

Hellmut Wollmann
Humboldt University of Berlin

UTILIZATION OF EVALUATION RESULTS IN LEGAL POLICY-MAKING AND ADMINISTRATION

The article is devoted to the analysis of evaluation method and its impact on legal policy development in governance, including in the area of law enforcement.

The author outlines the importance of question whether, when and how the evaluation results of public policies, programs and measures has been utilized (or not utilized) in policy making and policy implementation.

In the pursuit of this guiding question the article will come in five steps.

First, major variants of evaluation will be briefly presented particularly with an eye on their respective “utilization potential”.

Second, different concepts will be sketched that have been forwarded to capture the utilization of (social) science-generated knowledge by political, administrative and social actors. The author outlines a detailed description of types of evaluations: a retrospective, preliminary, ongoing, interactive, - as well as existing concepts of application of its results.

Third, looking at Germany, Switzerland and the European Union as “cases in point” pertinent research findings will be discussed. The article also subjected to a detailed analysis of the experience of Germany, Switzerland and the European Union in the field of utilization of evaluation results.

Fourth, a summary and conclusion will be given.

Fifth, some remarks on pertinent research needs will be made.

The author comes to the conclusion that that the rate of the utilization of evaluation-generated knowledge has so far turned out be, by and large, scarce. The author says that regarding the politically crucial decisions the political rationality and its underlying political will of the decision-makers prevail while concerning less important decisions evaluation-generated evidence does show some effect and, hence, a dose of scientific rationality comes to the fore.

There is also a need for further research on the subject of how should (and should not) use assessment results when making legal and political decisions.

Key words: evaluation, benchmarking, governance, enforcement, European Union, legal solutions.

Article info:

Received – 2016 October 29

Accepted – 2016 December 03

Available online - 2017 March 20

The utilization of (social) science-generated knowledge by political and social actors has been advocated and hailed since long as a crucial means and resource for the improvement of policy-making in modern times. *Max Weber* spoke of the “intellectualist rationalization through science”¹ (Weber 1922: 593) and *Harold Lasswell* proclaimed “policy sciences” as an all-out effort to mobilize the societally and interdisciplinarily available knowledge to promote political and social progress (see Lasswell 1951: 3, on this see also Wagner/Wittrock/Wollmann 1991). The idea of a “(social) scientification” of the political and societal activities has been voiced by *Donald Campbell* in his call for “reforms as experiments” (see Campbell 1969, see also Hellstern/Wollmann 1983, Danielson 2007) which aimed at guiding policy-making by scientifically accompanied and evaluated “social

¹ „intellektualistische Rationalisierung durch Wissenschaft und wissenschaftlich orientierte Technik“

experiments”). Moreover, the maxim of “evidence based policy-making” that was advanced by the New Labour Government in the U.K. in the late 1990s (see Cabinet Office 1999)² and has since gained (almost inflationary) international currency is targeted at grounding political decision-making on empirical evidence.

The evaluation of public policies, programs and measures that has first emerged as a policy strategy in the US during the 1960s and has since been pursued in many countries (see Furubo/ Sendahl eds. 2002, Wollmann ed. 2003, 2003c) has come to provide an enormous body and stock of empirical evaluative findings on the successes and failures of policies. This rapidly expanding fundus of internationally and nationally available policy-relevant knowledge rendered the question ever more salient and urgent as to whether, when and how such knowledge has been utilized (or not utilized) in policy making and policy implementation.

In the pursuit of this guiding question the article will come in five steps.

First, major variants of evaluation will be briefly presented particularly with an eye on their respective “utilization potential”.

Second, different concepts will be sketched that have been forwarded to capture the utilization of (social) science-generated knowledge by political, administrative and social actors.

Third, looking at Germany, Switzerland and the European Union as “cases in point” pertinent research findings will be discussed.

Fourth, a summary and conclusion will be given.

Fifth, some remarks on pertinent research needs will be made.

1. *Variants of policy evaluation and their respective utilization potential*

At the outset major variants of evaluation (see Wollmann 2007: 393, 2003a) and their respective utilization potential shall be sketched.

1.1. *Ex post evaluation*

Ex-post evaluation is the “classical” variant of evaluation to assess the goal attainment of policies and measures once they are terminated. Typically such evaluations are faced with two crucial conceptual and methodological problems. For one, the policy goals in terms of intended consequences need to be conceptualized by defining appropriate, if possible measurable indicators. At the same time, not intended consequences need to be taken into consideration. Second, the methodologically complex and demanding question needs to be coped with as to whether the observed effects have been “caused” by the policy or measure concerned.

As a rule, because of the research skills and capacity which the conduct of (fully fledged) ex-post policy evaluations requires they are mostly carried out by (external) researchers by way of commissioned/contractual research (on contractual research see Wollmann 2002). Once the (commissioned) evaluator has completed the study and delivered the results, it is, as a rule, entirely up to the (commissioning) political actors or agency to decide whether and how the results are going to be used. Usually the evaluator

1.2. *Ex-ante evaluation*

Ex-ante evaluation is directed at (in advance) assessing the effects and consequences of an envisaged policy, program or measure whereby possible different (alternative) paths of action (*scenarios*) are to be anticipated and explored. Also in the case of ex-ante evaluation it is, as a rule, left to the respective political actor to decide whether and how the ex-ante evaluation is used.

