

Nr. 4

Hans-Ulrich Derlien

Programme Evaluation as an Approach to the
Management of Public Policies in the
Federal Republic of Germany

Paper prepared for the Symposium on Management
of Public Policies, Centre d'Etudes Supérieures
du Management Public, Paris, 27th/28th October 1980

1. Programme Evaluation as a New Management Device

Programme Evaluation (PE) is a rather new device in the management of public policies. It is usually defined as the systematic (scientific methods employing) investigation of the effectiveness and the actual (foreseen and unforeseen, intended and unintended) effects of public intervention programmes.¹⁾

1.1 New Functions of the State and Deficiencies of Traditional Control Mechanisms

PE can be regarded as an instrumental - as opposed to social, political and juridical - feedback mechanism, which public administration uses to judge its performance. How can this innovation in the management of public policies be explained?

In the Federal Republic of Germany as well as in other countries, in particular in the United States of America, PE is closely linked to the emerging planning functions of the state.²⁾ In the FRG until about 1965 public planning was neither acknowledged nor accepted for ideological reasons. Only after the economic depression of 1965 and the formation of a grand coalition between the Christian Democrats (in power since 1949) and the Social Democrats in 1966 public planning conceived as active intervention in economic and social processes and structures became gradually accepted as one of the functions of the state. Concomitantly the quality of meta-policy-making changed as well: planning staffs, social indicators, forecasting, cost-benefit-analyses and evaluation were gradually established in Bonn.³⁾

The same development of planning and evaluation can be observed in the United States: after the launching of the "great society programmes" of the Kennedy/Johnson aera in the "war against poverty" and the initiation of the meanwhile historical Planning-Programming-Budgeting system PE mushroomed.⁴⁾ Obviously, there is a strain for structural con-

sistency of planning and PE as the adequate control mechanism.⁵⁾

In the liberalist conception of the state public policy is rather re-active; emphasis is on self-regulation of economy and of society, state functions were interpreted as merely subsidiary. The quality of public policy is viewed rather in terms of legacy and costs than of effectiveness. Thus, the corresponding control mechanisms are parliamentary and juridical controls, external accounting stressing orderliness of budgetary behaviour, incidental feedbacks from interest groups, the level of public expenditures, or at best some output measures (road or flats built e.g.). These mechanisms were bound to be deficient when macro-economic, target oriented interventions and social reform policies were introduced, as they do not provide information if programme goals have been met or not, whether positive or negative side effects or spill-overs occurred, and eventually what the causes of programme failures might have been and how programmes could possibly be amended.

Thus, there are first of all informational reasons for establishing PE as a regular function in the process of managing public affairs. Particularly with innovative reform programmes or even social experiments uncertainty about their effects and effectiveness increases. Uncertainty due to a change from incremental towards innovative and comprehensive, from medium to long-term planning is further enhanced, if the programmes cannot be logically derived from and the effects predicted on the ground of a valid scientific theory of the particular policy area. The less prediction of results can be accomplished, the less cost-benefit-analysis, conceivable as an ex-ante-control, can be employed and the more there is a need for ex-post-evaluation. We, thus, can assume that PE will be most needed in "soft" policy areas: social, educational and welfare policy, whereas, for instance, technical projects or economic policy can rather rely on prediction, cost-benefit-analysis, or, generally speaking, on feedforward mechanisms.

The development of PE subsequent to increased planning activities is, secondly, brought about by an increased need for re-allocating financial capacities from ineffective to effective programmes. As reform policies normally face financial restrictions, there is an additional need to legitimate new expenditures and/or the curbing of old programmes by proving them to be effective or ineffective, respectively.

1.2 The Institutionalization of Programme Evaluation in the Federal Bureaucracy

As in the USA and Great Britain⁶⁾ PE became institutionalized in the FRG only with a certain time lag after the planning system had evolved and new programmes were being implemented. The spreading of the concept can be traced, beginning in 1969. There are basically three forms PE takes on as a regular process:

- institutionalization of special evaluation units in individual ministries,
- institutionalization of PE with specific legislated programmes,
- institutionalization of PE with social experiments.⁷⁾

1.2.1 Evaluation units

PE involves socio-economic research and therefore requires know-how that is not easily available in the traditionally juridically trained federal bureaucracy. Therefore, some of the departments, which relatively often carry out evaluations, established special units for this task. This is not to mean, that the evaluation proper is carried through by administrative personnel; on the contrary, it is regularly accomplished by commissioned research, but one needs specialists for discussing research designs, providing data, keeping contact to the external research team, to comment on the research report, and, last not least, to contract researchers.

