pay more attention to theories in the real sense of the word. Not to the *input-throughput-output diagrams with lines and circles*, but to *propositions that link in a causal or correlational way contexts, mechanisms and outcomes*. The social and behavioural sciences, as well as the (neuro)biological sciences, have produced important 'middle-range' theories (Merton 1968; van Goozen & Fairchild 2006) that are quite relevant for evaluators.

The second recommendation is to combine paradigms instead of prepare for more wars. A recent example was given by van der Knaap, Leeuw & Bogaerts (2008). For interventions focused on reducing violence in the public arena, the authors showed how the Campbell Collaboration approach with its systematic research review and its hierarchy of evaluation designs can be combined fruitfully with the realist approach, unpacking contexts—mechanisms—outcomes. It was also found that this combined effort stimulated the use of the findings by the Ministry of Justice of the Netherlands (Leeuw, van der Knaap & Bogaerts 2007).

Notes

- 1 A draft version of this article was given as a keynote address at the September 2008 annual Australasian Evaluation Society (AES) International Conference in Perth.
- 2 Weiss and colleagues (2008) also show that nowadays evaluations in the USA often have to be used by decision-makers, due to sunset legislation and similar arrangements ('targeted or imposed use'). Instead of being positive about this development, the authors are critical. This is because 'imposed use may ... have warts, and much care has to be exercised by everyone involved to improve the quality of the evidence available and the procedures used to select the evidence that is worthy of attention' (p. 16). Therefore, the authors believe it is necessary to look at *alternatives for the practice of imposed use*. 'Rather than to accept the current version of imposed use as the best we can hope for, let us study the consequences of alternative requirements for integrating evidence into policy' (p. 16). The alternative they refer to is that '... deliberate processes can take the summary of relevant research as the foundational starting point, inviting in values, interests, and experiences as adjuncts to, and not substitutes for, the research base' (p. 16). I believe that it would have been wiser to give the strategy of *imposed use* more time to develop, instead of immediately debunking it and presenting persuasive arguments for these alternatives.
- 3 Leviton (2003) is of the opinion that the work done by Henry and Mark (2003) on the contribution of evaluations to social improvement advances the thinking to new and higher levels. These authors refer to the influence that evaluations have on society and identify 'three levels of influence: individual, interpersonal, and collective' (p. 525). At each level, Henry and Mark have identified between four and six mechanisms, derived from social science literature, that may occur as a result of an evaluation.
- 4 Ludwig von Mises was probably the first to point at this, long before evaluations were popular. von Mises links it with what nowadays would be called bureaupolitics, budget maximisation and similar 'public choice'—phenomena.
- 5 ECB is commonly defined as 'the intentional work to continuously create and sustain overall organisational processes that make quality evaluation and its uses routine' (Baizerman, Compton & Stockdill 2002, p. 14).

- 6 Recent developments like 3ie and NONIE (the network of networks of impact evaluation) try to reduce this gap.
- 7 See my chapter on historical aspects of experimental evaluations in Rieper, Leeuw & Ling (in press).
- 8 Former Stanford University researcher and professor emeritus of education Nathaniel Gage referred to the situation, speaking about 'raging wars'.
- 9 Pawson & Tilley (1997) describe in neutral terms debates between paradigms in 'A history of evaluation in 28½ pages'.
- 10 See <<http://www.indiadevelopmentblog.com/2008/02/randomistas-score-again.html>>.

