

Evaluation

<http://evi.sagepub.com/>

The Emerging Gap between Evaluation Research and Practice

Evaluation 2003 9: 340

DOI: 10.1177/13563890030093007

The online version of this article can be found at:

<http://evi.sagepub.com/content/9/3/340>

Published by:



<http://www.sagepublications.com>

On behalf of:



The Tavistock Institute

Additional services and information for *Evaluation* can be found at:

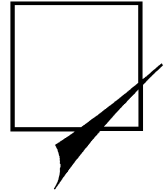
Email Alerts: <http://evi.sagepub.com/cgi/alerts>

Subscriptions: <http://evi.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations: <http://evi.sagepub.com/content/9/3/340.refs.html>



The Emerging Gap between Evaluation Research and Practice

ABDUL KHAKEE

Umeå University and Örebro University, Sweden

While evaluation practice is still disposed towards rational quantitative methods, evaluation research has increasingly utilized qualitative dialogical methods. In this article, the increasing gap between practice and research is examined by analysing how evaluation research has evolved from within three perspectives: a policy programme perspective, a welfare economics perspective and a planning theory perspective. The article also discusses the implications of the emerging gap between evaluation research and practice.

KEYWORDS: planning; policy analysis; positivism; post-positivism; welfare economics

Introduction¹

At the beginning of the 1990s, Mary Henkel, who has many years of experience as an evaluator in public audit agencies, published a book called *Government, Evaluation and Change* (Henkel, 1991). In her book, Henkel highlights existing tensions between evaluation research and evaluation practice. She argues that evaluation research is rapidly moving away from positivism and conventional measurement methods while evaluation practice is still firmly oriented within positivism and quantitative evaluation methods. According to Henkel, this is because politicians and government officials remain firmly convinced that things can be measured and evaluated in a rational way.

This article takes Henkel's remarks as a point of departure in order to discuss the emerging gap between evaluation research and practice. The main purposes of the article are: to analyse the convergent development in evaluative research, which runs counter to existing practice; and to examine the implications for public service evaluation. Since it is very difficult to get an overview of evaluation practice, a brief survey of evaluative practice carried out in a few policy areas in Sweden, including the Swedish National Audit Office, is presented. The major aim of this article is to analyse developments in evaluative research within three traditions in order to explain the nature of the challenge they pose to politicians, policy makers and public sector managers.

Evaluation research can be viewed from several different perspectives. This article focuses on three of them:

- evaluation research from within a policy programme perspective, e.g. education, healthcare, etc.;
- evaluation research from within a welfare economics perspective;
- evaluation research from within a planning theory perspective.

These three perspectives have been selected because they cover major traditions in evaluative research as well as in public service evaluation. These perspectives contribute towards our understanding and application of evaluative research in all important policy fields and in urban and regional planning.

Development in evaluative research within each of these three perspectives will be briefly described in the following sections and then the trends they share will be highlighted, followed by a brief survey of evaluative practice in Sweden. The concluding section of the article discusses the implications of the development in evaluative research. Prior to this I shall briefly discuss positivism and post-positivism and their impact on policy research especially with reference to what Henkel has in mind when she talks about the tensions between evaluation research and practice.

Positivism, Post-positivism and Evaluation

Evaluation theory and evaluation methods have, like much of the other policy research, been determined by positivism's ontological and epistemological assumptions. For example, Torgerson (1986: 35) writes that '[p]olicy analysis today bears the unmistakable imprint of the positivist heritage . . . [and] the influence of positivism has been persuasive not only in letter but also in spirit'. Post-positivism, however, is having a significant impact on current public policy research.

Ontologically, positivism adheres to the possibility of objective interpretations of reality through observation, scientific method and 'laws' of cause and effect. Post-positivism, whilst acknowledging that reality exists, argues that this reality cannot be fully understood or explained by observation alone, and that theory is needed to interpret the potential diversity of cause and effects (Boyd, 1991; Guba, 1990). This does not mean that post-positivist policy analysis has ended up in a hopeless dead end of relativism. On the contrary, post-positivist policy analysts have been engaged in 'reasoned critique of their normative assumptions and experiences' and social constructivism has been increasingly accepted in policy research (Fischer, 1993: 334).

