

Reforms as experiments: cases and experience in the international perspective

Prof. em. Dr. Hellmut Wollmann  
Humboldt-Universität Berlin  
hellmut.wollmann@rz.hu-berlin.de

Discussion paper

presented at the IPEN Conference „Reforms and Evaluation of Programs“  
held at Higher School of Economics, HSE, Moscow on September 27.- 28, 2007

## 1. “Reforms as experiment” – Origins, scope and key elements

“Reforms as experiment” was the title of a paper which was published by Donald Campbell in 1968 and which became the much-quoted lead article and “battle cry” of a school of political and social science thought that gained wide currency in the 1960s and 1970s. It echoes and reinforced a policy period and policy movement which struck new political and conceptual ground on a number of scores.

First, reforms became the political trademark of a policy period in the United States since the mid-1960s when, in reaction to increasing social tensions and race riots particularly in America’s big cities, a “War on Poverty” with Social Action Programs was launched under the Johnson Administration. This push for policy reforms was taken up in the course of the 1960s also in West European countries moving towards an “advanced welfare state”, especially in West Germany with a wave of infrastructural and social policies as well as in Sweden with the further build-up of the Swedish Welfare State. So “reforms”, that is welfare state reforms, became an emblem of an entire policy period.

Second, the launching of these reform policies was embedded and rooted in a concept of policy making in which, in order to strengthen the decision-making and operational capacity of the political and administrative system planning and information were assigned increasing importance. PPBS (Programming Planning Budgeting System) temporarily epitomised the central function which planning was given in forecasting and modelling future courses of action and policy. Within the “policy cycle” running from policy formulation through policy implementation to policy termination evaluation was “discovered” as a crucial analytical instrument to identify and report back (“feed back”) the results and goal achievement of reform policies. Policy evaluation swiftly became a standard operation in policy making.

Third, in designing, planning, implementing and, last not least, evaluating the reform policies (social) science was turned to and “tapped” on an unprecedented scale which holds true for the United States in the early 1960s as well as for West Germany in the late 1960s and beyond.

Fourth. The growing importance of analysis and information as well as the increasing involvement of (social) scientists in performing such analytical and informational activities broke the ground for the advance of the idea that policy making should be increasingly placed on a “scientific” basis.

In the traditional concept of policy making the “world of politics” is the central, if not exclusive actor. It is driven by the “political rationality” which is dominated by interests and political power. As Harold Lasswell put it in a much-quoted definition: “Politics is who gets what when and how”. Hence, following from this “political rationality” political decision-making is guided by political, economic etc. interests and power.

By contrast, the “world of science and scientists” revolves, ideally speaking, around its “scientific rationality”, that is around the autonomous search for truth. Following from its concept and postulate of a “scientification” of policy-making it has been stipulated that the “political rationality” of decision-making, based on interest and power, should be curbed, if not replaced by a the “scientific rationality”, that is, the principle that political decision-making should be based on scientifically generated information and evidence. Accordingly, science, that is, social science is called upon to exercise an “enlightenment” function as, correspondingly, the policy-makers are expected to accept such “enlightenment” through (social) science- generated information and evidence.

*Reforms as experiments – climax and epitome of the movement towards a “scientification” of policy making*

On the background of this development the call for “reforms as experiments” can be seen as its logical consequence and epitome since, in its core, it is the demand and vision that political decision-making (on reforms to be embarked upon) should be truly “scientificised” by essentially running through three stages.

Phase 1: Political decision making. The decision-making in terms of setting the goals and defining the policy measures and steps that are expected to achieve the goals lie with the policy makers. Yet within the concept of the “scientification” of policy making the scientists are seen to play an important role in the policy formulation phase for two reasons. First, any policy in its intention to link certain policy measures to certain policy goals implies a theory of action. To spell out this theory of action (social) scientists are seen to play an important

role, for instance, by providing available research knowledge to this process. Second, their contribution in this phase is already crucial in order to prepare phase 2, the experimental policy design proper.