² Government „must produce policies that really deal with problems, that are forward looking and are shaped by evidence rather than a response to short-term pressures: that tackle causes not symptoms“.

The *ex-ante* logic is characteristic also of various types of the impact (pre-) assessment, such as the *environmental impact assessment*, the *legal regulation impact assessment* (in Germany: *Gesetzesfolgenabschätzung*, see Konzendorf 2009) and the *regulatory impact assessment*. Moreover, hinging also on the *ex-ante* logic the *ex-ante cost benefit analysis* aims at, in advance, ascertaining both the achievement (benefits) and the costs of an envisaged policy or measure (possibly also pursuing alternative paths of action); the simultaneous assessment of the costs and benefits of a measure allows to weigh the pros and cons and provide a “balance sheet” thereof.

The comprehensive evaluation system which the European Union has, since the mid-1990s, put in place for the evaluation of its structural fund programs constitutes a somewhat exceptional variant of *ex-ante* assessment in that within each of the five-year program phases a sequence of *ex-ante*, *ongoing* (in the EU terminology: intermediate) and *ex-post* evaluation steps is prescribed and that it is conceptually and procedurally mandated to feed the results of the *ex-ante* evaluation operation into the subsequent program stages.

1.3 On-going evaluation

Ongoing evaluation which, in principle, sets in as soon as the policy or measure in question starts to be implemented has the purpose to ascertain (interim) effects generated by the respective policy and measure. A crucial task of ongoing evaluation is to transmit (feedback) the interim results to the policy-makers and/or project operators concerned in order to enable them to possibly rectify and modify the related policy design or implementation process. Hence, ongoing evaluation is essentially meant to enable and foster the utilization of the pertinent information.

However, as a rule, it is left to the operator of the program or measure to decide whether and which use is made of the incoming feedback information. By contrast, in the “interventionist” and “participatory” variant of ongoing evaluation the “evaluate” is actively involved in a mutual learning and utilization process (see below 1.4.)

Benchmarking

Benchmarking may be seen as a variant of ongoing evaluation in which, in periodically or non-periodically, monitoring relevant changes and effects, the results are put into comparative perspective, be it intra-organisationally or inter-organisationally (for the development of benchmarking in Germany see Korte 2004, Hollenrieder 2004, for an international overview see Kuhlmann 2010, Jäkel/ Kuhlmann 2014, Kuhlmann/Wollmann 2014: 226 ff.). The decision whether and how to make use of benchmarking again depends on the respective operative unit. It should be mentioned at this point that in the reform of German federalism in 2009 benchmarking received a constitutional recognition (in article 91b Federal Constitution) as a comparison-related tool to “assess and promote the performance of administration”.

New Public Management-inspired performance management.

A remarkable move towards institutionalizing the crucial feedback loop can be seen in the (indicator-based) performance management system which is part and parcel of the New Public Management (and its German derivate: “New Steering Model”, *Neues Steuerungsmodell*) concept. It hinges on the premise that (indicator-based) information on the current performance, including achievement as well as cost data, is (steadily) reported back (feedback) to the responsible operators and actors. Insofar as such indicator-based controlling is designed to function primarily intra-administratively, that is, within the respective administrative unit, it can be seen as a form of *self-evaluation* which, somewhat echoing Niklas Luhmann’s concept of *self-referentiality* (see Luhmann 1993), is designed to directly and steadily link the intra-administrative actors with the flow of feedback information. In order to connect the feedback loop with the extra-administrative world an (indicator-based) reporting system is put in place which is meant to inform the political decision-makers (for instance parliament and local council) as well as the political public at large about the state of

affairs (on Germany see Bogumil et al. 2007: 303, Kuhlmann et al. 2008). Because of assigning the feedback loop such as central evaluative function and of procedurally integrating it into the evaluation system NPM has been acknowledged by some as inaugurating a “new wave” of evaluation (see Wollmann 2003a).

1.4 *Interactive, participatory and user-focused evaluation*

Within *ongoing* or *accompanying* evaluation one can discern between a primarily *analytical* modality in which the evaluator remains, as it were, detached and distanced from the evaluated operation in order to assert the independence and *objectivity* of the evaluation process and its findings (see Wollmann 2007: 394). By contrast, the *interventionist* modality of ongoing/accompanying evaluation implies that the evaluator is assumed to actively engage in the implementation process in order to contribute to rectify shortcomings and give advice to the operator. Such *interventionist* orientation approximates the (social science) concept of *action research* (see Wollmann 2007: 394). While this *participatory* mode of evaluation jeopardizes and impairs the objectivity of its findings, it enhances the potential of mutual learning and the sharing of insights.

Similar strategies to bridge the hiatus between the role of the evaluator and that of the “evaluatee” have been advanced – with different accents on the role of the evaluatee – as *interactive* (see Balthasar 2012), *user-related* (see Patton 1997) or *empowerment-related* (see Fetterman et al. 1996) modes of evaluation. They have in common that the role of the evaluatee is enhanced in the generation and utilization of evaluation results (see also Wollmann 2013: 92).