References

- Astbury, B & Leeuw, F, 'Unpacking black boxes: mechanisms and theory-building in evaluation', paper (submitted manuscript), University of Melbourne and University of Maastricht, the Netherlands.
- Baizerman, M, Compton, DW & Stockdill, SH 2002, 'New directions for ECB', in DW Compton, Baizerman & SH Stockdill (eds), *The art, craft, and science of evaluation capacity building: New directions for evaluation*, 93, Jossey-Bass, San Francisco.
- Beere, D 2005, 'Evaluation capacity-building: a tale of value-adding', *Evaluation Journal of Australasia*, vol. 5, no. 1, pp. 41–47.
- Center for Global Development 2006, *When will we ever learn? improving lives through impact evaluation*, report of the Evaluation Gap Working Group, CGD, Washington, DC.
- Clements, P, Chianca, T & Sasaki, R 2008, 'Reducing world poverty by improving evaluation of development aid', *American Journal of Evaluation*, vol. 29, no. 2, pp. 195–214.
- Dahler-Larsen, P 2000, 'The administrative use of evaluations in Danish municipalities; surviving the routinization of evaluation', *Administration & Society*, vol. 32, no. 1, pp. 70–92.
- de Kogel, K 2008, *Hersenen in beeld*, Boom Juridische Uitgeverij/WODC, Den Haag.
- Feller, I 2002, 'Performance measurement redux', *American Journal of Evaluation*, vol. 23, no. 4, pp. 435–452.
- Fetterman, DM & Wandersman, A (eds) 2005, *Empowerment evaluation principles in practice*, Guilford, New York.
- Frey, BS 2006, *Evaluities. eine neue Krankheit*, Working Paper No. 293, Institute for Empirical Research in Economics, University of Zurich.
- Frey, BS & Osterloh, M 2006, *Evaluations: hidden costs, questionable benefits, and superior alternatives*, Working Paper No. 302, Institute for Empirical Research in Economics, University of Zurich.
- Furubo, JE, Rist, RC & Sandahl, R 2002, *International atlas of evaluation*, Comparative policy analysis series, Transaction Publishers, New Brunswick, New Jersey and London.
- Hellstern, GM 1985, 'Assessing evaluation research', in FX Kaufmann, G Majone & V Ostrom (eds), *Guidance, control, and evaluation in the public sector*, Walter de Gruyter, Berlin.
- Henry, GT & Mark, MM 2003, 'Beyond use: understanding evaluation's influence on attitudes and actions', *American Journal of Evaluation*, vol. 24, no. 3, pp. 293–314.
- Leeuw, FL 2003, 'External evaluations of schools and higher education institutes: trends, topics and theories', in D Kristoffinson (ed.), *Comparing educational evaluation systems*, Danish Evaluation Institute (EVA), Copenhagen.
- Leeuw, FL 2004, 'Diminishing returns in evaluation? Challenges for educational evaluators', *Lifelong Learning in Europe*, vol. 9, no. pp. 106–114.
- Leeuw, FL 2008, 'Gedragmechanismen achter overheidsinterventies en rechtsregels', Inaugural address, University of Maastricht, the Netherlands, 23 May.
- Leeuw, FL & Cooksy, L 2004, 'Evaluating the performance of development agencies: the role of meta-evaluations', in GK Pitman et al. (eds), *Evaluating development effectiveness*, World Bank series on evaluation and development, vol. 7, Transaction Publishers, New Brunswick, New Jersey and London.
- Leeuw, FL & Furubo, JE 2008, 'Evaluation systems: what are they and why study them?', *Evaluation*, vol. 14, no. 2, pp. 157–169.
- Leeuw, FL, Toulemonde, J & Brouwer, A 1999, 'Evaluation activities in Europe: a quick scan of the market in 1998', *Evaluation*, vol. 5, no. 4, pp. 487–496.
- Leeuw, FL, van der Knaap, LM & Bogaerts, S 2007, 'Reducing the knowledge–practice gap: a new method applied to crime prevention', *Public Money & Management*, vol. 27, no. 4, pp. 245–251.
- Leviton, L 2003, 'Evaluation use: advances, challenges and applications', *American Journal of Evaluation*, vol. 24, no. 4, pp. 525–535.
- Mark, M 2001, 'Evaluation's future: furor, futile, or fertile?', *American Journal of Evaluation*, vol. 22, no. 3, pp. 457–479.
- Merton, R 1968, *Social theory and social structure*, Free Press, New York.
- Mertens, D, 2005, *Research methods in education and psychology: integrating diversity with quantitative and qualitative approaches*, Sage, Thousand Oaks, California.
- Pawson, R & Tilley, N 1997, 'A history of evaluation in 28½ pages', in R Pawson & N Tilley, *Realistic evaluation*, Sage, London.
- Perloff, R 2003, 'A potpourri of cursory thoughts on evaluation', *Industrial-Organizational Psychologist*, vol. 40, no. 3, pp. 52–54.
- Power, M 1999, *The audit society*, Penguin Books, London.