Epistemologically, positivism's emphasis on neutrality of research implies that policy analysts are assumed to be 'detached, neutral observers of facts' who carry out value-free inquiry. Post-positivists question this assumption and assert that policy analysts' values mediate inquiry. In fact, post-positivism returns the analyst to the 'human world as an active participant' (Torgerson, 1986: 36, 40).

With regard to how knowledge is searched for, the positivists' claim about the superiority of empirical derivation and testing of knowledge has been challenged.

There is now a more probing and exploratory approach to searching for knowledge, using qualitative and interactive methods. For policy analysis and evaluation, it implies a much broader and flexible framework of acquiring and using knowledge.

For policy and evaluative research, this development has had significant impact. The domination of one single methodology whereby all performance can be measured in an objective manner has been increasingly called into question. Since the 1970s several alternative methodologies have been suggested with increasing uncertainty and doubts arising about the application of these methodologies (Palumbo and Nachmias, 1983). Issues concerning power relations in society, language and discourse, social constructivism and participatory orientation have become prominent in public policy research. This development has resulted in the situation that Henkel (1991) describes with regard to the increasing gap between evaluation research and practice.

Evaluation Research from within a Policy Programme Perspective

There is no comprehensive picture available of all the evaluation methods that have been developed from within this perspective. They have been developed within a wide range of policy fields, e.g. education, health, social welfare, etc. The development has not been uniform because of different premises, frames of references and theoretical assumptions. Several reviews, however, have sought to show that the evolution of evaluation research can be described in terms of three or four generations of evaluation (e.g. Guba and Lincoln, 1989; Vedung, 1991).

The first generation of evaluation, the 'measurement generation', corresponded to the period when extensive use of various types of measurements was made in order to evaluate school-children's performance determining their progress and measuring the actual impact of resources that were used in schools. Thus, in the early policy-programme evaluations, sound measurements were considered to be crucial. Such measurement methods came to be used in the recruitment of military personnel and in time and motion studies and piecework wage rates in manufacturing industries (Guba and Lincoln, 1989: 22-6).

The use of measurement methods persists today in the field of education, especially in school and college admission procedures. However, there was a gradual realization among researchers as well as practitioners that these measurement methods alone were inadequate. Performance had to be related to goals within any given policy area in order to relate resource input to desired outcomes (Madaus et al., 1983).

The second generation of evaluation methods, prevalent between the 1930s and 1950s, involved supplementing measurements with an account of goal achievement. These came to be used in several policy areas besides education including urban and regional development.

A difficulty existed in relation to defining goals because of the reluctance on the part of decision makers to state any more than 'a long list of pious and partly

incompatible platitudes' (Weiss, 1972: 25). Additionally there was an unwillingness to commit resources for long-term development, e.g. during the rapid growth of the welfare state. So it became all the more apparent that there was a need to judge development with the help of external standards. But if scientific evaluation was to be carried out, these standards had to be value-free (Guba and Lincoln, 1989: 30).

The demand for judgement in evaluation led to a large number of evaluations incorporating external standards to varying degrees. Scriven's Goal-free Model can be regarded as a representative of this generation of evaluation (Scriven, 1973). Scriven argues that judgement is at the heart of evaluation and describes how it could be incorporated in an objective, analytic framework. By neglecting goals, the evaluator can freely focus on a programme's total effects including often-neglected side effects.

The third generation of evaluation methods acquired considerable inspiration from scientific management and the methods that came into practice at the time when public agencies, under pressure from the neo-liberal ideology, increasingly reverted to market management techniques in policy making.