2. *Concepts of utilizing research-generated knowledge*

Dating back to the 1970s *knowledge utilization research* has developed in the USA as a social science research field addressing the question whether, how and why social science-generated knowledge has (or has not) been used and applied in the political, administrative and social practice. Carol H. Weiss was the perhaps most prominent and influential scholar in this field (see Weiss 1979, on this see also Wollmann 2009: 392 ff.). During the 1980s this research focus was also in Germany taken up by a research group that was funded by the German Research Council (*Deutsche Forschungsgemeinschaft*) and coordinated by Ulrich Beck (for the concept and results of this research group see Beck/Bonß eds. 1989).

In the course of the discussion on *knowledge utilization* different approaches have evolved which can be grouped along the distinction between *political* and *scientific rationality*. Ideal-typically speaking, the former may be seen to be characteristic of the “politico-administrative world” whose actors are essentially driven by their will to gain and retain power and to realize their own and their followers’ interests. By contrast, the *scientific rationality* can be regarded as specific of the science system and its members being (ideal-typically) committed to an independent, objective and impartial “search for truth”.

In an early phase of that debate concept of the “two worlds” prevailed (see Caplan et al. 1975) either being based on and driven by profoundly different and essentially incompatible rationalities and logics.

Within the range of concepts on the relations between the “world of politics” and the “world of science” one view perceives the decision-making in the political world as determined by the political will of the ruling majority and by its resolve to enforce their and their clientele’s interests while discarding scientific evidence. This *political rationality*-bound mode of decision-making corresponds with the *decisionist* model proposed by Jürgen Habermas (see Habermas 1968).

Closely bordering the *political rationality*-dominated decision-making model the concept of the *instrumental* knowledge utilization is based on the assumption that the political key decisions (on policy goals, resource allocation etc.) are determined *politically* that is, by *political rationality*, while social science research-generated knowledge is turned to by the political decision-makers to provide advice and information on the path and means how to

attain the politically pre-decided goals (see Alkin 2005). A similar *instrumental* concept and connotation of, as it were, ancillary knowledge utilization resonates in the *engineering* and *problem-solving* models (see Weiss 1979).

Similarly the concept of the symbolic use of knowledge tends towards the dominance of the *political rationality* as it assumes the relevant politico-administrative actors tap and use pertinent (social) science-generated knowledge and expertise in order to confirm and (post factum) legitimize pre-decided decisions (see also Balthasar 2009: 493).

In a similar vein, a *tactical* use of research-generated knowledge is guided by *political rationality* when political actors resort to commissioning research with the purpose to “buy time” and to “sit out” current conflicts (see Weiss 1979).

By contrast, the *scientific rationality* prevails in concepts according to which political decision-making is (largely) guided by (social) science-generated knowledge and evidence. A striking example can be found in the concept of and call for “reforms as experiments” as voiced and propagated by *Donald Campbell* (see Campbell 1969). His concept of “experimental policy-making” is based on the idea that the decision to adopt a certain policy or measure should finally be taken only after it was exposed to an “experimental real test” and to a scientifically rigorous evaluation. (For an overview see Hellstern/Wollmann 1983, Danielson 2007). The principle that political decisions should be essentially guided by scientific expertise corresponds to *Jürgen Habermas’ technocratic* model (see Habermas 1968).

Another school of thought ascribes to the *scientific rationality* a certain superiority and *enlightenment* potential *vis-à-vis* the “political world”. Hence, even when and if (evaluation) research-generated insights are not immediately and directly accepted and translated into political decisions, they may – in the form of “data, ideas, arguments” (Weiss 1991) – make their way and, as it were, “trickle into” the decision-making arena and actor constellation through multifarious diffusion paths and information networks, such as parliamentary hearings, professional workshops, and publications, mass media, informal contacts etc. (see Krautzberger/Wollmann 1988), before they finally “arrive”, in a possibly changed and “converted” (Caplan 1983) form.

Finally mention should be made of a kind of “compromise” concept in which the possible contradiction and conflict between the *political* and *scientific rationality* is bridged by a *dialogue* or *pragmatic model* (see Habermas 1968). It largely falls in line with the perception that in the “real world” of evaluation, particularly in its various “on-going” variants, the evaluation process unfolds as an interaction between the evaluator and the “evaluatee” and consequently as a mutual cognitive and learning process.

3. *Selective research on the utilization of evaluation results*

In the following chapter a brief account of the research will be given that has been conducted on the utilization of evaluation research. Besides Germany the overview includes Switzerland which since the late 1990s has seen a significant expansion of evaluation activities as well as the EU which has, since the mid-1990s, introduced a remarkably comprehensive evaluation system on its structural funding programs.

3.1. *Germany*

Since the late 1960s Germany was one of the frontrunners in the expansion of evaluation research among European countries and continued to rank among the evaluation-intensive countries (see Furubo/Sandahl 2002, Wollmann 2003b, 2003c, Widmer/Beywl/Fabian eds. 2009). More recently strong impulses to further promote evaluation have come, inter alia, from labour market policy, education policy (particularly under the OECD’s pressure conduct PISA-type surveys and assessments) as well as the EU’s push for comprehensive evaluation of its structural fund programs (for an overview see the policy reports in Stockmann ed. 2006). In the Federal Budgetary Regulation,

Bundeshaushaltsordnung (of 2001) it is stipulated that evaluations (in the terminology of that provision: “success controls”, *Erfolgskontrollen*) should be carried out on (all) “financially relevant measures”³. At the same time, the Federal Budgetary Regulation prescribes that, besides carrying out “success control” on “financially pertinent measures”, studies should be conducted as to whether the results of such “success controls” are implemented (*umsetzen*) by the ministry or agency concerned.

