Översikter och meddelanden

Social Experimentation and Social Reform: A Review of the 'Experimenting Years' in the USA

Evaluation research, as its practitioners well know, is a prickly business. Studies rarely produce definitive evidence about programme outcomes (Booth, 1988). They frequently come up against the methodological problems created by vague goals and weak effects (Rossi and Wright, 1977). Even when positive measures of impact are obtained they usually cannot be attributed unambiguously to the programme because of the impossibility of controlling for all the extraneous variables.

These problems have led many policy-makers to voice serious misgivings about the utility and worth of much evaluative research. Sir William Pile, for example, former Permanent Secretary at the Department of Education and Science in London, has bemoaned how rare it is 'to find a piece of research that really hits the nail on the head and tells you pretty clearly what is wrong or what is happening or what should be done.' (quoted in Broadfoot and Nisbet, 1980) Similar sentiments have emerged for similar reasons in the United States. The National Research Council, for example, following a review of applied social science research, concluded that the quality of the work, on average, was 'relatively undistinguished with only modest potential for useful application' (quoted in Lynn, 1978). Such conclusions lie behind what many evaluation researchers see as a crisis of under-utilisation: research is being done but it is not being used.

One response within the evaluation community to this situation has been to press the importance of putting evaluation research onto a more scientific footing. According to this view, a key reason why so much policy-driven research has lacked application is because it is technically flawed. If evaluations are to hold their own in the political arena they must be sufficiently robust to withstand the scorn of those who are disadvantaged by their results. Only tight designs and rigorous methods will yield firm answers to the kind of questions policy-makers ask.

One of these questions frequently asked of evaluation researchers is, 'Does it work?' As almost any textbook will declare, the most rigorous approach to measuring programme effects or policy outcomes is by the use of experimental designs. The implication would seem to be that if evaluators believe greater rigour will lead to greater utility they would do well to pay more attention to the method of social experimentation. This paper sets out to explore whether social experimentation does indeed promise a solution to the problem of utility in evaluation research.

Social experimentation has not been widely used so far in policy analysis or programme evaluation in Europe. By contrast, it has a long history of use in the United States where thinking is still much influenced by Campbell's (1969) powerful vision of an 'experimenting society' in which social programmes would be devised 'using the best of science' and retained, modified or discarded on the basis of a hard-nosed analysis of their effectiveness.

In seeking to address the question of whether social experimentation does indeed offer a methodology for improving the empirical foundations of policy-making and a way forward for evaluation research, this paper will review the lessons which may be drawn from the 'experimenting years' in the USA. First, however, it is necessary to clarify the basic characteristics of the experimental method.

The Method of Social Experimentation

Social experimentation is an attempt to extend the logic of the laboratory into the real world (Shaver and Staines, 1971). It has been promoted as a 'radical new strategy of social reform' (Rivlin, 1971) aimed at resolving the fundamental dilemmas of policy-making:

- the uncertainty of knowing beforehand whether new policies or programmes will achieve their intended results;
- the difficulties of establishing afterwards whether they have been successful or not;
- the bias towards self-justification which works against the righting of wrong decisions and so increases the risks of innovation.

Experimentation was seen as a way of tackling these dilemmas of reform and innovation by the simple expedient of 'suck it and see' – but under strictly controlled conditions. These conditions were modelled on the classical design for true experimental, laboratory research.

Experimental designs are used to infer causal relationships. They seek to estimate the effects of some treatment or intervention on a target group or outcome variable. This is done by assigning subjects to two groups: an experimental group and a control group. A 'pretest' measure is taken of the dependent variable (the one which might be expected to change after treatment) in order to establish a baseline against which the outcome may be compared. The experimental group is then exposed to the treatment while the control group is not. After the treatment, a 'posttest' measure is taken in both groups using the same instruments as before. The pretests and posttests for both the experimental and control groups are finally compared to identify precisely what, if any, changes might be attributed to the treatment as measured by actual differences in the recorded outcomes between the two groups.

Experimental designs can be classified into two main types depending on how assignment to treatment occurs: randomised experiments and quasi-experiments. In randomised experi-

ments, the subjects are assigned to the experimental or control group in such a way that each person has an equal chance of being selected for either group. The purpose of this procedure is to ensure that the two groups are comparable in every respect except for the treatment they receive. Random methods of assignment mean that they will differ from each other only by chance. As long as the sample size is adequate the groups may then be considered equal and alike for the purposes of comparison. Whatever differences in outcome appear between the pretest and the posttest measures of the experimental and control groups may therefore confidently be attributed to the effects of the intervention or treatment.