Lastly, the fourth generation of evaluation methods were developed as a reaction to the dominant positivist paradigm. Several approaches, including the naturalistic responsive approach (Guba and Lincoln, 1989), the multiplist model (Cook, 1985) and the design approach (Bobrow and Dryzek, 1987), take up the post-positivist challenge of interactive participatory evaluation.

These approaches in a way do not involve evaluation in the traditional sense since evaluation takes the form of a discourse between stakeholders who are directly or indirectly affected by a policy measure. It becomes a question of negotiations rather than of a search for a specific objective. The dialogue results in consensus on some issues and disagreements on others. The latter become the subject of further rounds of negotiations. Different perspectives and evaluation methods are used to enlighten that which is evaluated. Evaluation can be likened to an arena for spreading experience and perspective from one stakeholder to another. It is a learning, dialectic process during which reality is probed into rather than discovered (e.g. Khakee, 2002).

The fourth generation of evaluation cannot simply be included in the three previous generations of evaluation. It requires considerable resources in terms of time and money. Above all, it is critical of representative democracy, questions power relations that might be difficult to change and suggests a form of social inquiry within different political circumstances (Fischer, 1995).

Evaluation Research from within a Welfare Economics Perspective

The welfare economics perspective is based on the assumption that every public action should maximize the collective or societal value. The latter is an aggregation of utility of individuals making up the society. Individuals are assumed to be maximizers of their utility whatever their ethical or moral considerations are.

Söderbaum (1998) uses the degree of aggregation to distinguish between three

Evaluation 9(3)

sets of evaluation methods that have evolved since methods were first developed, in order to describe and quantify the social advantages and disadvantages, in monetary terms.

- Highly aggregated methods, e.g. cost-benefit analysis, cost-effectiveness analysis.
- Intermediate methods, e.g. Planning Balance Sheet (PBS), multi-criteria evaluation.
- Highly disaggregated methods, e.g. Positional Analysis, environmental impact assessment.

Highly Aggregated Methods

These methods aim to sum all impacts in terms of present value. An underlying implication is that there is a consensus in a society about 'specific valuation rules' (Söderbaum, 1998: 53). The best known example of this category of methods is cost-benefit analysis. Other examples are cost-effectiveness analysis and threshold analysis. Cost-benefit analysis became highly fashionable during the expansion of the welfare state and became dominant during the 1960s and 1970s.

In cost-benefit analysis the focus often is on the quantitative ratio of benefits and costs. It is essentially a monetary method even if non-monetary impacts are formally considered to be just as economic as the monetary ones. This method has been subjected to extensive criticism because of its theoretical assumptions and practical shortcomings (e.g. Peters, 1986). Besides the fundamental weakness in the maximization and aggregation assumptions, another essential argument against this method was its failure to evaluate intangibles adequately. Despite a long-standing debate about incorporating distributional effects into cost-benefit calculations, the weighting systems that have so far been proposed have not been easy to implement in practice. This is remarkable especially because the Pareto-criterion based compensation principle is basic for this method.

Intermediate Methods

These methods retain 'the idea of a single objective function and optimization' but emphasize that analysis should more explicitly consider various parties' interests. Lichfield's PBS (e.g. Lichfield et al., 1975) is a well-known example in this category of methods. It recommends clear separation between different groups of beneficiaries and non-beneficiaries and the need of qualitative indicators in order to capture indispensable effects. It also recognizes the ethical issues involved in community evaluation beside utility maximization. Lichfield's Community Impact Analysis (Lichfield, 1996) is a further improvement of the PBS and is a significant move towards a highly disaggregated method. However, it does not reject the idea of a consensus about one idea of efficiency concerning public issues.

Multi-criteria evaluation methods also belong to the intermediate category. These methods combine the assessment of the performance of alternative actions with the assignment of relative priorities to goals and criteria and observe their

respective effects on the resulting decision (e.g. Voogd, 1983). They enable an analysis of different alternatives in the face of heterogeneous judgement criteria.