Phase 2 centres around the design and conduct of the experimentation itself. This means to set up an experimental arrangement which is capable to “test” the policy program or policy measure (or even an alternative set of programs) as to whether the intended goal is attained or missed.

The basic logic of this is borrowed from nature science and also from small group psychology as the earliest social science field to which experimentation was introduced (Small wonder that Donald Campbell, as a key proponent of the experimental policy design was a leading psychologist!). In analogy to the natural science experiment the pivotal idea of social experimentation is this:

In order to find out whether a certain factor (called the “treatment” or in methodological terms: the “independent variable”) has an effect on a certain group or unit the experimental arrangement essentially consists, on the one hand, of the “experimental” group which is exposed to what in experimental terms is called the “treatment”. At the same time, a “control” group is set up which is not exposed to this treatment. A crucial element and precondition of the experimental arrangement is that steps are taken to make the two groups, that is the “experimental” group and the “control” group, to be similar or “identical” with regard to as many features (“variables”) as possible, ideally all other features and variables. In methodological parlance this crucial precondition is often called the “*ceteris paribus*” clause which, translated from Latin means, “all other variables being equal”. An important strategy to establish the “all other variables being equal” clause is composing the two groups by way of randomisation. If randomisation cannot be done the researcher may turn to so called “matching”, that is, trying to seek as much similarity as possible between the two groups or units. The underlying assumption is that, if the two groups are identical (“equal”) with regard to all features their possible effect can be seen as being “neutralised” which allows the conclusion that any difference which shows between the two groups after the occurrence of the treatment can be causally ascribed to the treatment. It is evident that the capacity of the “experimental arrangement” to produce reliable and “robust” results crucially hinges on the question whether and to which degree the “all other variables being equal” (*ceteris paribus*)

clause can be effected by way of establishing an “experimental” and a “control” group through “randomisation” or a methodologically sufficing “matching” procedure.

The third stage “belongs” again essentially to the policy-makers who are expected, within the experimental logic, to base their decisions upon the empirical evidence produced by the experimental arrangement and the evaluation. But also in this phase the participation of the scientists may be quite important in interpreting the experimental data and assisting policy makers in translating into their political decisions.

## *2. Contexts and examples of “experimental policy” in USA, Germany and Sweden*

In the second part of this paper some concrete examples of policy experiments shall be briefly presented in order to illustrate some of the policy fields in which they have been conducted and in order to hint at some specific methodological assets and liabilities of policy experiments. The examples will be drawn from the USA, Germany and Sweden since the USA, as it was just pointed out, were the pace-setter in this policy approach and since Germany and Sweden were among the “first wave” countries in adopting policy evaluation as a standard procedure in policy-making

### *2.1. Educational policies*

An early and classical field of policy experiments was education as a policy area which aims at influencing the behaviour and aptitude of young people (learning skills, social skills etc.) by changing the educational context.

#### *2.1.1. Head Start*

The earliest and perhaps best known example was the so called “Head Start program” which was launched in the United States back in 1964 which the aim to provide children from disadvantaged minority groups with a pre-school training in order to improve their ability to be integrated in the “normal” school system. The program was targeted at 500.000 children. Its purpose was to prepare the future decision as to whether the program should be continued and be put on a much wider basis..

In order to constitute the “experimental group” within a full fledged experimental design, in a first step, some 2000 children were selected by randomisation. In second step, in order to form the “control group”, another set of children was selected, again by randomisation, that did not take part in the Head start program. In an “before and after” design possible changes in the behaviour and skills of the children of both groups were measured. From the difference which the children of the experimental showed from those of the control group it was concluded whether and which impact the “Head Start program” (as the experimental “treatment”) had on them.

Immediately after the results of the Head Start study were published some problems which seem endemic in policy experimentation popped up. One typical problem appeared to be that those political actors who disliked the findings for political and ideological reasons turned to attacking the study first of all for methodological reasons in raising methodological objections particularly regarding

- the methodological rigour in establishing the “experimental” and the “control” groups in a manner meeting the “ceteris paribus” requirement,
- the reliability of the indicators used to measure the behaviour and skill changes of the children.

