

**Evaluation of joined-up policy initiatives:
Some difficulties in assessing ‘what works’**
การประเมินผลนโยบายที่หลายหน่วยงานร่วมมือดำเนินการ:
ข้อจำกัดบางประการในการประเมินสัมฤทธิ์ผลของนโยบาย

อมรศักดิ์ กิจชนานันท์
Amornsak Kitthananan

บทคัดย่อ

นโยบายสาธารณะในปัจจุบันได้พัฒนาไปสู่ระดับของการพยายามแก้ไขปัญหาการขาดการบูรณาการระหว่างงานต่างๆ ที่แตกกระจัดกระจายอยู่ในแต่ละหน่วยงาน ทั้งทั้งงานเหล่านี้มีความเชื่อมโยงสัมพันธ์กันทางใดทางหนึ่ง บทความนี้พยายามจะชี้ถึง สิ่งท้าทายในการประเมินผลนโยบายท่ามกลางทิศทางการเปลี่ยนแปลงไปสู่การกำหนดนโยบายเชิงบูรณาการ ปัญหาและข้อท้าทายที่พบมีทั้งในแง่ของข้อท้าทายทางทฤษฎีและระเบียบวิธีวิทยาการประเมิน การพยายามสร้างการมีส่วนร่วมของผู้มีส่วนได้ส่วนเสียทั้งหลายในกระบวนการ นโยบายได้กลายเป็นสิ่งที่มีความสำคัญยิ่งขึ้น ในขณะเดียวกัน “ตรรกะของการประเมินผล” แบบเดิมก็มีความจำกัดในการที่จะพยายามอธิบายว่า นโยบายใดได้ผลหรือไม่ รวมทั้งการประเมินผลลัพธ์และผลกระทบของมาตรการนโยบาย ที่มีต่อผู้มีส่วนได้ส่วนเสียแต่ละกลุ่มดูจะเป็นสิ่งที่แยกแยะได้ลำบากขึ้น

Abstract

Policy initiatives have moved to another stage, where the policy makers attempt to overcome the problem of disintegration in

policy formulation and delivery. This development poses more challenges to evaluation, both theoretically and methodologically. The evaluation procedures of the 'logic of evaluation' seem to be very limited in describing 'what works.' Difficulties can be found in both evaluating the policy outcomes and in the process of implementation. Stakeholder participation becomes increasingly important. Complex policy conditions have some important implications for evaluation, as they require us to challenge some basic assumptions that underpin traditional approaches to evaluation. In any case, fundamental inquiries are still waiting for the evaluators to tackle the new complex inter-agency conditions: identifying the effects of policy intervention in relation to different stakeholders, and understanding how such effects are brought about.

Introduction

Assessing the impact of social policy has become increasingly difficult as joined-up¹ activities have grown in popularity and more embedded as a major means of delivering public policy. The need for building adequate frameworks to monitor and assess such collaborative activities is increasingly profound. While inter-agency collaboration can take a number of forms, some of them demand frameworks that can accommodate the logic of evaluation in different ways to assess the outcomes of the cross-cutting collaborative mechanisms. This article discusses difficulties in assessing the success of policy initiatives and telling 'what works'. It employs the case of policy initiatives for alleviating poverty and social problems in Thailand as an example of such difficulties.

¹ Some other terms may be used interchangeably with this word such as collaborative, partnership, cross-cutting, and inter-agency.

The Logic of Joined-Up Policy Initiatives and the Challenges for Evaluation

Joined-up policy initiatives entail working across traditional departmental boundaries in order to tackle social problems in a more holistic fashion. The logic behind this joined-up approach to policy is that work in this collaborative manner is believed to help in resolving the problem of a fragmented public sector resulting from the specialisation principle of the ideal-type bureaucracy. This inter-agency collaborative measures helps avoid duplication, promotes adaptability and flexibility, and builds the capacity of mainstream agencies to adapt and change their ways of working to address complex social problems effectively. Further, it is also believed that maximum effectiveness in addressing complex cross-cutting policy problems can be achieved by coordinating and integrating mainstream policies, focusing them on the problems concerned, and providing supplementary/additional measures and resources only where these can enhance the positive impacts and benefits of existing policies and programmes (Sanderson, 2000). In other words, the core assumption of this approach involves getting the related agencies to work together in a coordinated way within a strategic framework to create synergies and to provide catalytic resources to lever out the benefits of joined-up working. To find out how far this assumption can be achieved in reality demands rigorous evaluation.