Another method of achieving comparability is by matching pairs of subjects on the basis of their relevant characteristics and then assigning them to the experimental or control group by random methods. The result of matching should be the same as for random assignment enabling the effects of the treatment to be isolated.

In quasi-experimental designs the researcher is unable to regulate who receives the treatment. In other words, the assignment of subjects is not random – usually because individuals select themselves for treatment or because officials act as gatekeepers. The same basic framework of before-and-after testing, however, is maintained. There are many sorts of quasi-experimental methods but, broadly speaking, they may be divided into time series designs and nonequivalent group designs.

The time series design involves taking a series of measurements before, during and after the experimental treatment or intervention (for example, a behaviour modification programme) and, by comparing the trial and before-and-after measures (say, frequency of absenteeism from school), identifying any change or rates of change in the subjects' behaviour (truancy). The nonequivalent group design involves taking pretest and posttest measures of both a treatment group and a comparison group whose characteristics resemble but are not strictly equivalent to those of the experimental subjects. Because random selec-

tion is lacking from both these methods it cannot be supposed that any observed change in the outcome as measured is due to the treatment: rival explanations cannot properly be ruled out and therefore valid inferences about causation cannot be made (Mark and Cook, 1984).

In line with the convention adopted by Riecken and Boruch (1974), true social experiments will here be regarded as those where provision is made for the random assignment of subjects to treatment and control groups. An example is provided by the experimental NHS Nursing Homes for Elderly People project initiated in 1983 (DHSS, 1985). Quasi-experimental trials of new programmes, including those lacking controls, will be called demonstration projects. Examples here include a number of 'central initiatives' launched by the (then) DHSS including the care in the community pilot projects, the opportunities for volunteering scheme and the creation of demonstration development districts for mental illness in old age (House of Commons, 1983).

Social experimentation, however, is advanced as more than just a rigorous approach to evaluation. For many of its more vociferous advocates it also represents a new, scientific approach to policy-making (Rivlin, 1971).

According to this view, one of the chief obstacles to producing better social services is that people are pushed in advance into taking up positions for and against any proposed reforms – which are always packaged as if they were certain to be successful. When changes are carried through, politicians are obliged to stand by their efficacy and uphold the correctness of their own decisions. Administrators are debarred from admitting their failures in order to avoid embarrassing their political masters and jeopardising their own reputations and careers. A new political posture is needed – or so the argument runs – one which fosters new ideas without the excess of commitment that 'blinds us to reality testing' (Campbell, 1969). A simple change that would help to make this possible, in Campbell's view, is 'to shift from the advocacy of a

specific reform to the advocacy of the seriousness of the problem'.

Experimentation offered a framework for rejigging the policy process along these lines. By setting up a controlled trial, politicians would be able to demonstrate their determination to tackle a problem without the risks of backing an untried solution. Equally, trapped administrators would be freed from the straitjacket of past commitments and airy promises, and encouraged to look for new solutions where previous ones had failed. In this light, experimentation can be seen as a scientific version of incrementalism; trial and error under laboratory-like conditions. In reviewing its strenghts and weaknesses, this paper will focus on the practical aspects of conducting social experiments in the real world rather than on epistemological issues.

Experimentation: Paradigm or Pitfall

Weighing up the lessons of the experimenting years, and the value of an experimental strategy, Rivlin (1973) identifies a series of dilemmas that have dogged most of the work done:

- *Design dilemmas* often arose from the conflict between the desire to obtain valid, reliable results and the need to produce them quickly and economically.
- *Implementation dilemmas* arose as people learned that, while the tightly regulated experiment may produce more clear-cut answers, it may also be an unreliable predictor of what will really happen in a messy world.
- *Evaluation dilemmas* arose because of the conflicts between involvement and objectivity, and between the goals of the programme and the requirements of the experiment.
- *Timing dilemmas* arose because results were often wanted quickly while good experiments take time. Pressures to release data early, so risking false conclusions, were counterbalanced by the dangers of

seeing the analysis through but not having the findings available when they were needed.

● *Moral dilemmas* focussed on the issue of whether it is ethical to experiment with people in ways that might disadvantage them, and on the equity of deliberately creating inequalities for experimental purposes.