Measured by the large (and ever expanding) volume of policy evaluations that have conducted (and are being conducted) on the federal and *Länder* levels and considering that since 2001 the studies on “implementation” (*Umsetzung*) of evaluation results have been given legal recognition and salience, the number and scope of available pertinent studies and information is, at least on the basis of our analysis, remarkably (and surprisingly) scarce.

In 1989, the President of the Federal Court of Audit (*Bundesrechnungshof*) - in his function as “Federal Mandatee for the (Economic) Efficiency in Administration” (*Bundesbeauftragter für Wirtschaftlichkeit in der Verwaltung*)” commissioned a comprehensive study which was directed at evaluating the implementation of the aforementioned Federal Budgetary Regulation of 2001 on its two crucial scores, that is, for one, whether “success controls” have conducted by the ministries concerned, and second, whether the results of such “success controls” have been implemented (*umsetzen*).

With regard to question whether evaluation (“success control”) results were implemented the pertinent investigation conducted in 1989 reached the conclusion that the utilization rate was “small” (*gering*) (see Bundesbeauftragter 1989: 30, see also Stockmann 2006: 33). “

Some ten years later the President of the Federal Court of Audit commissioned a follow-up investigation. Again the assessment of the utilization rate was largely negative. “The implementation of the results of the ‘success control’ (e.g. correction, continuation or termination of a measure) with the exception of few ministries varied greatly, but was in sum small (*gering*). In some cases measure were continued respectively terminated in defiance of the pertinent results of the ‘success control’ without explaining why” (Bundesbeauftragter 1998: 27).

Among the ministries only the Federal Ministry of International Economic Cooperation was positively set off in various aspects. For one it was pointed out that “insights into the causes of failed measures are taken notice of in follow-up projects”, including “learning by and from mistakes” (Bundesbeauftragter 1998: 28). Moreover it was accentuated that “in selected ... projects 10 to 15 years after completion a follow-up ex post evaluation (ex post ex post) was conducted in order to assess the lasting (*nachhaltig*) success of a measure” (Bundesbeauftragter 1989: 24).

The follow-up study of 1998 also addressed the question whether evaluative “success controls” (*Erfolgskontrolle*), as required by the Regulation of 2001, were carried out. It was stated that “success controls” failed to be conducted especially in politically sensitive cases. “In cases in which it was foreseeable that the political head (*Leitung*) (of the ministry) could be (negatively) affected (*berührt*) by a failure of the measure, the officials in charge of the ‘success control’ anticipated conflicts with the political head which they wanted to avoid” (Bundesbeauftragter 1998: 33).

Also in other regards the Federal Ministry of Economic Cooperation which got a positive rating in the afore-mentioned report has been standing out among federal agencies thanks to its initiatives and innovations in the field of evaluation. As early as at the beginning of the 1970s this ministry started to build up and institutionalize an evaluation system (see Stockmann 2006b, Zintl 2009). It was the first federal ministry to create an intra-ministerial *evaluation and inspection unit* which, from the outset, was assigned also the task to foster the

³ „... Erfolgskontrollen finanzwirksamer Massnahmen“

utilization of evaluation results by feeding them back into ongoing decision-making as well as into ministry-related vocational training (see Lorz 1984: 293). While, in an initial phase, evaluation studies were largely directed at individual development and aide projects, the ministry's evaluation strategy has, since a number of years, increasingly aimed at "accumulating" and "synthesizing" the collected evaluation knowledge and information – by way of cross-cutting evaluations (so called "meta evaluations"). From the latter general recommendations and criteria should be gleaned to serve as guidance for the selection, planning and conduct of future development projects (see Stockmann 2006b: 380). In addition, in a more recent move, evaluation activities have addressed the long term and lasting effects of development projects which hitherto were often neglected as the focus of evaluation was directed on short-term effects and results. In sum, the evaluation strategies and activities that have been pursued by the Federal Ministry of Economic Cooperation have proved exemplary of a ministry's handling of its evaluation efforts as well as of putting evaluation results to use.

Over the years, the academic (university based) social science and political science research community has exhibited hardly any interest in the study of the utilization of social science-generated knowledge, leave alone evaluation-generated findings. It is true, during the 1980s a research consortium that was coordinated by Ulrich Beck and funded by the German Research Council (*Deutsche Forschungsgemeinschaft*) conducted empirical research on the utilization (see Beck/Bonß eds. 1988). However, there has not been any noteworthy major university-based research project in this field since.

It should be noted, though, that quite recently a group of university-based social scientists embarked upon a (still small-scale) project which is set to study the utilization of research findings in German federal ministries (see Spiel/ Bergsmann 2009). On the basis of a (written) questionnaire (sent out to a very small sample of high level ministerial officials - with a low response rate, see Spiel/Bergsmann 2009: 483) it was concluded that the "discrepancy between the importance which the utilization of evaluation results has and its practice is very large" (Spiel/Bergsmann 2009: 465).