The core purpose of evaluation is to determine the effects or effectiveness of a policy intervention. Although evaluation almost always has a political dimension when policy makers need evaluation to demonstrate real improvement in the living conditions of the people and to justify their initiatives, it is still the key activity to determining 'what works' in

practice. Evaluation comprises a chain of activities that enable the evaluator to make valid judgements about the matter under consideration. Such chain of activities is termed the *'logic of evaluation'* (Fournier, 1995) consisting four elements (or steps): establishing criteria, constructing standards, measuring performance, and synthesising and integrating evidence into a judgement of worth. It is very simple to say that by following these steps logically, we will come up with a good judgement over the policy interventions being evaluated. However, when the policy initiatives become more joined up, the task of assessing the policy outcomes/impacts is not simple anymore. This policy context presents a number of implications for evaluation. I will discuss these implications after introducing the case of recent social Problems and Poverty Alleviation Programme in Thailand.

The Integrated Social Problems and Poverty Alleviation Programme

The government who took the administration in 2001 committed in its election campaign to fighting poverty. Thus, following the national meeting on poverty alleviation on 1st October 2003 and 15th November 2003, the Thai government launched a nationwide poverty and social problems registration programme initially lasting from 5th January to 31st March 2004. This programme was called the Integrated Social Problems and Poverty Alleviation (ISPA) programme. The programme allowed peoples who need assistance from the government to identify their problems according to the categories pre-identified by the government as (1) lack of agriculture lands, (2) homelessness, (3) illegal occupation, (4) student unemployment, (5) frauds, (6) personal debts, (7) housing, and (8) other problems². Anyone

² This approach of poverty registration by people themselves was under serious criticisms by many academics and NGOs. However, the aim of this article is not about discussing such criticisms in depth. I leave this issue for others to pursue.

who considered him/herself had problems as defined was invited to report him/herself at the District Office. The person filled out a form stating details of his/her problems and major reasons for having them. The District Office then passed on the information and roster of the registered poor to the village where they were from and asked the Village Assembly to review and assess whether the information was genuine. The confirmed information and revised roster were passed to the central National Center for Poverty Eradication (NCPE) for assistance in alleviating the identified problems. Table 1 presents an overview of the registration at the end of September 2004. In total, 8,258,275 people came to register (13.15 percent of total population) and 12,198,333 cases of problems were identified³. It was very interesting that the number of self-assessed poor from this programme exceeded the number of income-poor according to the national poverty line (poverty incidence was at 9.8 percent in 2002 national survey).

³ Number of cases exceeded the number of registered people because one registered person may have more than one problem.

Tabel 1: Summary of the Social Problem and Poverty Registration Programme

Problems	Numbers	Percentage	Lead agency
Personal debt	4,680,832	38.37	Ministry of Finance
Lack of agricultural lands	3,968,119	32.53	Agricultural Land Reform Office
Housing	1,892,131	15.51	National Housing Authority
Students Unemployment	240,345	1.97	Ministry of Labour
Frauds	82,720	0.68	N/A
Illegal occupations	7,546	0.06	N/A
Homelessness	5,064	0.04	N/A
Other problems	1,321,576	10.83	N/A
Total	12,198,333	100.00	-