- Power, M 2004, *The risk management of everything: rethinking the politics of uncertainty*, Demos, London.
- Preskill, H 2008, 'Evaluation's second act: a spotlight on learning', *American Journal of Evaluation*, vol. 29, no. 2, pp. 127–138.
- Put, V 2005, 'Normen in performance audits van rekenkamers: een casestudie bij de Algemene Rekenkamer en het National Audit Office (2002–2005)', PhD thesis, University of Leuven, Belgium.
- Rieper, O, Leeuw, FL & Ling, T (eds) in press, *The evidence movement*, Transaction Publishers, New Brunswick, New Jersey and London.
- Rist, RC & Stame, N (eds) 2006, *From studies to streams: managing evaluative systems*, Transaction Publishers, Piscataway, New Jersey.
- Ryan, B 2003, 'Death by evaluation? Reflections on monitoring and evaluation in Australia and New Zealand', *Evaluation Journal of Australasia*, vol. 3, no. 1, pp. 6–16.
- Scriven, M 1998, 'Minimalist theory: the least theory that practice requires', *American Journal of Evaluation*, vol. 19, no. 1, pp. 57–70.
- Shulha, LM & Cousins, JB 1997, 'Evaluation use: theory, research, and practice since 1986', *American Journal of Evaluation*, vol. 18, no. 1, pp. 195–208.
- Smith, NL 2007, 'Empowerment evaluation as evaluation ideology', *American Journal of Evaluation*, vol. 28, no. 2, pp. 169–178.
- Speer, S 2009, 'Economics and evaluation', in J Vaessen & FL Leeuw (eds), *Mind the gap: perspectives on policy evaluation and the social sciences*, Transaction Publishers, New Brunswick, New Jersey and London.
- Stern, E 2008, 'Evaluation: critical for whom and connected to what?', *Evaluation*, vol. 14, no. 2, pp. 249–257.
- Tashakkori, A & Teddlie, C 1998, *Mixed methodology: combining qualitative and quantitative approaches*, Sage, Thousand Oaks, California.
- Vaessen, J & Leeuw, FL (eds) 2009, *Mind the gap: perspectives on policy evaluation and the social sciences*, Transaction Publishers, New Brunswick, New Jersey and London.
- van der Knaap, L, Leeuw, FL & Bogaerts, S 2008, 'Combining Campbell standards and the realist evaluation approach: the best of two worlds?', *American Journal of Evaluation*, vol. 29, pp. 48–57.
- van Goozen, SHM & Fairchild, G 2006, 'Neuroendocrine and neurotransmitter correlates in children with antisocial behavior', *Hormones and Behavior*, vol. 50, no. 4, pp. 647–654.
- van Thiel, S & Leeuw, FL 2002, 'The performance paradox in the public sector', *Public Productivity & Management Review*, vol. 25, no. 3, pp. 267–281.
- Walters, R 2008, 'Government manipulation of criminological knowledge and policies of deceit', in W McMahon (ed.), *Critical thinking about the uses of research*, Centre for Crime and Justice Studies, King's College London.
- Weiss, CH 1997, 'Theory-based evaluation: past, present, and future', in DJ Rog & D Fournier (eds), *Progress and future directions in evaluation: perspectives on theory, practice and methods: New directions for evaluation*, 76, Jossey-Bass, San Francisco.
- Weiss, CH 1998, *Evaluation: methods for studying programs and policies*, 2nd edn, Prentice Hall, Upper Saddle River, New Jersey.
- Weiss, C, Murphy-Graham, E, Petrosino, A. & Gandhi, AG 2008, 'The Fairy Godmother—and her warts: making the dream of evidence-based policy come true', *American Journal of Evaluation*, vol. 29, no. 1, pp. 29–47.
- World Bank 2004, *Influential evaluations: evaluations that improved performance and impacts of development programs*, Operations Evaluation Department, the World Bank, Washington, DC.