Finally, a study should be mentioned which dealt, in a kind of case study, with the utilization of evaluation-generated information in the legislative activities of a federal ministry (that is, the Federal Ministry of Urban Development (*Bundesbauministerium*) (see Krautzberger/ Wollmann 1988). Based on "anecdotal" (rather than systematic) evidence (including personal participation of one of the authors in the respective legislative process) the study sheds some light on the complex setting and course of legislative decision-making in which the evaluation-generated information flow is only one of several relevant information streams and channels to which the responsible legislative actors are exposed and which provide them with often conflicting (interest-laden) information and data. Regarding the substance of decision-making the study argues that evaluation-generated knowledge hardly exert noticeable influence on legislative decisions that rank high on the political agenda of the acting government and its ministers, thus letting the *political rationality* prevail. Yet, when it comes to settling minor issues in the elaboration of legislative drafts evaluation-generated information, *instrumental* knowledge and advice often do play a relevant role. The study also gives some insights into the peculiarities of the ministerial and parliamentary legislative process as relevant evaluation-generated information may not find direct access to legislative decisions, but may go through delays and detours, may be stored and filed in ministerial archives or minds, before they may be put to use, perhaps in an altered or converted form, if and when a "(legislative) window of opportunity" finally opens.

3.2 Switzerland

In Switzerland, since the late 1990s policy evaluation has experienced a strong expansion which has lifted it to a leading position among European countries (see Widmer/Beywl 2009: 515). This development was significantly promoted in 1999 by the

adoption of article 170 in the Federal Constitution in which the federal parliament (*Bundesversammlung*) is mandated to “see to it that the effectiveness of the measures undertaken by the Federation be reviewed”⁴ (see Mader 2009: 53). Since then the evaluation function has been anchored in a great number of legal provisions, such as evaluation clauses, parliamentary supervision of administrative operations, federal finance control (*Eidgenössische Finanzkontrolle*). On the federal level some 500 evaluation studies have been carried out between 1999 and 2002 (see Balthasar 2007, Mader 2009: 60).

In the wake of this massive expansion of evaluation activities on the federal and regional levels a distinct interest in studying the utilization of evaluation results followed suit. In a survey which was conducted among the officials in charge of evaluation in the federal administration two thirds of them indicated that the “utilization rate” was “high or fairly high” (*hoch oder eher hoch*) (see Balthasar 2009: 497). In about half of the evaluations with a “high or fairly high” utilization rate the utilization mode was judged as directly implementing the recommendations given by the evaluators (see Balthasar 2009: 498). To explain the comparatively high utilization rate *Andreas Balthasar* argues, in referring to *Michael Patton*’s “process-related” utilization concept, that “a high utilization rate can be expected if and when the evaluator closely cooperated with the evaluatee” whence he derives his notion of role of the evaluator as “critical friend”. By contrast, Balthasar does not find “any evidence that the chance of knowledge utilization improves when in the institutionalization of the evaluation function the distance and independence between evaluator and evaluatee are given priority” (Balthasar 2008: 243).

3.3. European Union

For the evaluation of its structural funds the European Union has, since 1995, introduced an evaluation system which, envisaging a series of five-year periods (the first between 1996 and 2000), is based, within each program period, on a systematic sequence of ex-ante, intermediate (on-going) and ex-post evaluation (see Leeuw 2006:72). The complexity and scope of the evaluation system is enhanced by the provision that the evaluation of the European Union is meant to be carried out both on the level of the European Commission and on the level of the Member States.

On the European Commission level each Directorate-General is held to evaluate (or commission to evaluate) the policy and funding programmes within its respective jurisdiction. During the funding period of 1996-2000 some 470 evaluation studies were commissioned (see European Commission 2001: 29). In 2000 the costs thereof amounted to 14 million €. The “Annual Evaluation Reviews” give an account of the current evaluation projects (see European Commission 2001, 2010).

On the national level the Member States (in Germany, because of its federal structure, this applies first of all to the regional States) are in turn obliged to conduct evaluations on the structural funding they receive. They, too, are held to follow the three phase evaluation cycle (ex- ante, intermediate and ex-post) in pursuing the five-year rhythm (see European Commission 2001: 239, 2010: 1). In all “receiver” countries this has resulted in the emergence of evaluation research “landscape” to carry out (and to be financed through) these evaluation projects (for Germany see Toepel/Schwab 2005).

In the following our discussion will focus on and be restricted to the conduct (and the utilization) of evaluation falling under the mandate and jurisdiction of the European Commission and its Directorates-General. Our discussion will draw on the empirical study which a research team (consisting of *K. Williams, B. deLaat, G. Bastian and E. Stern*) was commissioned by the European Commission to do on the “utilization of evaluation results in the Commission” (see EPEC 2005).

⁴ „dafür zu sorgen, dass die Maßnahmen des Bundes auf ihre Wirksamkeit überprüft werden“

On the basis of its comprehensive work the research team concludes that, although the evaluation results did affect the operation (management and implementation) of the funding measures, they had no noticeable influence on politically sensitive and controversial issues, such as the distribution and assignment of financial resources. “Evaluation has a function as a tool for the management and implementation of interventions, but by and large they play a very minor role in budget allocations which remain a highly political process rather than one in which evidence stemming from evaluation has a significant input” (EPEC 2005: 11).