Source: National Centre of Poverty Eradication as of 22nd September 2004

NCPE was established on 19th November 2003 to become a national central institution to oversee and manage the process of solving problems related to poverty, but at the first stage, it focused on the problems and cases identified by this registration programme. The centre was chaired by the Deputy Prime Minister. Its executive committee consisted of representatives from all relevant partners-including business, civil society, and ministries-but members from the government agencies formed the majority of the committee. Its main responsibilities were to formulate policies, strategies and measures, as well as to monitor and evaluate government policies on poverty alleviation. The concept of joined-up initiatives was adopted here (the term 'integrated' was used instead of joined-up). Because *social problems and poverty* in this programme were defined in many dimensions as categorised into seven problems, cross-cutting agency collaboration was implemented. A couple of subcommittees were

established under the centre to look after the various issues related to the problems people registered. Each of them was chaired by the government agency that held a mandate over each specific problem while its members came from all related agencies. The subcommittees were:

- Subcommittee for promoting occupational development and employment, led by the Ministry of Labour;
- Subcommittee for Housing, led by the National Housing Authority;
- Subcommittee for restructuring debt for the poor, led by the Ministry of Finance;
- Subcommittee for Land Allocation, led by the Agricultural Land Reform Office;
- Subcommittee for reviewing the structure of economy, education, law and regulations related to poverty eradication, led by the Office of the Council of State;
- Subcommittee for formulating community development plan for sustainable poverty alleviation, led by the Local Development Institute (a famous local non-governmental organisation); and
- Subcommittee for Water Resource Allocation, led by the Department of Water Resources (Ministry of National Resources and Environment).

Although people registered individually, the policy initiatives were taken on the area-based constituency. Each subcommittee operated at the district/village level because it was the area where people actually lived and it helped the subcommittees to operationalise the measures. This inter-agency operation made the coalition more complex. For example, in a village, the operational team of the Housing subcommittee came to

contact the registered person who reported his/her problem in lacking of housing while the same person may report his/her personal indebtedness which was tackled by a team from the Debt Restructuring subcommittee. The same person could also be contacted by a team from the Ministry of Labour if he/she reported as an unemployed. This complexity of the policy initiatives in the context of inter-agency collaboration made assessment more difficult as policy outputs cannot be looked in isolation on the basis of single agency. This posed some difficulties for evaluating the outcomes of the programme and to tell 'what works'.

Evaluating Outcomes and Process Evaluation

An evaluation of this joined up policy could be considered by distinguishing between implementation process evaluation and outcomes evaluation. We need to be able to tell if the outputs and outcomes have been attained, and trace a pathway between the outputs/outcomes and the activities that contributed to them. However, difficulties can be found in both evaluating the policy outcomes and the process of implementation.

Evaluating the outcomes is defined as an assessment of the benefits the participants have received from the products of the programme. The evaluation of outcomes is technically difficult when we attempt to assess whether the given interventions have brought about the expected benefits to the target participants of the programme (Sullivan and Skelcher, 2002). There are major conceptual problems associated with formulating causal relationships between interventions and outcomes, and specifying indicators towards progress also presents significant difficulties. Where indicators are specified, they may in fact operate in tension with each other. For example, improving the quality of life of people who need good housing in the ISPA programme may involve the demolition of high-rise flats

and their replacement by low-rise accommodation at a lower density. To realise this, the blocks have to be emptied and families temporarily found alternative accommodation. This process may have adverse effects on the school attendance and performance of children, thus detrimentally affecting an outcome indicator about improving local educational attainment which is under the responsibility of the Ministry of Education. Further, many of the outcomes sought will be in relation to cross-cutting issues that require concerted action by various agencies. This poses a problem in disentangling the causes of outcomes and who can legitimately 'claim' the benefits and avoid the problem of double counting.

For the process evaluation, the concern is that evaluation does not take place in a static environment. Instead, changes in circumstances and the behavior of individuals participate in the programme can contribute positively or negatively to the achievement of outcomes, and thus need to be accommodated in the evaluation. Only checking afterwards whether objectives of the policy have been achieved is not enough. Evaluation practice is now required to study service delivery mechanisms that they were in the past treated as a 'black box' of the policy delivery. However, evaluating the process can be highly complicated when multiple actors participate in the implementation process and many agencies attempt to reach the same group of people. In the case of ISPA, it was very likely that one-third of the registered people was in debt and lacked agricultural lands, and 15 percent of them needed good housing. These groups of people were very likely to be approached by teams from four subcommittees. This condition caused the difficulty of telling 'who makes what'.