Intermediate methods have been applied with varying success but their inability to pay sufficient attention to the conflicting values of individuals has been increasingly criticized. The latter may be guided to a considerable extent by egoistic motives but ethical, moral and ideological values also play an increasingly important role, e.g. in the current debate on global and local environment issues. The plurality of values requires mutual capacity of learning and understanding and evaluation methods prepared by experts for professional decision makers become less relevant for this purpose.

Highly Disaggregated Methods

Methods in this category aim to elicit consensus in resolving conflicting claims inherent in many public issues, e.g. those involving ecological sustainability and economic development. Since the aim is not to capture a single collective value or to reduce effects to a simplified scheme, these methods are multi-dimensional and flexible. The design of the methods adapts to the changing context and not only are the results important but also the way of arriving at them. The methods combine inductive and deductive analysis and make use of quantitative as well as qualitative information.

Positional Analysis interprets views and opinions about the issue under evaluation of all involved parties with the possibility of modifying or reconsidering values when faced with 'possible valuation standpoints'. It visualizes 'conflicts of interest' and clarifies how activities will be influenced in choosing alternative paths of action. The method makes possible an open discussion of ideology (Söderbaum, 1998: 64–9). Another example of disaggregated method is a version of environmental impact assessment. In this case the reductionist approach, which arrives at an existence value of natural resources involving willingness to pay, is replaced by a form of communicative practice that combines the assessment of alternative ways of preserving natural resources with the assignment of relative priorities to various community interests.

Evaluation Research from within a Planning Theory Perspective

From within this perspective two distinct paradigms, consisting of clear and distinct theoretical and empirical propositions, have determined planning theory research; these are rational planning and communicative planning. The two planning theories are both descriptive and normative. They not only explain the nature of planning and the processes thus involved. They also guide various processes including evaluation (Lichfield et al., 1975, 1998).

There are several theoretical phenomena that were either developed as a response to the rational planning model or as ideas leading to the communicative planning theory. To the former belong incrementalism, implementation, advocacy and strategic planning, whereas transactive planning and negotiative planning belong to the latter. In this section evaluation research will only be

discussed in relation to the two paradigms. Readers interested in evaluation research in relation to the above-mentioned theoretical propositions are referred to Khakee (1998).

Rational Planning

Rational planning, which has dominated planning research for more than half a century, is based on instrumental rationality, whereby decision makers decide on goals and put questions about policy measures to professional planners and other experts who then formulate alternative plan proposals. Instrumental rationality implies that the most favourable relationship between goal achievement and resource use is obtained. This requires that goals are carefully specified and that goal achievement implies the minimization of expenditure or use of resources.

Evaluation within rational planning corresponds to 'optimization'. In practice it is extremely difficult, if not impossible, to identify an optimal plan or programme. Problems are poorly defined, many goals can only be formulated qualitatively, relationships between goals and means are poor due to value uncertainty and scarcity of knowledge, and goal formulation is not exclusively an analytical process – it contains a great deal of politics.

Advocates of the rational planning approach nevertheless contend that despite these limitations, evaluation should try to emulate optimization procedure as far as possible. In this way one can expose all the assumptions, new knowledge can be generated, a better understanding can be achieved and the whole process can look like optimization but with certain restrictions. Herbert Simon (1976) describes such optimization as 'satisficing' (e.g. Faludi, 1987).

Communicative Planning

Communicative planning, which is a much more recently developed approach, is based on communicative rationality. To a large extent planning is an interactive communicative activity, where information is presented in many different ways but with a varying degree of distortion. The latter is not coincidental and unpremeditated. Distortion is often systematic, structural and institutional and depends on the economic and political structural order in a society. A dialogue, based on communicative rationality, is the only way to combat distortion (Forester, 1989). Communicative rationality is characterized by comprehensibility, integrity, legitimacy and truthfulness. Planners can use these criteria in order to contribute towards a progressive planning practice that at the same time challenges the power that hinders such planning (see Healey [1997] for further discussion about various theoretical propositions about communicative planning).