4. Summary and conclusions

4.1. Further expansion of evaluation activities

In recent years evaluation approaches and activities have still further expanded. This holds true for the “classical” ex-post evaluation of policies, programs and measures on which our discussion in this article has focused in picking Germany, Switzerland and the EU as “cases in point”. But this expansion applies to other approaches and variants of evaluation as well, not least thanks to the advances of indicator-based monitoring, benchmarking and performance management procedures and tools for which the generation and feed-back of evaluative results is pivotal.

Thus, the stock of evaluation generated information and knowledge about the effects of policies and measures has seen an all but exponential growth which makes the call for *evidence-based policy-making* and for utilizing the abundantly available research findings all the more mandatory and urgent.

4.2. Scope and limits of evaluation-generated knowledge utilization

In focusing on fields and examples of “classical” ex-post evaluation in Germany, Switzerland and the EU our account arrived at the conclusion that the rate of the utilization of evaluation-generated knowledge has so far turned out be, by and large, scarce. On the top of it, the referred-to studies suggest that, insofar as evaluation findings were implemented, they did not pertain to the politically crucial key decisions of policy-makers, but instead to minor (operational etc.) ones. Drawing on our earlier conceptual debate it can be said that regarding the politically crucial decisions the *political rationality* and its underlying *political* will of the decision-makers prevail while concerning less important decisions evaluation-generated evidence does show some effect and, hence, a dose of *scientific rationality* comes to the fore.

However, this assessment and interpretation arguably need some modification on sundry scores.

For one, as to the conclusiveness of the interpretation of the referred-to studies it should be borne in mind that they were largely guided by a conceptual lens which analytically recognizes only the cases of immediately effectuated utilization of evaluation results. Hence they tend to ignore modes and events of utilization that may occur later in the course of the decision-making process, such as “sleeping” effects. The referred-to sketch on the legislative activities unfolding inside a German federal ministry (see Krautzberger/Wollmann 1989) hinted at the complicated process of relevant (and possibly conflicting) information finding access (sort of trickling) into the decision-making process – in line with the *conceptual* utilization scheme with a dose of (*scientific rationality*-transmitted) “enlightenment” (see Weiss 1989).

Second, as to range of the fields of ex-post evaluation discussed in this article their coverage is admittedly limited. Major policy fields (and their evaluation repertoire) have not been included in the present account. This applies not least to the field of education and university reforms on which, in the wake of OECD’s PISA initiatives, nationally and internationally large-scale surveys and assessments have been conducted which certainly deserve attention, not least under the aspect of knowledge utilization.

Another nationally and internationally advancing evaluative track which looks promising for further discussion and research is the *benchmarking* scheme (see Kuhlmann

2010, Jäkel/Kuhlmann 2013). Empirical evidence shows that benchmarking has enhanced the operational and financial transparency of administrative activities and has induced what has called “benchlearning”. However, such overture to more transparency on the operational and finance performance has so far been often limited to the intra-administrative world whereas the political decision-makers (in the local council etc.) and the public at large have so far been not (or insufficiently) connected (see Hollender 2003: 158 ff.).

4. *Research desiderata.*

The politically, administratively and socially relevant evaluation-generated knowledge has continued to grow in an all but exponential rate. It has been propelled on the different evaluation tracks, be it ex-post and ex-ante evaluation, be it monitoring or ongoing evaluation schemes.

Whereas some empirical evidence on the utilization potential of the varied evaluation strategies and modalities is, no doubt, already available it is still fragmentary and “episodical” rather than systematic and comprehensive. The need is palpable to have further research on the question as to why and how evaluation results have been utilized (and why not).

University-based political/social science research and researchers have in the past shown remarkable (and regrettable) restraint in paying attention to the issue of the utilization of evaluation-generated knowledge in the political, administrative and social practice. Except for the emergence and blossoming of *knowledge utilization research* in the US during the 1970s and a short-lived upsurge in Germany during the 1980s (see Beck/Bonß eds.1989) there has since been no noteworthy engagement of political/social science researchers in this subject matter.

Political/social science research should move to close this research gap on two scores.

For one, the research should be resumed and pursued to empirically explore knowledge utilization on the different tracks and fields of evaluation.

Second, to overcome the fragmentation of research results and knowledge the research should move forward and attempt, somewhat reminiscent of *Harold Lasswell’s* vision of *policy science* (see Laswell 1951), to systematize and generalize the findings on why and how (and why not) pertinent research-generated knowledge is utilized.

In pursuit of this agenda political/social science research should, normatively speaking, make it point, in the tension between *political rationality* and *scientific rationality*, to advocate the latter and to, thus, in *Aaron Wildavsky’s* famous book title, “speak truth to power” (Wildavsky 179).¹⁰