● *Dilemmas of confidentiality and openness* clustered around the issues of how to protect the privacy of the participants in the experiment and how much they should be told about the reasons for the experiment bearing in mind that such knowledge might influence their behaviour and bias the results.

Design Dilemmas

The design dilemmas struck at the heart of the experimental method. After all, its selling point was that policy-makers could bank their reputations on the results. If experimental designs could not be guaranteed to hold up in practice then this claim rang-hollow. In fact, experience showed there were inherent difficulties in the method which rendered it unsuitable for the evaluation of broad-aim, social action programmes (Weiss and Rein, 1970).

It proved difficult to select satisfactory criteria of success. The aims of social programmes were often only vaguely formulated, if at all; articulated in different ways by different sets of interests; and changed or were reinterpreted over time. This made it difficult to know what baseline data to collect and how to measure the changes that followed. Experimenters often responded by imposing their own narrow criteria on the programme (Guttentag, 1973), measuring what they could count and ignoring any unanticipated consequences (whose importance frequently rivalled and sometimes outweighed the initial aims). In this way, they were often unwittingly led into misrepresenting the programme's true impact.'

In field settings, the situation is essentially uncontrolled. Random assignment is not enough to control for the effects of exogenous variables. Field experiments were limited in their scale by their cost. Usually they were carried out within an institution or a local community. These institutions or communities were rarely randomly selected: they had to be willing to host the experiment and this willingness distinguished them from other places or locales. Consequently, there was no way of being sure that the results were not influenced by special characteristics attached to those sites or that the results would be the same if the experiment was repeated elsewhere or on a larger scale (Arrow, 1976). In other words, experimentation rarely allowed for confident generalisations to be made about the operation of a full-blown programme.

Even when experiments were conducted in a variety of locations, in an attempt to control for such context variables, they met with another problem: *they could not ensure the treatments were standardised.* As Edwards and Guttentag (1975) say, 'it is certainly nonsense to assume that some program, implemented in different ways by different people in different places, is a single entity simply because it is called by a single name and perhaps funded from a single source of money ... such variations make the experimental approach to evaluation difficult to apply'. Experimenters were pushed into assuming that what actually took place was what was supposed to take place. Boruch (1973), for instance, acknowledges that 'in the field, treatments may be poorly imposed, may not be delivered, may be delivered inappropriately to members of the control group ...'

A last, and important, shortcoming in the experimental design itself is that it was shown to be *seriously limited in the information it could produce for policy-making.* The science got in the way of the sense, demonstrating the force of Cohen and Garet's (1975) observation that 'methodologically superior knowledge is often more complex, arcane and hard to interpret'. The demands of rigour and precision obliged the experimenters to focus on tightly-defined questions about measurable variables using carefully calibrated instruments to collect quantitative data that could be statistically tested. Policy-makers just wanted to know

what happened. They were not interested merely in how far a programme had achieved its goals but in what forces had shaped its development, what opposition it had encountered, the reasons for its success or failure and any unanticipated consequences it may have had. Broad-brush matters such as these did not fall within the ambit of experimental designs and reduced their utility accordingly.

Furthermore, the results from experiments turned out to be much less convincing and authoritative than their advocates had promised (Berk et al, 1985). Few produced completely unambiguous findings (Boruch, 1973); many encountered problems of execution so serious they had to be converted into quasi-experiments; while others, especially the larger, more high-profile ones, provoked furious controversy among social scientists about their methodological adequacy, leaving policy-makers no wiser about the validity of their conclusions. A common problem was 'experimental mortality'. High drop out rates in a lot of studies – especially among non-captive subjects outside of institutional settings, and among the control groups – undermined their statistical foundations (Kramer and Shapiro, 1984). At this point, design difficulties began to merge into implementation dilemmas.

Implementation Dilemmas

One of the major threats to the integrity of an experiment was failure to adhere to the design (Riecken and Boruch, 1974). Such hiccups arose, as Marris and Rein (1967) have pointed out, because the demands of action and experimental research are not all the same and their claims are hard to reconcile. Research calls for a clear and unwavering purpose, and a clear definition of the means by which it is to be pursued – which then must be exactly and consistently followed, without revision, until the experiment has been completed. Action, on the other hand, is tentative, non-committal and adaptive. It is responsive to changes as events proceed. Their different bearings make it difficult for each to be carried out as part of

the same operation. The testing of social action programmes in the field generally imposes a more explicit and rigorous definition of means and ends than administrators in charge of running them can comfortably sustain (Marris and Rein, 1967).