Communicative planning emphasizes both interaction and iteration, which take place in an extensive institutional context, and where the aim is to obtain commitment and consensus among all the stakeholders. Therefore the central issues of evaluation are how best to organize an inclusive discourse, to promote a learning process which is emancipatory and expedites progress, and to emulate political, social and intellectual capital. A central aspect of evaluation is to focus on both the quality of the planning process and the programme of actions.

Evaluation is thus a question not only of effectiveness and legitimacy but also of integrity and mutual understanding. Evaluation itself becomes a form of interactive discourse where all those involved can explain their values, problems and concerns. This results in a set of recommendations and value judgements, but those problems and issues for which no mutual consensus is available are then part of the subsequent discursive process.

Common Characteristics in the Evolution of Evaluation Research

Evaluation, from within all three perspectives described above, is becoming part of a broad political process in which the premises of the traditional representative democracy are increasingly challenged. The development of evaluation research within these perspectives is summarized in Table 1.

The following outlines key characteristics in the evolution. Firstly, evaluation has become much more comprehensive. This is not only a question of knowledge used to rank different plans or programmes but also of values. The fourth generation of evaluation or disaggregated welfare economics evaluation or communicative planning evaluation does not adhere to any reductionist principle but explicitly recognizes value pluralism, value conflicts and reassessment of values.

The new evaluation methods distinguish between nominal and real clients. Nominal clients may be politicians or government officials or others with an official mandate, but the real clients are all stakeholders – beneficiaries as well as non-beneficiaries. Mapping of these stakeholders and efforts to include them in the interactive process are an important element of the evaluation process.

The role of politics in specifying indicators and interpreting their meaning has changed. The effectiveness requirement has been supplemented by other important requirements: justice, legitimacy, mutual understanding, integration of professional and experiential knowledge and democratic pluralism. The perception that evaluation is carried out with the explicit aim of finding an instrumental means–ends relationship is superseded by evaluation becoming part of the policy process.

Evaluation's expanded domain shows the necessity to supplement quantitative effectiveness-based methods with qualitative analysis. Models that facilitate our understanding of what happens during the policy process have replaced models based on some mechanical input–output principles. This is because it is no longer

Table 1. Evolution within Three Evaluative Research Perspectives

<i>Policy programme perspective</i>	<i>Welfare economics perspective</i>	<i>Planning theory perspective</i>
Measurement	Highly aggregated	Rational paradigm
Description	Intermediate	Communicative paradigm
Judgement	Highly disaggregated	
Responsive constructivism		

accepted that the output of a policy process is limited to policy proposals and plans. Interactive learning, institutional capital and stakeholder inclusion have come into focus in evaluation.

Epistemologically the development implies that the assumption about value-free knowledge is replaced by an acceptance that knowledge is politically influenced, and that all phases of policy making need both professional and experiential knowledge (see Khakee et al. [2000] for further discussion of this issue).

Another important characteristic of the evolution is that it is inadequate and incomplete to regard evaluation in terms of organization theory (for example, incompleteness in information, lack of consensus on policy goals, and so on). It has become necessary to adapt a critical theory perspective in order to understand the importance of the uneven division of power, ambiguity in value systems and defectiveness in social structure (e.g. Fischer, 1995).

Evaluation Practice in Sweden

Åberg (1997) argues that despite a tremendous increase since the 1980s in evaluative activities within central and local governments in Sweden, there is no comprehensive picture of these activities. Moreover, very few evaluation reports account for methodologies applied. In his survey of 19 evaluations in public health and hospital services, Åberg states that methods were predominantly aggregative based on the assumption about policy makers' economic rationality. Effectiveness and efficient management were key premises whenever methods were described.

A similar picture emerges in the meta-evaluation of labour market policy evaluation (Nyberg, 2000). In the evaluation studies carried out in the 1980s and 1990s, highly aggregated methods were used in order to assess income effects and the macro-economic impact of labour market education. Statistical methods dominated and whenever interview analysis was applied, the aim was to obtain quantifiable information about wages, employment, etc. (see the Institute of Labour Policy Evaluation website: www.ifau.se). In all current reports, quantitative surveys dominate.