Procedural Difficulties in Evaluating Joined-Up Policy Initiatives

Another way of understanding difficulties in evaluating inter-agency initiative is by considering a chain of activities mentioned in the first section. We can find difficulties in evaluating the joined-up policy at every components of the 'logic of evaluation'-establishing criteria, constructing standards, measuring performance, and making judgement.

In the stage of establishing criteria, evaluation activities start with defining what counts as the criteria for judgment, what evidence to collect, and how that data is to be analysed. Different approaches to evaluation would specify different requirements. However, in the conditions posed by evaluation of joined-up initiatives, which are normally conducted by various agencies, the dimension of evaluation criteria and who is involved in determining such criteria become a critical issue of evaluation. When the governance structure and delivery of the policy initiatives are increasingly the responsibility of various agencies and sectors, this collaboration of multiple agencies could initiate value conflict in evaluation as the criteria of policy aims and consequences may vary amongst agencies involved. The Minister of Labor paid high attention to the criteria of employment rate while the National Housing Authority focused on boosting the volume of accommodation for register people. It was very difficult for both agencies to establish a common base line with which all policy impacts could be assessed. Further, the issue of whose agenda should be used as evaluation criteria is very difficult to agree on. This applies to the ISPA programme which we can expect more complicated process and industry needed to set agreement on the evaluation criteria amongst the seven agencies involved.

The next step is the constructing of standards-the determination of standards of performance in relation to each criterion is determined to inform judgments of what works. Generally, setting objective standards for policy evaluation is often complicated by the difficulty of defining tangible measures of performance and the need to take account of many other extraneous factors that may have influenced performance. An attempt to identify any causal relationships in the inter-agency initiative is extremely difficult because there are many interventions that have been engaged in the same target group. It is hard to deny that the same group of people in each constituency is likely to be the beneficiaries of policy interventions from other departments that are not involved in the ISPA programme. Thus it posed the difficulty of disentangling the influence of other policies which are directed at the same target people. For example, the Ministry of Commerce might employ a policy to support the property market to boost up economic growth by implementing incentives for housing constructors to reduce the price. The price reduction in property market attracted the new homebuyers who were very likely to involve those registered people in the ISPA programme, who were also a target group of the work of the Housing Subcommittee. How could we set defined standards for the work of the Housing Subcommittee that can assure the outcomes of policy were caused only by the work of Housing Subcommittee? The inter-agency initiatives make the evaluation impossible to insulate the influence of interventions of other policies. Further, when the ISPA programme was applied across the country at the same time, it posed another difficulty of evaluation. The government was implementing the programme throughout the county but it left the decision in regard to the details of measures employed by each agency to the operational teams at the ground level, the standards could be very different according to the different measures. Thus, it was uneasy

to apply a single set of performance standards to produce an estimate of the net outcomes of the programme as a whole.

When proceeding to the measuring of performance, the evaluators need to establish ways of articulating the expected impact of collaboration on goal achievement and the value added by the act of joined up. Among the most important obstacles to measuring change is the difficulty of establishing a reliable baseline of assessment (Martin and Sanderson, 1999). The ISPA programme had been implemented across the country thus it was so large that collecting detailed primary data about all of the initiatives would be prohibitively expensive. Further, there were no reliable sources of appropriate secondary data of the conditions directly related to the target areas or people. Such constraints could cause the before-after evaluation design not applicable in this case. Another factor effecting the measuring of the outcomes is the timing of evaluation. It seems very unlikely that measurable improvements in outcomes observed over the early years of the programme, especially the community development initiatives, will reflect the real impact of the programme. An achievement of the short-term performance is unlikely to guarantee any major change in the later years. Releasing people from poverty needs a longer timeframe of evaluation to assure that the poor have been completely relieved. For example, helping people from their indebtedness requires a longer time for evaluators to see if they will not return to be indebted again.