Although the “Head Start program” was drawn, at once, into the political strife and into the underlying deep ideological conflicts and although it laid bare the methodological vulnerability of this new policy approach, it broke the conceptual and political ground for further large-scale social experiments that came to be conducted in the USA subsequently in the later 1960s and 1970s.

#### *Traditional school versus comprehensive school experiments in Germany*

Another instructive case in point in the field of education can be found in West Germany where, in the course of the 1960s, the discussion about a radical reform of the existing educational system emerged manifesting strong conceptual and ideological conflicts. The crucial question was whether the traditional “stratified” school system (differentiating between elementary school, secondary school, including “Gymnasium”) should be retained or whether the “comprehensive school” (Gesamtschule) should be introduced. In most of the Länder that in Germany are responsible for the public school system large-scale policy experiments were started in juxtaposing the traditional and the comprehensive school forms.

Although the experiments were carried out with a remarkable degree of methodological sophistication and great empirical input the thus generated evidence was regarded as proving not sufficiently conclusive to either “ideological camp” whereby the objections were typically made on methodological grounds. But, politically, the experimental phase and the great political attention which it aroused contributed, no doubt, greatly to the effect that the controversial concept of the “comprehensive school” was introduced in most Länder as an alternative school mode besides the “traditional” school form.

## *2.2. Social policy*

One of the most comprehensive and ambitious policy experiments that were conducted in the USA during the 1970s was the “Experimental Housing Allowance Program”. It was started in 1972 in order to find out, on a large scale experimental basis, which effect the provision of households with a (social policy related) housing allowance has on the housing market both on its demand side (that is, for the renters) and on its supply side (that is, for the investors, landlords etc.). Again a large number of rental households were selected, by way of randomisation, that either received such housing allowance (as the experimental “treatment”) or did not receive such subsidy (as the control group).

## *2.3. Traffic policy: speed limit control experiments*

In West Germany, in the early 1970s, a political discussion emerged on the question whether speed limits should be imposed on the national roads and autoroutes (Autobahnen) in an attempt to reduce the number of traffic accidents. Ensuingly, large-scale policy experiments were started in order to “test” whether the introduction of speed limits (100 kilometers per hour on the national roads and 130 kilometers per hour on the Autobahnen) would have the desired effect. The political debate was conducted with strong ideological overtones as the opponents of any speed limit were politically and economically firmly entrenched in the automobile industry and in the national associations of automobile drivers.

Under methodological auspices the experimental design of speed limit experiments was much easier to put place than experiments in educational or social policy particularly on two scores.

First, it is relatively easy to define and “operationalise” the policy goal, that is, increasing traffic safety, which can be fairly reliably measured by the number of accidents, particularly fatal accidents per, say, 1.000 traffic movements.

Second, this holds also true for the “experimental” arrangement of identifying road connections which are similar or “equal” with regard to number of lanes, road conditions, time of the year and of the day etc. and which thus may come close to satisfying the “ceteris paribus” requirement..

In fact, the findings of these large scale experiments were generally not questioned on methodological grounds. But in translating the findings into decision-making the political considerations and interests came unabatedly to the fore. While the speed limit was introduced on normal national roads, it has still not been imposed on the autoroutes (Autobahnen). Germany continues to be practically the only European country where such speed limit does not exist on the “Autobahnen” to this very day. The reason for this is largely as the German automobile industry and its lobby have so far been quite successful in preventing policy makers from putting a speed limit on German Autobahnen in place.

#### *2.4. Institutional reform policy: Local government reform in Sweden*

In 1974 a major territorial reform of Sweden’s municipalities was carried out which resulted to reduce the number of some 2000 municipalities (in the early 1950s) to less than 300 (with average size of some 34.000 inhabitants). The reform was meant to increase the political and administrative capacity of local government on the municipal level.