References

1. Weber M. *Wirtschaft und Gesellschaft*. Tübingen, Mohr Siebeck, 1922.
2. Lasswell H.D. The Policy Orientation, in: Lerner D., Lasswell H.D. (eds.). *The Policy Sciences*. Stanford University Press, 1951, pp. 3–15.
3. Wagner P., Weiss C.H., Wittrock B., Wollmann H. The policy orientation: Legacy and perspective, in: Wagner P., Weiss C.H., Wittrock B., Wollmann H. (eds.). *Social Science and Modern States*. Cambridge University Press, 1991, pp. 2–19.
4. Campbell D.T. Reforms as Experiments. *American Psychologist*, 1969, Vol. 24, no. 4, pp. 409–429.
5. Hellstern G.-M., Wollmann H. (Hrsg.). *Experimentelle Politik – Reformstrohfeuer oder Lernstrategie*. Opladen, Westdeutscher Verlag, 1983.
6. Danielson C. Social Experiments and Public Policy, in: Fischer F., Miller G.J., Sidney M.S. (eds.). *Handbook of Public Policy Analysis*. CRC Press, 2007, pp. 381–392.
7. Furubo J.-E., Sandahl R. A Diffusion Perspective on Global Development in Evaluation, in: Furubo J.-E., Rist R.C., Sandahl R. (eds.). *International Atlas of Evaluation*. New Brunswick, Transaction, 2002, pp. 1–26.
8. Wollmann H. (ed.). *Evaluation in Public Sector Reform*. Edward Elgar, 2003.

9. Wollmann H. Evaluation in Public Sector Reform. Trends, Potentials and Limits, in: Wollmann H. (ed.). *Evaluation in Public Sector Reform*. Edward Elgar, 2003, pp. 231–258.
10. Wollmann H. Policy Evaluation and Evaluation Research, in: Fischer F., Miller G.J. Sidney M.S. (eds.). *Hand-book of Public Policy Analysis*. CRC Press, 2007, pp. 393–404.
11. Wollmann H. Evaluation in Public Sector Reform. Towards a “new wave” of evaluation, in: Wollmann H. (ed.). *Evaluation in Public Sector Reform*. Edward Elgar, 2003, pp. 1–11.
12. Wollmann H. Contractual Research and Policy Knowledge, in: *International Encyclopedia of Social and Behavioral Sciences*. Vol. 5. 2002, pp. 11574–11578.
13. Konzendorf G. Institutionelle Einbettung der Evaluationsfunktion in Politik und Verwaltung in Deutschland, in: Widmer Th., Beywl W., Carlo F. (Hrsg.). *Evaluation*. Wiesbaden, VS Verlag, 2009, ss. 27–33.
14. Korte R. Ziele, Aufwand und lokaler Nutzen von Kennzahlen und Kennzahlvergleichen für Kommunen aus der Sicht interkommunaler Vergleichsringe, in: Kuhlmann S., Bogumil J., Wollmann H. (Hrsg.). *Leistungsmessung und –vergleich in Politik und Verwaltung*. Wiesbaden, VS Verlag, 2004, ss. 123–137.
15. Hollenrieder J. Benchmarking für soziale Dienstleistungen im Rahem der Arbeitsmarkt- und Sozialpolitik in: Kuhlmann S., Bogumil J., Wollmann H. (Hrsg.). *Leistungsmessung und –vergleich in Politik und Verwaltung*. Wiesbaden, VS Verlag, 2004, ss. 150–161.
16. Kuhlmann S. Performance measurement in European local governments: a comparative analysis of reform experiences in Great Britain, France, Sweden and Germany. *International Review of Administrative Sciences*, 2010, Vol. 76, no. 2, pp. 1–15.
17. Jäkel T., Kuhlmann S. Lernen von den Besten? Steuerung und Nutzung von Leistungsvergleichen in europäischen Verwaltungen, in: Kropp S., Kuhlmann S. (Hrsg.). *Wissen und Expertise in Politik und Verwaltung, der modern staat*, Sonderheft 1/2013. Verlag B. Budrich, 2014, pp. 141–160.
18. Kuhlmann S., Wollmann H. *Introduction to Comparative Public Administration*. Elgar, 2014.
19. Luhmann N. *Soziale Systeme*. Frankfurt, 1993.
20. Bogumil J., Grohs S., Kuhlmann S., Ohm A. *Zehn Jahre Neues Steuerungsmodell*. Berlin, Sigma, 2007.
21. Kuhlmann S., Bogumil J., Grohs S. Evaluating administrative modernization in German local government: success or failure of the “New Steering Model”. *Public Administrative Review*, 2008, vol. 68, no. 5, pp. 851–863.
22. Balthasar A. Fremd- und Selbstevaluation kombinieren: der ‘Critical Friend Approach’ als Option. *Zeitschrift für Evaluation*, 2012, Vol. 11, no. 2, ss. 173–192.
23. Patton M.Q. *Utilization-focused evaluation*, 3rd ed. Thousand Oaks, Sage, 1997.
24. Wollmann H. Evaluation in Politik und Verwaltung, Ansätze, Ergebnisse, Verwendung, in: Reichard Ch., Schröter E. (Hrsg.). *Zur Organisation öffentlicher Aufgaben*. Leverkusen, Verlag B. Budrich, 2013.
25. Weiss C.H. The many meanings of research utilization. *Public Administration Review*, 1979, Vol. 39, no. 5, pp. 563–574.
26. Wollmann H. Kontrolle in Politik und Verwaltung: Evaluation, Controlling und Wissensnutzung, in: Schu-bert K., Brandelow N. (Hrsg.). *Lehrbuch der Politikfeldanalyse 2.0*, 2. Aufl. München, Oldenbourg Verlag, 2009, ss. 379–400.
27. Beck U., Bonß W. (Hrsg.). *Verwendung, Verwandlung, Verwissenschaftlichung*. Frankfurt am Main, Suhr-kamp, 1989.
28. Caplan N., Morrison A., Stambaugh R. *The Use of Social Science Knowledge in Policy Decisions at the Na-tional Level*. Ann Arbor, 1975.
29. Habermas J. Verwissenschaftlichte Politik und öffentliche Meinung, in: *Technik und Wissenschaft als Ideo-logie*. Frankfurt am Main, Suhrkamp, 1968.
30. Alkin M.C. Utilization of evaluation, in: Mathison S. (ed.). *Encyclopedia of Evaluation*. Thousand Oaks, Sage, 2005, pp. 234–236.
31. Balthasar A. Evaluationen in der Schweiz. Verbreitung und Verwendung, in: Widmer T., Beywl W., Carlo F. (Hrsg.). *Evaluation. Ein systematisches Handbuch*. Wiesbaden, VS Verlag, 2009, ss. 486–498.
32. Weiss C.H. Policy Research: data, ideas or arguments, in: Wagner P., Weiss C.H., Wittrock B., Wollmann H. (eds.). *Social Sciences and Modern States*. Cambridge University Press, 1991, pp. 290–306.