Performance audit is closely related to evaluation, though they are carried out under different conditions. In a comparative survey of performance audit in five western European countries including Sweden, Pollitt and his colleagues show that economy, efficiency and effectiveness are central criteria in such audit. So much so that the General Director of the Swedish National Audit Office, together with a colleague, argued: 'A one-sided interest in productivity may result in the ever-better performance of the wrong tasks' (Ahlenius and Jonsson, 1995: 7; Pollitt et al., 1999: 90).

An examination of methods used in audit by the Swedish National Audit Office shows that analyses of documents, existing data and interviews are the usual methods employed. Consultation with third parties, in the form of seminars and hearings, are limited to experts, researchers and government officials (Pollitt et al., 1999: 129–37). The use of new methods has been 'incremental' rather than 'revolutionary' for several reasons. The prime objective of performance audit is

to satisfy parliamentarians and the government. The absence of much pressure from that direction may have acted as a discouragement to innovate. Time and economic constraints also affect the range of methods that could be employed. All this implies that more participative evidence collection has been quite rare (Pollitt et al., 1999: 137, 143).

Forss and Carlsson (1997) reviewed 273 evaluation reports about Swedish development aid to less developed countries. The authors are primarily interested in finding out whether these evaluations made use of adequate and structured methods. However, their study also shows that the major aim in these evaluations was open-ended interviews with project staff and beneficiaries to uncover the effectiveness of the development aid.

In a special issue of *Nordisk Administrativ Tidsskrift* [*Nordic Journal of Public Administration*] a survey of evaluative practice in Sweden, Norway, Finland and Denmark shows that early evaluation was influenced by the American programme-budgeting practice. More recently New Public Management inspired methods, aiming at improving input and output efficiency, characterize Swedish evaluative practice as well as that of other Nordic countries (*Nordisk Administrativ Tidsskrift*, 2000).

Finally, in a case study of the Swedish Medical Information System, Hanberger (2001) shows that the leadership and programme managers' major interest in letting Umeå Centre for Evaluation Research carry out an evaluation was for their multimedia concept to gain legitimacy. Every attempt by the evaluators to promote interactive learning and empower citizens failed, leading the author to conclude that there is a gap between technocrat and citizen rationality.

For reasons stated above, our brief survey of Swedish evaluative practice is not comprehensive. Nevertheless, there is good evidence that rational, quantitative and aggregated methods characterize the evaluative practice.

Implications of the Emerging Gap

In this article I have described the growing convergence in evaluation research towards stakeholder-oriented, communicative, disaggregated and multi-dimensional methods. On the other hand, public agencies still demand of their policy evaluators quantitative, aggregated, (often uni-dimensional) expert products. Although some Swedish public agencies are willing to use new methods and approaches that include public participation, the real aim is to facilitate legitimization and implementation (Hanberger, 2001). Thus the increasing gap between evaluation research and evaluation practice poses some major challenges to politicians, policy makers and public sector managers.

Evaluation as an interactive learning process has far-reaching consequences on organization of evaluation process, choice of methods, the role of evaluator, requirement of resources and the use of evaluation results.

Even if it is not a question of fully adopting a responsive constructivist or a highly disaggregated evaluation approach, audit institutions as well as evaluation agencies in the public sector would need an open framework for evaluation. This means that those commissioning evaluation should not exercise complete control

over what questions evaluations should pursue, how information should be collected and interpreted and to whom the findings should be disseminated. As Guba and Lincoln (1989: 47) aptly express, preoccupation with exercising control implies that 'we fail to empower the very people who we are putatively trying to serve'. Such an open framework also implies that the evaluator is prepared to take independent action if new policy issues arise in the course of evaluation. From the very beginning, evaluators must have a clear picture of the democratic premises of their commission and should be prepared to represent those whose concerns and issues are not otherwise taken into consideration.