In the early 1980s the so called “Free Communes Experiment” was introduced by national legislation which intended to give local authorities, on an “experimental” basis, the right to have a more flexible structure of political institutions and to apply for being exempted from central government regulations. So, generally speaking, the so called “free commune experiments” were meant to “test” a higher degree of organisational autonomy of the local authorities. A large scale evaluation study ensued on the effects of this experimental provision. It was largely on the basis of these findings that, in 1991, a significant reform of the Local Government Act was adopted in which many of the changes which the “free commune experiment” dealt with were translated into permanent legislation.



The “free commune experiment” approach has been picked up by other Scandinavian countries (such as Norway and Finland) and has proved a remarkably successful strategy within local government reform policies.

## *2.5. Experimental legislation (“sunset legislation”)*

In the international discussion the concept of having legislative provisions adopted for a limited duration has gained momentum under the term “sunset legislation”. As this image (“sunset”) suggests the legal provisions are put in force only for a limited duration at the eclipse of which the decision needs to be taken as to whether the termination of the legal provision should be final or whether it should be set in force again, in the previous or an amended version (be it again for another limited duration or without time limit). In case the piece of “sunset” legislation is adopted under the condition that an evaluation should be conducted to find out the effects of the legislation the scheme can be seen as “policy experimentation” in the narrow sense. In case a methodologically “rigorous” evaluation procedure is not put in place, it can be seen as a “softer” form of policy experimentation with the general intention to find out “whether and how things work”. .

Be it in the more rigorous or in the softer form, another reason for passing a legal provision only on a temporary basis lies in the political process. It appears sometimes easier for conflicting political parties and interests to agree, by way of a political compromise, on the adoption of a legal provision if it is on a only temporary formula rather than on a permanent basis- As the final decision is postponed to a later date, each of the (opposing) political camps that are involved may hope that at the end its interest and political goals will prevail.

An instructive German example pertains to the federal rental legislation in the early 1970s when a major reform of the federal rental law was effected, by way of a compromise between conflicting political parties and interest groups, that is between landlords and renters. So it was agreed to adopt the legislative reform (which conspicuously strengthened the renters’ rights) with a time limit (of 4 years) and to have it evaluated by a large scale study before the decision would be made whether to continue or to discontinue the provision. At the end the legal provision was extended (without time limit), for largely political reasons because the originally opposing political party saw it politically opportune to go along with the “renter-friendly” legislation. .

Another more update example can be drawn, in Germany, from the recent fundamental reform through which the social security and the unemployment benefit scheme have been merged and in the connection with which the question needed to be decided by which institutional structure (the Federal Employment Agency or the local government level) the new integrated benefit scheme would have to be implemented. In to cope with the underlying interest conflict (for instance between the Federal Agency and the local authorities) an “experimental clause” has been introduced under which a number of local authorities may “opt” for taking over the entire administrative and functional responsibility for the integrated benefit scheme, including the related labour market activities. The duration of the experimental phase has been set (“sunset”) to be four years. At the same time, a large scale evaluation program has been started. The legislation decision on whether and how to decide on the future implementation model (whether federal or local) is thought to be based on evaluation findings at the end of the “experimental” period.

## 2.6. “*Model projects*”

Finally, as an important branch and variant of “policy experiments” in a “soft” or “light” version mention should be made of so called “model projects” or “pilot projects” which, being installed and funded at the local level on a limited duration, have often the purpose of “testing” the capacity and the impact of certain policy approaches or policy measures. With regard to the degree to which they are combined with an evaluation scheme such pilot or model projects differ considerably.

It might be added that, in the case of Germany and its federal system in which the federal level is constitutionally barred, in principle, from having administrative units of its own on the regional and local levels, the federal level has “discovered” and employed the initiation and funding of local level model projects and pilot projects as strategy and vehicle to get a practical foothold of its own, and be it only temporarily, at the local government level. A similar interest and strategy can be observed with the European Union and its Commission which is prevented from having operating units of its own within the EU members countries and for which, funding local level model projects and pilot projects is used as a strategy and expedient to have some direct access to, and link with the local level.