In the case of the FRG an additional cause contributed to the evolution of evaluation units in some of the departments: since in the 1969 budget reform cost-benefit-analyses were prescribed and could be required by the ministry of finance in preparing future budgets,⁸⁾ especially those ministries with big budget shares containing large investment programmes were particularly exposed to pressure and inclined to establish these new units:

- The Ministry of Agriculture with its huge subsidy programmes in 1973 changed its organisation structure according to programmes and subsequently institutionalised a section in the planning unit in order to specialise on ex-ante and ex-post analyses. PE of less sophisticated nature are regularly required by the units being in charge of a programme to justify budget proposals within the department of agriculture.⁹⁾
- The Ministry of Transport, one of the big investors as well, initially created a section for cost-benefit-analyses in its planning unit in 1970, and extended its task to incorporate ex-post evaluations in 1978. Furthermore, in one of the dependent agencies of the ministry concerned with federal roads there were also specialists occupied with the evaluation of a number of safety regulations, particularly of experiments with speed limitations (1971).
- The Ministry of Developmental Aid was the first ministry to establish an evaluation unit to survey its developmental projects - not so much for budgetary reasons, but rather because of particular control needs and because of the longer tradition evaluation has internationally in this policy area. To a certain extent it was also parliamentary pressure on the ministry that led it to have this unit in 1970.

- Also due to prolonged parliamentary pressure (1968-1974) the Ministry of Technology instituted an evaluation unit, which, however, is less engaged in impact research, but rather in reviewing research institutions and the implementation of funded projects.
- Information programmes are to be evaluated by special staffs in the Federal Agency for Press Relations (1968) and in the Federal Agency for Health Information (1974).

1.2.2 Evaluation of Specific Legislated Programmes

As might be obvious from what has been said so far, parliament played an important role in adapting the control structure to the functional requirements of planning. Uncertainty about the substantive impact and financial implications of new programmes are the predominant motives for claiming legitimation of programmes by proving their effectiveness and efficiency. Apart from political agreements to evaluate individual programmes parliament after 1969 increasingly obliged the federal government to report on implementation and impacts of programmes by instituting evaluations in the very programme legislation. A similar trend can be observed in the USA, too.¹⁰⁾

Examples are the PE of

- joint federal-state programme to improve the regional economic structure (1970),
- labour market and employment act (1969),
- law to continue the payment of wages in case of illness (1969),
- amendment of the social subsidies act (1969),
- hospital investment programme (1971),
- postgraduate grants law (1970),
- general educational grants law,
- legal abortion law (1975),
- legal restrictions for the cancellation of flat-rent-treaties (1971),
- reports on subsidies and taxes.

1.2.3 Evaluation of Experimental Policy

A special type of reform policy consists of carrying through social experiments; of course, experiments logically imply the necessity to evaluate them in order to learn from their results, before general regulations are enacted. Experiments have initially been carried out in house construction in the mid 1960s;¹¹⁾ later on the idea of experimental policy has spread to other policy areas, in particular in education and health policy.¹²⁾ These are obviously "soft" policies to which PE is most suitable; but it was also for a political reason that the federal government started to carry out experiments under the label "model programmes" in these areas: legislation in education and health is normally a task constitutionally assigned to the individual states. Federal government is merely authorised to regulate basic questions in so called frame legislations. Thus, launching experiments was a way to practice reform policy on "foreign territory" and, secondly, to do it without legislation, but by simple administrative (executive) agreements between federal and individual state governments; the justification being that experiences should be accumulated in order to clarify basic questions which are within the jurisdiction of the federal government.¹³⁾

Examples for experiments to be evaluated are

- model stations to give advice before legal abortion,
- as an alternative to nursing schools having small children educated within other families during the day,
- comprehensive schools as an alternative to the traditional three-stage-school system,
- practicing an alternative to traditional juridical university training.