Experience of social experimentation repeatedly showed that, as the a priori ideas of action programmes were put to the test of a field trial, the expectations, convictions and knowledge of programme managers and their staffs altered considerably, with the understandable result that they pressed for or introduced changes in how things were done (Riecken and Boruch, 1974). As soon as staff discovered better ways of running the project or serving their clients they tended to adapt their procedures, methods and techniques accordingly. It was almost impossible to combine good, flexible management on the one hand with scientific, controlled research on the other (Clarke and Cornish, 1972).

In the frequent battles that broke out between administrators and researchers the former usually won the day. When all was said and done, they had their hands on the wheel and, so it seemed, common sense on their side. One recurring point of conflict was over control of the randomisation process. Practitioners and professionals often resisted or subverted the idea of random allocation of subjects to treatment (Boruch, 1987; Roos et al, 1977; Landy and Wechsler, 1960). Moreover, it was not easy to argue that what managers clearly saw as a failure in a programme should not be corrected until the experimental results had verified their judgement; or that a potentially successful programme should be allowed to crash in the interests of science for want of a bit of tinkering. The cost of common sense, however, was the loss of experimental rigour.

Evaluation Dilemmas

Conflicts between researchers and administrators were also a feature of what Rivlin (1973) called the evaluation dilemmas accompanying social experimentation.

Objectivity required that the research side of an experimental programme should be detached from its operational side. The evaluators could not effectively wear both hats. This split often grew into a breach. Field staff whose first loyalty was to the programme showed a lack of commitment to the research effort. Evaluation staff, locked in their eyries away from the scene of action, were insulated from what was happening in the field (Radin, 1977). Research directors chafed at the inconsistency and incoherence of much that was done in the front line. Programme managers cavilled about the theoretical preoccupations of the research team and the methodological constraints placed on their day-to-day work. Misunderstandings flourished in the crossfire.

Two particular consequences had worrisome implications for the validity of evaluation findings (Weiss and Rein, 1970). First, there were the likely effects on the reliability of records maintained for research purposes by uncommitted field staff who neither appreciated nor valued their importance. Second, there was the danger of goal displacement: where field staff chase the outcome indicators chosen by the researchers to measure the programme's success. As Campbell (1975) has warned, 'the more any quantitative social indicator is used for social decision-making, the more subject it will be to corrupting pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor'. In other words, the very act of operationalising the programme's aims carried with it the risk of influencing the behaviour of field staff and so thwarting the experiment.

But the problems of evaluation went deeper still. The mounting of an experiment is a political act, not primarily a scientific decision (Riecken et al, 1974). It involves a big commitment of resources so it must deal with a policy issue in which its sponsors already have a substantial political investment. Its implementation often affects local alliances, awakens public interest and creates its own constituency of support. As Senator Moynihan has

observed, 'The general problem of government in the experimental mode is that experiments (innovative trials) create interests which make for the perpetuation of all manner of activities' (quoted in Cronbach, 1980). In short, as Thomas (1985) has observed, its 'very existence contains the seeds of its continuation'. Bernstein and Freeman (1975) take the same point a step further by suggesting that many experiments were set up as a pretext for obtaining funding for the operating programmes. In such cases, the evaluation components were little more than superficial trimmings added on so as to appease the doubters.

Following through this line of reasoning, the bigger experiments, and many were very big indeed, should also properly be regarded as programmes in their own right. That is to say, they lived or died in the world of politics more than in the world of science (Yarmolinsky, in Abt, 1976). Considerations of legitimacy, feasibility and support were as crucial as those of impact and effectiveness, if not more so, in deciding their eventual fate. For this reason, the validity of a social experiment turned out in many cases to be almost irrelevant in predicting whether its policy recommendations would be adopted. Social scientists were misguided in assuming that more sophisticated, and technically sound, research would necessarily lead to greater utilisation (Higgins, 1980).

Timing Dilemmas

Another obstacle to the use of experimental results for policy-making was timing. Experimentation is a lengthy business. The New Jersey Income Maintenance Experiment, for instance, lasted six years from the award of the contract to the submission of the final report. The Seattle-Denver NIT Experiment was launched in 1970 and issued the final report in 1983. Such a timescale carries with it a real risk that issues on the agenda when an experiment is launched will have faded into insignificance before the findings appear. To be useful, therefore, experimentation should really

be concerned with policies that may appear on the political agenda of the future – probably well beyond the political horizon (Rossi, 1975). Few administrations would be prepared to back a hunch about priorities so far ahead.