### *3. Methodological problems of and barriers to experimental policies*

The examples of experimental programs and projects that were given earlier should have demonstrated and illustrated the thorny methodological problems and difficult hurdles experimental policy designs are confronted with in most policy fields on two scores, namely, first in the definition and operationalisation of the policy goals the attainment of which shall be measured and, second and even much more difficult, to put an experimental arrangement in place in which, by establishing the “experimental group” and the “control group”, the methodological requirements and demands are sufficiently met, particularly with regard to the “all other variables being equal” (*ceteris paribus*)-clause.

The examples suggest that it is much easier to have these requirements met in “technical” policy fields (such as traffic speed control policy with relatively easy to operationalise and measure variables) rather than in education policy or social policy matters (where the structure of variables is much more complex both in the goal dimension as well as in the dimension of composing the “experimental” and the “control groups”).

It was also shown in the antecedent examples that a possible way out of these methodological demands of “rigorous” policy experimentation has been (quite successfully) seen in the adoption and conduct of “softer” and “lighter” versions of policy experimentation which, among others, aim at initiating and ascertaining “good practice” in a set of cases and settings.

### *5. Concluding remarks*

- The design and conduct of policy experiments has become almost a standard procedure in policy making in Western democracies.
- However, instead of large scale social experiments that were conducted in the early period of the 1960s experimental design have been pursued which are smaller in scale and methodologically less rigorous by turning, for instance, to the “sunset legislation” type (with evaluation without control group design) or model projects or pilot projects (with a “good practice” and diffusion of “good practice” strategy).
- On the one hand, the high flying hopes of the early days of “reforms as experiments” to attain a penetrating “scientification” of the policy-making process have, no doubt, largely

founded, as the interest-related and power-rooted logic of the decision-making process often still prevails. Yet unmistakably there has been a trend and movement towards “evidence-based policy making” which has been demanded, promoted and buttressed by a plurality of political, economic and social actors in whose view the generation as well as utilisation of empirical evidence, not least of the science-generated sort, has become an indispensable instrument and process to make the political system more transparent and more accountable.

- It is in this context that the experimental policy design – be it in a rigorous or be it in a “softer” and “lighter” version - has (normatively as well as empirically) become a key element in contemporary political decision making.

#### Selected references

Baldersheim, Harald/ Stahlberg, Krister 1994, Towards the Self-Regulating Municipality. Free communes and modernisation in Scandinavia, Aldershot: Ashgate

Campbell, Donald 1968, Reforms as Experiments, reprinted in Caro, F.G. (ed.) Readings in Evaluation Research, New York 1971, p. 233 ff.

Danielson, Caroline 2007, Social Experiments and Public Policy, in: Fischer, Frank et al. (eds.), Handbook of Public Policy Analysis, CRC Press: Boca Baton etc., pp. 381-390

Hellstern, Gerd-Michael 1983, Bilanz – Reformexperimente, wissenschaftliche Begleitung und politische Realität, in: Hellstern, Gerd-Michael/ Wollmann, Hellmut (eds.), Experimentelle Politik – Reformstrohfeuer oder Lernstrategie, Westdeutscher Verlag: Opladen, S. 1-78

Wagner, Peter/ Weiss, Carol Hirschon/ Wittrock, Björn/ Wollmann, Hellmut 1991, The policy orientation. Legacy and promise, in: Wagner, Peter/ Weiss, Carol/ Wittrock, Björn/ Wollmann, Hellmut (eds.), Social Sciences and Modern States, Cambridge U Press, pp. 2-27

Weiss, Carol H. 1986, The many meanings of research utilization, in: Bulmer, Martin (ed.), Social Science and Social Policy, London: Sage

Wollmann, Hellmut 2003, Evaluation in public sector reform. Towards a “third wave” of evaluation, in: Wollmann, Hellmut (ed.), Evaluation in Public Sector Reform, Elgar, pp. 1-11