At the last stage, synthesising and integrating evidence into a judgment of worth which means the judgments of the outcome of the evaluation have to be made on the basis of the evidence gathered and regardless of any inherent or preferred findings. Evaluation attempts to establish the causal relationships between the interventions and the policy outcomes and this could be technically difficult. The problem involved in assessing the

impacts of joined-up policy intervention concerns the 'value' that should be placed on different kinds of outcomes (Martin and Sanderson, 1999). Ministers, local politicians, senior executives, professional groups, front-line staff, and service users will almost certainly place different values on different kinds of outcomes. Especially, in the absence of a clear definition of how priorities are to be determined, it is not easy to weight different kinds of outcomes or to reflect accurately the spectrum of competing views and values. Further, the joined-up policy initiatives are multi-party activities and consequently the normal processes of debate and disagreement at the end of any evaluation could be exaggerated. The disagreement can be reconciled only if the criteria and standards established at the early stage of the evaluation were agreed by all parties. But it is not an easy task to achieve for the case of ISPA as I have explained.

Theoretical Difficulties in Evaluating Joined-Up Policy Initiatives

This joined-up policy-making context presents a number of theoretical difficulties for evaluation. Firstly, faced with such complex policy situations, evaluators encounter substantial difficulties in operationalising theory-based approaches to evaluation (Sanderson, 2000). Theories comprise propositions about how policies and programmes generate or produce effects and expected outcomes. An important purpose of theory-based approach to evaluation is to test such propositions by drawing the plausible inferences of evidence about what is likely to work elsewhere. Difficulties in operationalising such approach arise due to conventional conceptions of the process of theory development as a deductive process of hypothesis testing. In a joined-up policy setting where potentially many programmes interact and hypotheses about their effects are likely to be heavily context-dependent, the conventional evaluation designs which need to control the

conditions for rigorous theory testing are hard to achieve. The difficulty arises from the practical constraints faced in specific evaluation contexts that often conspire against rigorous hypothesis testing.

Secondly, the context plays an important influence in evaluation. It raises the issue of context dependency which the 'realistic' evaluators advocate (Pawson and Tilley, 1997). They argue that the evaluation of social programme takes place within dynamic environments in which the environment (its location, institutions, and so on), that the intervention operates, is as significant as the intervention itself in determining programme outcomes. Their approach emphasises the complexity of establishing cause-effect relationships in environments by focusing its attention on increasing understanding of the programme under consideration rather than attempting to construct control or comparison sites against which to judge the programme. Individual components of policy interventions cannot be evaluated effectively if they are taken out of the context. It implies the necessity to identify and understand the influence of the key contextual factors and the 'historical' nature of the task. Sanderson (2000) suggests that, therefore, a key element to understand the influence of context will be comparative analyses over time of carefully selected instances of similar policy initiatives implemented in different contextual circumstances. Under such conditions, controlled (quasi-) experimental comparisons are unlikely to be feasible.

Thirdly, furthermore, the evaluation of policy intervention in the context of complex policy systems, especially the current emphasis on joined-up policy initiatives, must cope with the 'embeddedness' of such intervention within a wider range of social systems (Schmid, 1996). The approach to evaluation must be able to accommodate complex, multi-programme initiatives in which there is limited potential to analyse specific

measures in a controlled way. In the past, policy evaluation had focused too much on the effects of single programmes and on effects at the level of the individual agent and this has been a key factor in the problem of lack of external validity (Sanderson, 2000). This means in the context of complex policy systems, if evaluation is to be useful in providing meaningful lessons about 'what works' in various policy conditions, it must be able to provide evidence of effects at the macro-level. The problem related to this issue is the provision of such evidence is a long-term venture. Sabatier (1986) proposed the problem of *premature assessment* of the effects of policies and programmes. The kinds of changes as the result of policy interventions will not be manifested for a much longer period than initiatives typically track. An example of community initiatives, it is not very common for the planning, capacity building and start-up phases of an initiative to take three or more years. This means that an evaluation which tracks an initiative for its first five years will cover only a relatively short period during which the initiative is functioning at full capacity. In the case of ISPA, which the evaluation was done after the first year of implementation, it was only able to evaluate the output attainment of the programme with no capacity to draw the evidence to the outcomes/impacts/. However, extending the time frame of the evaluation would raise other problems. It is very difficult to maintain the focus of the evaluation, as well as the initiative, over the long term, and retaining the interest and commitment of the funders (Auspos and Kubisch, 2004).