What happened instead was that politicians rarely financed an experiment until the problem it addressed landed on their doorstep, or rarely took seriously a new proposal unless there was a chance of doing something about it soon. So, often, the same political conditions that made it possible for federal agencies to fund a field experiment of an innovative policy were also the ones that favoured the introduction of the policy itself (Rossi and Wright, 1977). In such cases, the pressures were for quick results. The danger was that the experiment would not be finished before the proposal was enacted.

The scheduling of experiments rarely geared with the timetabling of decisions. Unless they were very fortunate, experimenters often had to choose between the imperatives of science and those of policy. The former option risked missing the boat. By grabbing the moment to inform decisions, however, they risked invalidating the experiment. This conundrum showed once again that policy usually would not wait on definitive research.

Ethical Dilemmas

Moral dilemmas are not unique to the method of social experimentation; they arise in all forms of social research. Indeed, it can be argued that the ethical problems of experimental evaluation largely mirror those of social reform itself: because it amounts to no more than a controlled trial of a new programme (Riecken and Boruch, 1974). As such, whatever 'price' an experiment imposes on those subjected to it may unknowingly be paid anyway if the programme is implemented without testing (National Academy of Sciences/Social Science Research Council, 1969).

Nevertheless, in order to justify experimentation in the eyes of sponsors and the wider public its practitioners have had to accommo-

date their moral qualms about some features of the method. Broadly, these concerns fall into four categories:

● Concerns about the possible *manipulation* of experimental subjects or, more sinister still, about the manipulation of the method itself by establishment forces seeking ways of regulating the lives and opinions of citizens.
● Concerns about the possibility of exposing people to *damaging or detrimental treatments* for experimental purposes.
● Concerns about the *fairness* of experimental programmes where the benefits are arbitrarily denied to some of the participants in the trial.
● Concerns about the effects on recipients of *withdrawing* the benefits of the programme once the trial is completed.

Such ethical issues could place political limitations on the use of the experimental method (see Rivlin and Timpane, 1975 for an extended discussion of these issues). Random assignment to treatment and control groups was not always easy to justify. Administrators and officials risked public censure for allocating scarce resources on the basis of chance rather than some more acceptable criterion such as need or merit or even 'first come, first served' (Rivlin, 1971). Deliberately treating one group differently from another group could easily provoke a backlash from those who felt they were getting the raw end of the deal.

Moreover, because ethical considerations prevented the withholding of existing services for which people were eligible, experimentation could only be used for field testing add-ons or innovations. It was not a method for evaluating the status quo.

Governments or agencies not interested in innovation were unlikely to be interested in experimentation.

But perhaps the most serious ethical dilemmas had to do with issues of confidentiality and openness. Social experiments may be more vulnerable to threats to the confidentiality of their data than other kinds of social research (Riecken and Boruch, 1974). Opera-

ting at the cutting edge of new ideas in social policy, they were often caught up in political controversy and legal argument. In these battles, the status of the data could be seriously challenged.

The New Jersey Income Maintenance Experiment ran into just such problems. Shortly after the experiment began the county welfare authorities started looking into the misappropriation of state welfare payments by families receiving overlapping benefits under the experimental negative income tax scheme. Legal proceedings were instituted which rumbled on for the better part of two years culminating in a four-month grand jury investigation. In the process, the experimenters were subpoenaed to produce the confidential records of some of their families (a case eventually resolved out of court), and later a number of families were wrongly pilloried in the local press as welfare frauds. As the researchers concluded, these happenings 'illustrate the potential vulnerability of any social experiment to the antagonism of local officials'. (Kershaw and Fair, 1976).

The issue of openness presents a different sort of dilemma. It is generally held that, on ethical grounds, social experiments using people as subjects should only proceed with their informed consent. They should not be deceived about the purposes of the experiment or their part in it. Such awareness, however, raises difficulties for the researcher. People who know they are being watched may behave differently. This brings into play a host of familiar threats to the working of the experiment. Among the more important of these are:

- *Hawthorne effects* where the subjects make special efforts to live up to the expectations of the researchers.
- *Placebo effects* where the subjects act as if the idea being tested were true so turning it into a self-fulfilling hypothesis.
- *Volunteer effects* where subjects, knowing the purpose of the experiment, either selectively decline to participate or, later, selectively withdraw so destroying the compara-

bility of the experimental and control groups.