1.2.4 Distribution of Functions in the Evaluation Process

The systematic ex-post evaluation of experimental and reform policies is obviously a task which can hardly be accomplished by traditional external control institutions (parliament, general accounting office) nor by the ministerial bureaucracy. Even those ministries disposing of special evaluation units in general do not carry through the actual research work, i.e. data collection and analysis. Evaluation research is basically external (commercial and academic) commissioned research.

To conclude this chapter of the paper the question should be asked, which functions in the evaluation process can be attributed to the political and administrative institutions on the federal level.¹⁴⁾

In an ideal-typical way the scenery can be depicted as follows: PE is primarily initiated by parliament, possibly the general accounting office, and the Chancellor's office; however, relatively seldomly by the departments in charge of a programme.

The function of the departments and their sections is rather to administer the external evaluation research and to write parliamentary reports on the basis of the evaluation studies, including normally suggestions for programme amendments.

The relative decentralisation of the evaluation function onto the departmental level can be explained structurally: first of all, the corresponding planning functions are not centralised in the FRG; the Chancellor's office rather confines itself to coordinating functions. Secondly, this decentralisation is supported by the constitution which gives the individual ministries a good deal of autonomy in initiating and executing their policies, whereas the cabinet is merely involved in programmes to be submitted to legislation and in fundamental policy questions. Departmental autonomy is particularly large in coalition governments with the ministers of one or the other fraction being in a veto position.

Centralising planning and evaluation onto the ministry of finance as an alternative would not be meaningful in the FRG,

because budgeting and planning are not integrated, but can be conceived of in terms of the traditional two-track-system.

2. Methodological Problems of Programme Evaluation

PE in order to function properly has first of all to cope with a number of methodological problems, which are quite characteristic for this type of applied research. Meanwhile there are lots of text-books on the methodology of evaluation research,¹⁵⁾ which do not fall short of taking account of these specific problems, and practitioners tend to speak of the iron law of evaluation flaws.¹⁶⁾ I, therefore, shall confine myself to the most important methodological issues one has to come to grips with when doing this type of policy analysis:

- specification of impact indicators (operationalisation problem)
- measuring, i.e. collecting data (data problem)
- explaining the effects measured by referring to properties of the programme (attribution problem).

2.1 Specifying Indicators of Programme Effectiveness

Logically the effectiveness of a programme should be measured in terms of its goal achievement. However, some programmes do not have explicit goal descriptions; others do have goal descriptions, but too ambitious to be ever accomplished, too ambiguously formulated or even informationally empty ones.

There are often political reasons for this lack of precision in formulating programme goals:

- formulations referring to general welfare and happiness of the majority are likely to secure political support, as the electorate can interpret them arbitrarily and believe that individual expectations will be met;¹⁷⁾
- the more ambiguous, the less operational goal descriptions are, the easier is political consensus building;¹⁸⁾
- some programmes might even be launched for symbolic purposes with no specific effects to be achieved at all;¹⁹⁾

- as a consequence of political bargaining goals are formulated in an abstract way in order to disguise basic political controversies, the solution of which is, sometimes intentionally, deferred into the implementation process;
- because of these factual functions of goal formulations they are regularly not formulated at the beginning, but rather at the end of the policy process.

Furthermore, even if it is possible to derive indicators of goal achievement, these do only measure the positive and intended effects of a programme. Uncertainty being one of the motives for PE, unforeseen positive and negative impacts, however, should be investigated, as well.

Facing this situation the researcher is bound to be selective in measuring the impact of a programme. He is also likely to stimulate political conflicts when trying to specify what might have been (positively or negatively) expected by programme proponents and opponents. Case studies show, that evaluation research that sticks to goal formulations and tries to operationalise them without taking into respect the various and often conflicting political expectations, or that is too selective in investigating unforeseen consequences, does not succeed in being politically accepted afterwards; is blamed not to have measured what one had intended to achieve with the programme under scrutiny; or to be affirmative by neglecting negative side-effects which have emerged since the programme had been implemented.

2.2 Collecting Data

Even if the evaluator succeeds in satisfactorily operationalising programme goals, this does not mean that he will manage to collect the required data. Official statistics and administrative implementation data are regularly too crude to measure socio-economic impacts. Only when there are

evaluation units in the departments that participate in programme development and thereby safeguard that at an early point official statistical surveys are initiated and implementing agencies instructed which data to collect, the researcher can use these data.