33. Krautzberger M., Wollmann H. Verwendung sozialwissenschaftlichen Wissens in der Gesetzgebung, in: *Zeit-schrift für Rechtssoziologie*. 1988, ss. 188–205.
34. Caplan N. Knowledge Conversion and Utilization, in: Holzner B., Knorr K., Strasser H. (Hrsg.). *Realizing social Science Knowledge*. Wien, 1983.
35. Wollmann H. Evaluation and public sector reform in Germany: Leaps and lags, in: Wollmann H. (ed.). *Eval-uation in Public Sector Reform*, Edward Elgar, 2003, pp. 118–132.
36. Widmer T., Beywl W., Carlo F. (Hrsg.). *Evaluation. Ein systematisches Handbuch*. Wiesbaden, VS Verlag, 2009, ss. 117–128.
37. Stockmann R. (Hrsg.). *Evaluationsforschung*, 3. Aufl. Münster, Waxmann, 2006.
38. Stockmann R. Evaluation staatlicher Entwicklungspolitik, in: Stockmann R. (Hrsg.). *Evaluationsforschung*, 3. Aufl. Münster, Waxmann, 2006, ss. 278–416.
39. Zintl M. Evaluierung in der deutschen Entwicklungszusammenarbeit, in: Widmer T., Beywl W., Carlo F. (Hrsg.). *Evaluation. Ein systematisches Handbuch*. Wiesbaden, VS Verlag, 2009, ss. 245–255.
40. Lorz R.E. Das Inspektionsreferat des Bundesministeriums für wirtschaftliche Zusammenarbeit, in: Hellstern G.-M., Wollmann H. (Hrsg.). *Handbuch zur Evaluierungsforschung*. Bd. 1. Opladen, Westdeutscher Verlag, 1984.
41. Spiel C., Bergsmann E. Zur Nutzung der Evaluationsfunktion. Partielle Bestandsaufnahme und Pilotstudie in österreichischen und deutschen Bundesministerien, in: Widmer T., Beywl W., Carlo F. (Hrsg.). *Evaluation. Ein sys-tematisches Handbuch*. Wiesbaden, VS Verlag, 2009, ss. 478–485.
42. Mader L. Die institutionelle Einbettung der Evaluationsfunktion in der Schweiz, in: Widmer T., Beywl W., Carlo F. (Hrsg.). *Evaluation. Ein systematisches Handbuch*. Wiesbaden, VS Verlag, 2009.
43. Balthasar A. *Institutionelle Verankerung und Verwendung von Evaluationen. Praxis und Verwendung von Evaluationen in der Schweizer Bundesverwaltung*. Zürich, Rüegger, 2007.
44. Balthasar A. Der Einfluss unterschiedlicher Instrumentalisierungsmodelle auf die Verwendung von Evalua-tionen: Erfahrungen aus der Schweiz. *Zeitschrift für Evaluation*, 2008, Vol. 7, no. 2, ss. 243–271.
45. Leeuw F.L. Evaluation in Europa, in: Stockmann R. (Hrsg.). *Evaluationsforschung*, 3. Aufl. Münster, Waxmann, 2006, ss. 64–84.
46. Toepel K., Schwab O. Evaluation in der europäischen Strukturpolitik. *Zeitschrift für Evaluation*, 2005, Vol. 1, ss. 63–74.
47. Wildavsky A. *Speaking Truth to Power*. Little Brown, 1979.

Information about the author

Hellmut Wollmann,
 Prof. (em.) Dr. jur.,
 Humboldt University of Berlin,
 c/o Humboldt-Universität zu Berlin, Institut für
 Sozialwissenschaft,
 Unter den Linden 6, D-10099 Berlin,
 e-mail: hellmut.wollmann@rz.hu-berlin.de,
 ORCID: orcid.org/0000-0003-1678-9715.

Bibliographic description

Wollmann H. Utilization of evaluation results in
 legal policy-making and administration.
Pravoprimerenienie = Law Enforcement Review,
 2017, vol. 1, no. 1, pp. . –
 DOI: 10.24147/2542-1514.2017.1(1).7-18.
 (In Russ.).