- *Limited duration effects* where the expectation that the experimental treatment will be only of a limited duration may cause participants to behave differently than they would in a regular programme.
- *Sabotage effects* where subjects deliberately behave in a way contrary to that expected or seek to manipulate the outcome for their own ends. Kenneth Clarke, for example, rejected the idea of pilot testing his NHS reforms proposed in the White Paper *Working for Patients* because he believed that vested interests would be intent on ensuring they failed (The Guardian, 1989).

Problems of this sort once again throw doubts on just how far the results of an experimental programme provide a valid basis for predicting what would happen if it was extended to cover the country. While technical methods can be found for mastering these effects they all tend to increase the complexity, the cost and the length of experiments: considerations which themselves were powerful drawbacks to the widespread use of experimentation.

The Politics of Rigour

Do these lessons from US experience support the idea that the pursuit of rigour will improve the utility of evaluation research for policymakers? Should evaluators and social researchers invest more effort in conducting social experiments? Would policy-makers be more inclined to heed the findings from experimentation? The answers give no great cause for hope.

Enough has been learned to show that policy-making could not be turned into a laboratory science. As Riecken has observed (in Abt, 1976), it may be only when an innovative programme is one to which relatively few people are willing to give serious consideration that an experiment can be run without it being captured by political forces. In other words, experimentation works best in scientific terms

under conditions when it is least likely to command support or attention. Other options apart from the narrow pursuit of rigour must be followed in order to meet the objectives of policy relevance and usefulness (Booth, 1986).

So many unforeseen factors can derail even the most carefully planned experimental design that the quality of the information eventually produced, often after years of research, has rarely justified the effort. Experience has shown that social experiments are only feasible and worthwhile where (Rossi, 1975; Heclo and Rein, 1980):

- the programme under trial is a simple one with clearly-defined aims;
- there is a need to establish its effectiveness;
- the inputs are specific and measurable;
- people can agree on how the outcomes should be measured;
- randomisation is both politically feasible and administratively possible;
- ethical objections do not intrude;
- non-co-operation or attrition can be kept within acceptable bounds;
- and the results are likely to be useful and timely.

Too few innovatory programmes meet these guidelines for social experimentation ever to play a significant role in evaluation research or policy development.

On balance, it must be concluded from US experience that experimental designs have not stood up well to the buffetings of the action setting. This is not to say social experiments have been a complete waste of time. They have produced tangible returns in the form of administrative knowledge. They have helped to link policy-abstractions with practice complexities (Goodwin, 1975); created opportunities for trying out new services without having to make long-term commitments; and stimulated innovation. But all these benefits can be had in other ways: by demonstration projects, pilot schemes, simple field trials or case studies. The extra costs of scientific rigour have been shown to add too little in the our have been shown to add too little in the way of usable knowledge. This surely must be

the key lesson for practising social researchers today.

Tim Booth

References

Abt, C., 1976, The Evaluation of Social Programs, Beverly Hills: Sage Publications.

Arrow, K., 1976, Discussion: evaluation of social experiments, in C. Abt, The Evaluation of Social Programs, Beverly Hills: Sage Publications.

Berk, R., Chambers, R., Chambers, D., Rossi, R. and White, A., 1985, Social policy experimentation: a position paper, Evaluation Review, 9(4), 387–429.

Bernstein, I. and Freeman, H., 1975, Academic and Entrepreneurial Research, New York: Russell Sage.

Booth, T., 1979, Research and policy making in local authority social services, Public Administration, 57, Summer, 173–186.

Booth, T., 1986, Social research and policy relevance, Research, Policy and Planning, 4(1–2), 15–18.

Booth, T., 1988, Developing Policy Research, Aldershot: Gower.

Boruch, R., 1973, Problems in research utilization: use of social experiments, experimental results and auxiliary data in experiments, Annals of the New York Academy of Sciences, 218, 56–77.

Boruch, R., 1974, Bibliography: illustrated randomized field experiments for program planning and evaluation, Evaluation, 2(1), 83–87.

Boruch, R., 1987, Comparative aspects of randomized experiments for planning and evaluation, in M. Bulmer (ed.), Social Science Research and Government, Cambridge: Cambridge University Press.