The current evaluative research points to the need of accommodating value pluralism and rejecting the preoccupation with so-called objective reality. For commissioners of evaluation studies as well as evaluators, one of the major issues in evaluation should be negotiations over value differences. Rejecting the idea of value-free evaluation requires explicit means to accommodate value differences.

Epistemologically the current development in evaluative research implies reducing excessive commitments to the scientific paradigm of inquiry (see Guba and Lincoln, 1989: 35–8). For evaluation commissioners as well as evaluators this means reducing the use of quantitative measurements, becoming more sceptical to the so-called scientific evidence, paying more attention to contextual factors and taking moral responsibility for the results of the evaluation.

Contemporary evaluation research's orientation towards social inquiry and all that it implies in terms of questioning the power structure in society make it less palpable for public agencies and political establishment. To echo Henkel's concern about the current widening gap between evaluative research and practice, the result of the gap may be that politicians, policy makers and public managers will be inclined to ignore the current development in evaluation research in the hope that it is short-lived and they will carry on their policy analysis in a 'business as usual' manner.

Notes

I would like to thank the two anonymous referees and the editor of this journal for many useful suggestions to improve the final version of the article. Anders Hanberger has helped me to collect works on the Swedish evaluative practice. The final version of the article was prepared at the Barbanente family's country estate at Cozze; I send my heartfelt thanks to the Barbanentes for their kind hospitality.

1. An earlier version of this article was presented at a symposium arranged by University College London in order to commemorate the 85th birthday of Professor Nathaniel (Nat) Lichfield. I would like to dedicate this article to Nat. No one has been such an important source of inspiration in the field of evaluation in planning as Nat has been for scholars and practitioners in many countries.

References

- Åberg, J.-O. (1997) 'Det rationell och det legitima. En studie av utvärderingars teori och praktik [The Rational and the Legitimate: A Study of the Theory and