If these provisions are not made, especially when old programmes are to be evaluated, the evaluator often has to use proxy measures, for instance statistical data which can only be used as secondary indicators or which have to be re-interpreted. Or he has to rely on special surveys which primarily measure opinions and attitudes but not behavioural changes, for instance.

Last not least, a lack of evaluation planning almost always leads to a failure in measuring the state of affairs before the particular programme came into operation. This, however, means that it is impossible to state exactly the change that has been brought about by the programme.

2.3 The Attribution Problem

This leads us to the question how impacts, however measured, can be causally attributed to the forces of a particular programme or individual of its elements. There are always rival hypotheses for explaining the impact measured, be it the way the programme was being implemented, be it the working of external, uncontrolled factors in the environment, or be it the assumption that change incurred by chance.

Obviously a sufficiently valid explanation can only be given if the programme was designed experimentally with

- measurement of the goal variables before and after programme affection,
- with an experimental group, exposed to the programme, and a control group, which is not affected by the programme.²⁰⁾

These conditions are seldomly met, and even with experimental programmes it is difficult to control all potential external factors in a field experiment.

At least in Germany many programmes are legislated and, therefore, cannot be withheld from a control group. Secondly, social experiments are mostly merely quasi-experiments:

- the geographical location often is the result of a political decision rather than of systematic thinking striving for representativeness;
- the programme variables are not kept constant, as for instance in school experiments teachers would not wait to improve situations until the research results are published;
- therefore, systematic variation of programme elements in order to learn about the relative effectiveness and efficiency of alternative programme designs is hardly accomplishable;
- there is no systematic sample of those participating for instance in a school experiment. Hawthorne effects, i.e. self-fulfilling prophecies, are likely when most highly motivated teachers and children as well as parents in favour of comprehensive school do recruit themselves into the experiment.²¹⁾

3. Political Obstacles to Evaluation

When outlining the institutionalisation of the evaluation function I emphasized the role parliament played in this process. This fact can be taken as a hint that the federal bureaucracy is not particularly fond of PE.

3.1 Resistance against Evaluation

Evaluation is normally regarded by practitioners as a control procedure, and this is in accordance with the rationale of this management tool. But as in the FRG PE often is termed in the traditional control terminology (Erfolgskontrolle, Inspection), it is associated with person-oriented performance measurements or juridical controls of individual decisions. Programme failures, thus, tend to be personalized and interpreted by administrators in terms of guilt and responsibility instead of cause and effect. Furthermore, if the evaluation

function is not located in special units, initiating the evaluation of the programme he is in charge of would mean to a minister or section head not only the possibility of incurring negative political or administrative sanctions, but also to be prepared for cognitive dissonances if the programme would not produce the expected (and promised!) results. As Wildavsky has put it: the proper evaluator should be a kind of official eunuch.²²⁾

Unless legally instituted it, therefore, needs considerable public criticism to induce an ad-hoc-evaluation. Resistance against PE will be the stronger the older the programme is, as vested interests of programme beneficiaries and self-confidence of the administration tend to neglect or doubt all criticisms.

3.2 Political Reactions to Evaluation Results

Particularly with externally ad-hoc initiated evaluations the likelihood of negative results is relatively great, as there is always some truth in political feedback mechanisms. Past experience of this sort seems to have conditioned a basically defensive attitude the more.

Usually the research process is already negatively affected by opposition and defense of those in charge of a programme, the result being increased methodological difficulties for the external evaluators as cooperation in giving access to files and data may be lacking.

Negative evaluation results constitute a threat to the legitimacy of an ongoing programme, the political success of a minister, and the administrative career of the official. In this situation the normal reaction is to question the validity of the research results. Be it that the methodological weaknesses are evident, be it that external counter-advice is asked for, the iron law of evaluation flaws offers lots of possibilities to maintain that the inefficiency of a programme has not been validly demonstrated.