Boruch, R., McSweeny, A. and Soderstrom, E., 1978, Randomized field experiments for program development and evaluation: an illustrative bibliography, Evaluation Quarterly, 4, 455–695.

Broadfoot, P. and Nisbet, J., 1980, The Impact of Research on Policy and Practice in Education, Aberdeen: Aberdeen University Press.

Campbell, D., 1969, Reforms as experiments, American Psychologist, 24(4), April, 409–429.

Campbell, D., 1975, Assessing the impact of planned social change, in G. Lyons (ed.) *Social Research and Public Policies,* Hanover NH: Public Affairs Center, Dartmouth College.

Clarke, R. and Cornish, D., 1972, *The Controlled Trial in Institutional Research,* Home Office Research Studies No. 15, London: HMSO.

Cohen, D. and Garet, M., 1975, Reforming educational policy with applied social research, *Harvard Educational Review,* 45(1), February, 17–43.

Cronbach, L., Ambron, S., Dornbusch, S., Hess, R., Hornick, R., Phillips, D., Walker, D. and Weiner, S., 1980, *Toward Reform of Program Evaluation,* San Francisco: Jossey-Bass.

DHSS, 1985, *The Experimental National Health Service Nursing Homes for Elderly People: An Outline,* April, London: Department of Health and Social Security.

Edwards, W. and Guttentag, M., 1975, Experiments and evaluations: a re-examination, in C. Bennett and A. Lumsdaine, *Evaluation and Experiment,* New York: Academic Press.

The Guardian, 1989, Tests that Mr. Clarke must face, February 13.

Guttentag, M., 1973, Evaluation of social intervention programs, *Annals of the New York Academy of Sciences,* 218, June, 3–13.

Guttentag, M. and Soar, S., 1977, *Evaluation Studies Review Annual,* Vol. 2, London: Sage Publications.

Heclo, H. and Rein, M., 1980, Social science and negative income taxation, in Organization for Economic Co-operation and Development, *The Utilization of the Social Sciences in Policy Making in the United States,* Paris: OECD.

Higgins, J., 1980, The unfulfilled promise of policy research, *Social Policy and Administration,* 14(3), Autumn, 195–208.

House of Commons, 1983, *Public Expenditure on the Social Services: Minutes of Evidence,* Social Services Committee, 321–i and ii, Session 1982–83, London: HMSO.

Kershaw, D. and Fair, J., 1976, *The New Jersey Income Maintenance Experiment,* Vol. I, New York: Academic Press.

Kramer, M, and Shapiro, S., 1984, Scientific challenges in the application of randomized trials, *Journal of the American Medical Association,* 252, 2739–2745.

Landy, D. and Wechsler, H., 1960, Common assumptions, dimensions and problems of path-

way organizations, *Journal of Social Issues,* 16(2), 70–78.

Lynn, L., 1978, The question of relevance, in L. Lynn (ed.), *Knowledge and Policy: The Uncertain Connection,* Washington DC: National Academy of Sciences, Vol. 5.

Mark, M. and Cook, T., 1984, Design of randomised experiments and quasi-experiments, in L. Rutman (ed.), *Evaluation Research Methods,* Beverly Hills: Sage Publications.

Marris, P. and Rein, M., 1967, *Dilemmas of Social Reform,* London: Routledge and Kegan Paul.

National Academy of Sciences/Social Science Research Council, 1969, *The Behavioural and Social Sciences: Outlook and Needs,* Englewood Cliffs: Prentice-Hall.

Phillips, M., 1991, If it's human to err, why not admit it?, *The Guardian,* Friday, 1 March, p. 20.

Radin, B., 1977, Political relationships in evaluation: the case of the experimental schools program, *Evaluation,* 4, 201–204.

Riecken, H. and Boruch, R. (eds.), 1974, *Social Experimentation: A Method for Planning and Evaluating Social Intervention,* New York: Academic Press.

Rivlin, A. and Timpane, M. (eds.), 1975, *Ethical and Legal Issues of Social Experimentation,* Washington DC: The Brookings Institution.

Rivlin, A., 1971, *Systematic Thinking for Social Action,* Washington DC: The Brookings Institution.

Rivlin, A., 1972, Why can't we get things done?, *The Brookings Bulletin,* 9(2), 5–9.