Such timeframe is needed to be able to distinguish, on the one hand, the effects of the implementation structure and, on the other hand, the validity of the relevant policy theory. The latter is crucially dependent upon longer-term impacts of policy. It means that, in the policy context, there may be a long period of little discernible change in response to

policy intervention before the conditions become right for synergy effects of the joined-up policies to produce significant change. This presents a potentially serious problem for policy pilots in testing policy interventions (Sanderson, 2000). In the relatively short timeframe that is politically demanded for such pilots and their evaluation, it is likely that policy impacts will not be discernable and that the focus will necessarily be on implementation processes.

Methodological Difficulties in Evaluating Joined-Up Policy Initiatives

Quantitative methodologies maintain their dominance in different evaluation approaches. Positivist assumptions privilege the notion of objective, value-free knowledge derived through quantitative social research methodologies which claim to provide relatively reliable, hard and factual data (Guba and Lincoln, 1989; Van der Knaap, 1995). However, Albaek (1995) argues that the dominance of such assumptions has been weakened as the influence of positivism in the social science has been challenged while the phenomenological and hermeneutic traditions have gained strength, providing the growing credibility for qualitative approaches. Quantitative methodologies in evaluation are confined by an experimental (or quasi-experimental) design. However, the pool of potential 'matching cases' of the similar size, from the same region, and under the same sociopolitical conditions is harder to define. Experimentation and quasi-experimentation are normally ruled out as a mean of establishing the counterfactuals in doing evaluations of social programme because of a lack of appropriate control groups. When the experimentation and quasi-experimentation is difficult to apply, the before-after comparison analysis of the progress in the area made by the programme initiatives becomes an alternative. How-

ever, this approach still does have a problem in differentiating between genuinely new initiatives and the pre-date intervention of other policies. It is hard to disentangle the effects of other policies that are implemented in the same area. The qualitative approach to evaluation is then used when there are problems that are operationalised in quantitative designs or in a supplementary way to investigate implementation processes (Shaw, 1999). Nevertheless, qualitative approaches do not escape criticism as they lack preciseness in disentangling effects and confound situational with systemic effects, so there are no clear results to be used for programme design (Hellstern, 1986 cited in Sanderson, 2000).

Since both quantitative and qualitative approaches each have their own limitations, methodological pluralism is now gaining support in evaluation as we need to collect evidence from different angles and to construct an intelligible picture of the intervention effects (Schmid, 1996). An implication of this position is the need to include in the scope of evaluation all relevant stakeholders, recognising their interest in the policy and seeking to interpret their perspectives and arguments in relation to their institutional and social context (Connell *et al.*, 1995). In the context of evaluating joined-up policy, this demands well-developed skills in triangulation on the part of evaluators and the ability to exercise balanced analysis of the policy effects.

Evaluation of the inter-agency initiatives means by itself that the evaluation has to be conducted amongst a range of stakeholders who normally have very varied views about the purposes and design of the evaluation as well as the interpretation of its results. This tends to make evaluators operate in a more complex condition. Thomas and Palfrey (1996) identify three groups of stakeholders: those who pay (such as the government on behalf of the taxpayer, beneficiaries, other customers), those who

benefit (for example clients), and those who provide (including professionals, managers and politicians). The valid evaluation requires the consideration of a range of these stakeholders. However, each stakeholder certainly has different views and differential influence over the evaluation process. Those who exercise most influence over how evaluations are conducted and what is evaluated are those with some responsibility for both payment and provision. While those with the least amount of influence over the evaluation process are those who are intended beneficiaries of the policy intervention or individual customer. The imbalance of power relationship is considerably important in determining the criteria and use of evaluations involving many stakeholders. Thus bringing more people into evaluation means creating a bigger room for confusion and argument.