The inclination to politicise evaluation results is particularly great when they imply an overall negative judgement, which would lead to the termination of the programme. These studies, which merely measure impacts without explaining them by certain programme elements or features of the implementation process, can be called macro-negative; on the opposite, micro-positive studies are those pointing at internal causes and thereby variables which can be improved. The latter type of study is much more favourably perceived by the administrator, because it gives him hints how to amend the programme and helps him to save it politically. Nothing is better than an amendment, political rationality tells us.

4. Recent Trends in Policy Analysis

This being the roughly sketched experience with PE in the FRG, there are certain new developments emerging which at the same time treat PE more carefully and broaden the inventory of policy analysis.

4.1 Evaluability Studies

It has become recognized by evaluators that the methodological problems tremendously increase, when you start to evaluate old programmes, and that research is likely to run into political struggles. Therefore, the first device is to test, if a programme is evaluable at all from a methodological point of view.²³⁾

4.2 Termination of Programmes

In order to increase the propensity of the executive branch to carry through PE, in the USA the so called "sunset legislation" was introduced recently, i.e. the automatic termination of a programme after a couple of years; it is only extended to the next period, if its effectiveness has been demonstrated by an evaluation study.²⁴⁾ Termination, therefore, can be regarded as an inbuilt mechanism to bring about PE.

4.3 Selectivity of Evaluation

PE being a threat to administration it should not be done covering all and every programme. Even with special departmental units it should be selective in substantive and in time respects. The evaluation units should plan, which programme when to evaluate, and try to reach internal agreement on this plan.

4.4 Emphasis on Implementation Process

As the possibility to take action is so important for the political reception of a study, much more emphasis has been put on implementation studies, recently.²⁵⁾ On the one hand they reveal weaknesses which can be amended in relatively short time, on the other hand this can be done quite early after the programme was launched, whereas impact evaluations necessarily can be carried out only after years of operation because results will not be observable before.

4.5 Testing Programme Practicability

Obviously the time perspective of administrators is limited. If therefore implementation studies gain importance, it is only logical to save time by simulating the implementation process. This has been done in the FRG with respect to the effects of a programme particularly with new tax regulations and changes of the health and social insurance system. Quite recently, however, the implementation process, too, was simulated in the cases of the City-planning-law (1971), Building Regulation law (1973, 1975) and the Youth Protection law (1977).²⁶⁾

5. Conclusion

Although PE was recognized as a necessary counterpart of planning and as an integral part of the management cycle in public policy, practical experience shows that sophisticated methodological devices, which are required for valid evaluation studies, can hardly be realized in practice. There is therefore a trend toward robust techniques, and more emphasis is put on implementation studies and the simulation of effects and the implementation process.

Footnotes

- 1) For a definition see, for instance, Carol H. Weiss, Evaluation Research. Methods of Assessing Program Effectiveness, Englewood Cliffs 1972, p.4.
- 2) See Hans-Ulrich Derlien, Die Erfolgskontrolle staatlicher Planung, Baden-Baden 1976, p. 82-86; Allen Schick, From Analysis to Evaluation, in: The Annals of the American Academy of Political and Social Science 1971, p. 57-71.
- 3) See Heribert Schatz, Auf der Suche nach neuen Problemlösungsstrategien: Die Entwicklung der politischen Planung auf Bundesebene, in: Renate Mayntz/Fritz Scharpf (eds.), Planungsorganisation, München 1973, p. 9-67.
- 4) See S. Pamela Horst/J.N.Nay/J.W.Scanlon/J.S.Wholey, Program Management and the Federal Evaluator, in: Public Administration Review 34 (1974), p. 300-308.
- 5) Cf. Renate Mayntz, Probleme der inneren Kontrolle in der planenden Verwaltung, in: Prakseologica 1971, p. 343-351.
- 6) See Christopher Pollitt, The Central Policy Review Staff 1970-1974, in: Public Administration 52 (1974), p. 375-392.
- 7) For details see Hans-Ulrich Derlien, op.cit., p. 52-82.
- 8) See Hans-Ulrich Derlien, Die Effizienz von Entscheidungsinstrumenten für die staatliche Ressourcenallokation, in: Hans-Christian Pfohl/Bert Rürup, Anwendungsprobleme moderner Planungs- und Entscheidungstechniken, Köln 1978, p. 311-326.
- 9) See Hans-Ulrich Derlien, Probleme des neuen Planungssystems im Bundesministerium für Ernährung, Landwirtschaft und Forsten, in: Die Verwaltung 8 (1975), p. 363-371; Horst Willer, Agrarpolitische Planung und politisch-administrative Praxis, in: Berichte über Landwirtschaft 55 (1977), p. 177-213.
- 10) See Henry W. Riecken/Robert F. Boruch (eds.), Social Experimentation, New York 1974, p. 118.
- 11) See R. Bohnsack et al., Modelleinrichtungen in der Sozialpolitik. Experimentelle Reformverfahren im Rahmen der Jugendhilfe, in: Franz X. Kaufmann (ed.), Bürgernahe Gestaltung der sozialen Umwelt, Meisenheim 1977, P. 150-193.
- 12) Riecken/Boruch (eds.), op.cit., confer in their appendix 60 abstracts of social experiments carried out by 1974.
- 13) See Hans-Ulrich Derlien, Die Erfolgskontrolle ..., op.cit., p. 114-118.
- 14) See for a more detailed analysis Hans-Ulrich Derlien, Organisatorische Aspekte der Programmevaluation, in: Verwaltung und Fortbildung 1978, p. 51-61.