Khakee: The Emerging Gap between Evaluation Research and Practice

- Practice of Evaluations', PhD Thesis. Göteborg: Department of Sociology, Göteborg University.
- Ahlenius, I.-B. and G. Jonsson (1995) 'A New Strategy for Performance Audits of Social Security Systems', *International Journal of Accounting* 4(1): 7-9.
- Bobrow, D. B. and J. S. Dryzek (1987) *Policy Analysis by Design*. Pittsburgh, PA: University of Pittsburgh Press.
- Boyd, R. (1991) 'Confirmation, Semantics, and the Interpretation of Scientific Theories', in R. Boyd, P. Gasper and J. D. Trout (eds) *The Philosophy of Science*. Cambridge, MA: The MIT Press.
- Cook, T. P. (1985) 'Postpositivist Critical Multiplism', in R. L. Shortland and M. M. Mark (eds) *Social Science and Social Policy*, pp. 129-46. Newbury Park, CA: Sage.
- Faludi, A. (1987) *A Decision-centred View of Environmental Planning*. Oxford: Pergamon Press.
- Fischer, F. (1993) 'Restructuring Policy Analysis: A Postpositivist Perspective', *Policy Sciences* 25(4): 333-9.
- Fischer, F. (1995) *Evaluating Public Policy*. Chicago, IL: Nelson-Hall Publishers.
- Forester, J. (1989) *Planning in the Face of Power*. Berkeley, CA: University of California Press.
- Forss, K. and J. Carlsson (1997) 'The Quest for Quality - Or Can Evaluation Findings Be Trusted?', *Evaluation* 3(4): 481-501.
- Guba, E. G. (1990) *The Paradigm Dialog*. Newbury Park, CA: Sage.
- Guba, E. G. and Y. S. Lincoln (1989) *Fourth Generation Evaluation*. Newbury Park, CA: Sage.
- Hanberger, A. (2001) 'What is the Policy Problem? Methodological Challenges in Policy Evaluation', *Evaluation* 7(1): 45-62.
- Healey, P. (1997) *Collaborative Planning*. London: Macmillan.
- Henkel, M. (1991) *Government, Evaluation and Change*. London: Jessica Kingsley.
- Khakee, A. (1998) 'Evaluation and Planning: Inseparable Concepts', *Town Planning Review* 69(4): 359-74.
- Khakee, A. (2002) 'Assessing Institutional Capital Building in a Local Agenda 21 Process in Göteborg', *Planning Theory and Practice* 3(1): 53-68.
- Khakee, A., A. Barbanente and D. Borri (2000) 'Expert and Experiential Knowledge in Planning', *Journal of the Operational Research Society* 51(7): 776-88.
- Lichfield, N. (1996) *Community Impact Evaluation*. London: UCL Press.
- Lichfield, N., P. Kettle and M. Whitbread (1975) *Evaluation in the Planning Process*. Oxford: Pergamon Press.
- Lichfield, N., A. Barbanente, D. Borri, A. Khakee and A. Prat, eds (1998) *Evaluation in Planning. Facing the Challenge of Complexity*. Dordrecht: Kluwer.
- Madaus, G. F., M. Scriven and D. L. Stufflebeam, eds (1983) *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*. The Hague: Kluwer-Nijhoff.
- Nordisk Administrativ Tidskrift [*Nordic Journal of Public Administration*] (2000) Special Issue on Evaluation in Nordic Countries 81(2): 107-61.
- Nyberg, E. (2000) 'Möjligheter och begränsningar för arbetsmarknadsutbildning vid universitet och högskolor' [Possibilities and Limitations for Labour Market Education in Universities and University Colleges], in *En effektivare arbetsmarknadsutbildning [An Effective Labour Market Education]*, Public Commission Report Appendix 2. Stockholm: Ministry of Industry.
- Palumbo, D. J. and D. Nachmias (1983) 'The Pre-conditions for Successful Evaluation: Is There an Ideal Type?', *Policy Sciences* 16(1): 67-79.
- Peters, B. G. (1986) *Public Policy in America: Process and Performance*. London: Macmillan.

Evaluation 9(3)

- Pollitt, C., X. Girre, J. Lonsdale, R. Mul, H. Summa and M. Waerness (1999) *Performance or Compliance? Performance Audit and Public Management in Five Countries*. Oxford: Oxford University Press.
- Scriven, M. (1973) 'Goal-free Evaluation', in E. R. House (ed.) *School Evaluation: The Politics and Process*, pp. 319–28. Berkeley, CA: McCutchan.
- Simon, H. A. (1976) *Administrative Behaviour*, 3rd edn. New York: Free Press.
- Söderbaum, P. (1998) 'Economics and Ecological Sustainability. An Actor-network Approach to Evaluation', in N. Lichfield, A. Barbanente, D. Borri, A. Khakee and A. Prat (eds) *Evaluation in Planning: Facing the Challenge of Complexity*, pp. 51–72. Dordrecht: Kluwer.
- Torgerson, D. (1986) 'Between Knowledge and Politics: Three Faces of Policy Analysis', *Policy Sciences* 19(1): 33–59.
- Vedung, E. (1991) *Utvärdering i politik och förvaltning* [Evaluation in Government and Administration]. Lund: Studentlitteratur.
- Voogd, H. (1983) *Multicriteria Evaluation for Urban and Regional Planning*. London: Pion.
- Weiss, C. H. (1972) *Evaluation Research Methods of Assessing Program Effectiveness*. Englewood Cliff, NJ: Prentice Hall.

ABDUL KHAKEE is Professor in Political Science in the Department of Social Sciences, Örebro University and in the Department of Political Science, Umeå University, Sweden. He is also a member of the board of the Umeå Centre for Evaluation Research and the co-ordinator of the international network in Evaluation in Planning. Please address correspondence to: Umeå Centre for Evaluation Research, Umeå University, SE-901 87 Umeå, Sweden.
[email: abdul.khakee@pol.umu.se]