Conclusion

This article aims to elaborate the difficulties to evaluating the joined-up policy initiatives with reference to a case of ISPA programme in Thailand. The policy initiatives have moved to another stage where the policy makers attempt to overcome the problem of disintegration in policy formulation and delivery. This development poses more challenges to evaluation theoretically and methodologically. The evaluation procedures of the 'logic of evaluation' seem to be very limited in telling 'what works'. Difficulties can be found in both evaluating the policy outcomes and the process of implementation. Stakeholder participation becomes increasingly important. Methodological pluralism is gaining increasing support in evaluation. Complex policy conditions have some important implications for evaluation, as they require us to challenge some basic assumptions that underpin traditional approaches to evaluation. We now have to realise that

evaluation is becoming necessarily a highly complicated endeavour. Evaluation must be seen as an exercise in crafting an approach comprising a range of methods appropriate to particular circumstances, which will provide some understanding of the policy initiatives in promoting change. In any case, fundamental inquiries are still waiting for the evaluators to tackle in the new complex inter-agency conditions: identifying the effects of a policy intervention in relation to different stakeholders, and the understanding how such effects are brought about.

References

- Albaek, E. 1995. "Policy evaluation: Design and utilization." in *Policy evaluation: Linking theory to practice.*, ed. R. C. Rist. Aldershot: Edward Elgar.
- Auspos, P., & Kubisch, A. C. 2004. *Building knowledge about community change: Moving beyond evaluations.* New York: Aspen Institute Roundtable on Community Change.
- Clarke, A., & Dawson, R. 2000. *Evaluation research: An introduction to principles, methods and practice.* London: Sage.
- Connell, J. P. and others. *New approaches to evaluating community initiatives, : Concepts, methods and contexts.* Washington DC: Aspen Institute.
- Coote, A., Allen, J., & Woodhead, D. 2004. *Finding out what works: Building knowledge about complex, community-based initiatives.* London: King's Fund
- Fournier, D. 1995. "Establishing evaluative conclusions: A distinction between general and working logic." *New Directions for Program Evaluation* 68: 15-32.
- Guba, E. G., & Lincoln, Y. S. 1989. *Fourth generation evaluation.* Newbury Park, CA: Sage.
- Martin, S., & Sanderson, I. 1999. "Evaluating public policy experiments: Measuring outcomes, monitoring processes or managing pilots?" *Evaluation* 5, no. 3: 245-258.

-
- Owen, J. M., & Rogers, P. J. 1999. *Program evaluation: Forms and approaches*. London: Sage.
- Pawson, R., & Tilley, N. 1997. *Realistic evaluation*. London: Sage.
- Sabatier, P. A. 1986. "What can we learn from implementation studies?" in *Guidance, control and evaluation in the public sector*, eds. F. Kaufman, M. Majone, V. Ostrom & W. Wirth. Berlin: Walter de Gruyter.
- Sanderson, I. 2000. "Evaluation in complex policy systems." *Evaluation* 6, no. 4: 433-454.
- Schmid, G. 1996. "Process evaluation: Policy formulation and implementation." in *International handbook of labour market policy and evaluation*, eds. G. Schmid, J. O'Reilly & K. Schomann. Cheltenham: Edward Elgar.
- Shaw, I. 1999. *Qualitative evaluation*. London: Sage.
- Sullivan, H., & Skelcher, C. 2002. *Working across boundaries: Collaboration in public services*. New York: Palgrave Macmillan.
- Thomas, P., & Palfrey, G 1996. "Evaluation: Stakeholders-focused criteria." *Social Policy and Administration* 30, no. 2: 125-142.
- Van der Knaap, P. 1995. "Policy evaluation and learning: Feedback, enlightenment or argumentation?" *Evaluation* 1, no. 2: 189-216.