- 15) See the recent publications by Clark C. Abt (ed.), The Evaluation of Social Programs, London 1977; Marvin G. Alkin/R.D. Daillak/P. White, Using Evaluations: Does Evaluation make a Difference?, London 1979; Michael Q. Patton, Utilization - Focused Evaluation, London 1978; Peter Rossi/H.E. Freeman/Sonia R. Wright, Evaluation. A Systematic Approach, London 1979. See also several volumes of Evaluation Studies Annual Review published by Sage, London and the journal Evaluation Quarterly.
- 16) See Walter Williams, Social Policy Research and Analysis, New York 1971, p. 123.
- 17) Cf. David Braybrooke/Charles E. Lindblom, A Strategy of Decision. Policy Evaluation as a Social Process, New York/London 1970 (1963).
- 18) Ibidem
- 19) See Murray Edelman, The Symbolic Uses of Politics, Urbana 1964.
- 20) Cf. Riecken/Boruch, op.cit., p. 87-116.
- 21) For a survey of threats to validity see Donald T. Campbell/J.C. Stanley, Experimental and Quasiexperimental Designs for Research, Chicago 1966; Philip A. AuClaire, Informing Social Policy: The Limits of Social Experimentation, in: Sociological Practice 2 (1977), p. 24-37.
- 22) Jeanne Nienaber/Aaron Wildavsky, The Budgeting and Evaluation of Federal Recreation Programs or Money doesn't grow on Trees, New York 1973, p. 6.
- 23) See Joseph S. Wholey, The Role of Evaluation and the Evaluator in Improving Public Programs: The Bad News, the Good News, and a Bicentennial Challenge, in: Public Administration Review 1976, p. 679-683.
- 24) See B. Adams/B. Sherman, Sunset Implementation: A Positive Partnership to make Government Work, in: Public Administration Review 1978, p. 78-81; Douglas Bothun/John C. Comer, The Politics of Termination: Concepts and Process, in: Policy Studies Journal 7 (1979), p. 540-553; Werner Hugger, Legislative Effektivitätssteigerung: Von den Grenzen der Gesetzesevaluierbarkeit zum Gesetz auf Zeit, in: Politische Vierteljahresschrift 20 (1979), p. 202-220.
- 25) See Renate Mayntz, Die Implementation politischer Programme. Theoretische Überlegungen zu einem neuen Forschungsgebiet, in: Die Verwaltung 1977, p. 51-66; eadem (ed.), Implementation politischer Programme, Königstein 1980; Helmut Wollmann (ed.), Politik im Dickicht der Bürokratie, Opladen 1980; for a recent American publication see Daniel Mazmanian/Paul A. Sabatier (eds.) Effective Policy Implementation, Lexington/Mass. 1980.
- 26) See Carl Böhret/Werner Hugger, Bessere Gesetze durch Test der Entwürfe? in: Zeitschrift für Parlamentsfragen 1979, p. 245-259; idem, Der Praxistest von Gesetzentwürfen - am Beispiel des Referentenentwurfs zum Jugendhilfegesetz, Baden-Baden 1980.