Rivlin, A., 1973, Social experiments: the promise and the problems, *Evaluation,* 1(2), 77–78.

Roos, L., Roos, N. and McKinley, B., 1977, Implementing randomization, *Policy Analysis,* 3(4), 547–559.

Rossi, P., 1975, Field experiments in social programs: problems and prospects, in G. Lyons (ed.), *Social Research and Public Policies,* Hanover, NH: Public Affairs Center, Dartmouth College.

Rossi, P. and Wright, S., 1977, Evaluation research: an assessment of theory, practice and politics, *Evaluation Quarterly,* 1(1), 5–51.

Salasin, S., 1973, Experimentation revisited: a conversation with Donald T. Campbell, *Evaluation,* 1, 7–13.

Shaver, P. and Staines, G., 1971, Problems facing Campbell's 'experimenting society', *Urban Affairs Quarterly,* 7(2), December, 173–185.

Thomas, P., 1985, *The Aims and Outcomes of Social Policy Research*, Beckenham: Croom Helm.

Weiss, R. and Rein, M., 1970, The evaluation of broad-aim programs: experimental design, its difficulties and an alternative, *Administrative Science Quarterly*, 15(1), March, 97–109.

Williams, W., 1980, *Domestic Program Policy Analysis: Some Observations on its Practice at the Top of the Government in the US and the UK*, Discussion Paper No. 4, Institute of Governmental Research, University of Washington.

Lokal självstyrelse i Centraleuropa – en underskattad hörnsten i reformarbetet

Inledning[1]

Två frågor har dominerat debatten kring den pågående samhällsomvandlingen i Central- och Östeuropa[2]: demokratiseringen av det nationella politiska systemet och ekonomins marknadsorientering. Dessa två frågor tycks överskugga alla andra i betydelse. Demokratin måste förankras och marknadsorienteringen uppfylla några av de förväntningar som ställts på den, innan regionens auktoritära historiska arv, realsocialistiska förflutna, etnisk-sociala instabilitet och ekologiska kris kan börja övervinnas tycks de flesta bedömare mena.

Syftet med föreliggande artikel är att lyfta fram ett tredje tema som hittills, åtminstone hos betraktare i väst, intagit en mera undanskymd position, nämligen framväxten, eller kanske snarare återerövrandet av kommunal självstyrelse. Utan en bred, lokal förankring av den nya politiken är det i högsta grad osäkert vad för slags samhällen som kommer att resa sig ur ruinerna efter de realsocialistiska experimenten. Det finns starka skäl av både deskriptivt och normativt slag för en sådan ståndpunkt.

Efter en konstitutionellt orienterad bakgrundsteckning, för vi fram argumentet att den lokala självstyrelsen har en potential som förmedlande länk mellan staten och det civila samhället. Därefter sammanfattar vi den loka-

la styrelsens roll under realsocialismen, ger sedan översikter över de senaste årens självstyrelsereformer och lokala val i Polen, Ungern och Tjeckoslovakien för att därpå lyfta fram några av de problem, orosmoln och ljuspunkter som åtföljer rekonstruktionen av den kommunala självstyrelsen i de tre länderna. I ett sammanfattande slutord berör vi också frågan vilket intresse den komparativa statsvetenskapliga forskningen kan ha av att också inkludera de centraleuropeiska staternas kommuner som ett av sina studieobjekt.

Konstitutionella reformer

De nya regimerna kämpar mot tiden. Även om de nationella och etniska motsättningarna i Centraleuropa ännu inte har samma omfattning och intensitet som i Jugoslavien och Sovjetunionen, saknas inte risker för att inre motsättningar av olika slag skall sätta stopp för demokratiseringsprocessen och öppna vägen för nya former av diktatur. I värsta fall kan vi få se en ny järnridå dras ned över halva Europa. Denna gång handlar det om en järnridå inte mellan två konkurrerande samhällssystem, utan mellan en västlig sfär av välstånd och stabilitet och en östlig sfär i upplösning. Splitter från den exploderande bomben skulle sprida sig över Västeuropa i form av en våg av flyktingar – om väst är humanitärt nog att ta emot dessa.

Flera faktorer har framhållits som viktiga för att ge omvandlingsprocesserna i Central- och Östeuropa ett gynnsamt förlopp. Ekonomer har talat sig varma för ekonomiska "chockpaket" – en marknadsekonomi bör betraktas som en helhet och går inte att genomföra stegvis.