

Evaluation in Germany: An Overview

Gerlinde Struhkamp

Historical Perspective

Even though often considered being a “late-starter” concerning evaluation, the beginnings for systematic inquiry into impacts of governmental programs in Germany parallel the developments in the U.S. where evaluation became particularly prominent during the 1960s and from then on. As a modern democracy, too, German government and public administration were concerned with the effects of their actions, be it in the form of laws or programs or other measures of public intervention. So, for example, in 1970 the federal law was passed that there had to be “success controls” (“Erfolgskontrollen”) for governmental measures. Indeed, this law caused a leap in the market for such “success control” studies, though for the most part not academics but commercial research and consultancy firms succeeded to produce the lion’s share of the evaluation research funding (Wollmann, 1997). The relative absence of the academic world regarding evaluations sustained for another 20 years. That does not mean, however, that research and research findings did not pick up matters of broad public interest. For example, large-scale studies were undertaken to explore the effects of different forms of schooling starting at the end of the 1960s and coming to an end early 1980s (Stockmann, 2004b: 29). However, seldom such studies were called evaluation or program evaluation or evaluation research. The

peculiar term, literally translated, meaning “accompanying research” (“wissenschaftliche Begleitung”, “Begleitforschung”) emerged instead and is widely used till today. To clarify the concept: this term is not meant as action research or any form of incorporating advocative elements into the research task (at least not per se). Using evaluation jargon: such “accompanying research” highlights conceptual use and knowledge gain over instrumental purposes. On the other hand, this idea of research does involve in the field, does connect to practice. So there is the notion of feeding research results back into the ongoing process and by that possibly improving the object of analysis—much like in a formative evaluation. In fact, both terms, “accompanying research” and evaluation, are used in Germany till today, the latter becoming more prominent, though. Sometimes they are used in a way exposing visible differences (e.g., an evaluation generating explicit value judgments), sometimes they are used interchangeably. Then, it can only be fathomed in which ways such an approach to applied research overlaps with evaluation and to what extent. A thorough debate on this is still pending.

So, even though the legal and executive interests in evaluation activities did exist in Germany just as in the U.S., an evaluation *profession* did not evolve until the mid 1990s, at least not under the term ‘evaluation’, unlike in the U.S. That does not mean government did not involve any measures of control or accountability. For example, the institution of the “Bundesrechnungshof” (self-portrayal: “The Bundesrechnungshof is a supreme federal authority [...] an independent body of government auditing”¹), does conduct regular extensive checks on government spending, or ministries, e.g., the Federal Ministry for Economic Cooperation and Development, did set up a central evaluation department (Stockmann, 2004b, p.

¹ <http://www.bundesrechnungshof.de/en/1024.html>

31). Finally, a last sweep towards a broader institutionalization of evaluation came with the introduction of New Public Management also in Germany (Stockmann, 2004b, p. 32).

Therefore, looking at the history of evaluation in Germany, by and large and despite certain ups and downs there has been continuity concerning such tasks as evaluation studies. In contrast to this, though, has been the lack of professionalization towards an evaluation discipline, not only as a field of application (cf. above Wollmann's analysis of consultancies conducting "success control" studies), but also as an academic discipline.

What makes it look like that there had been relatively little interest in evaluation is partly due to the difference in language. The term 'evaluation' (spelled the same in German and pronounced only slightly differently) has become used only, roughly speaking, throughout the last decade. Before that, neither in the academic nor the political world this term really caught on. In fact, you can count almost on one hand the books that carried the term 'evaluation' (or the German adaptation 'Evaluierung', meaning the same) in the title. If so, it occurred mainly in conjunction with close ties to U.S. developments and U.S. authors. For example, as early as in 1972 Wulf published a reader mainly presenting translations of articles by U.S. authors, among them, e.g., Scriven's "Methodology of Evaluation". In a similar vain Hellstern and Wollmann issued an extensive publication in 1984 with chapters by German, but also again a number of U.S. authors. Weiss's "Evaluation Research" from 1972 was published in German language in 1974, and in 1988 Hofmann translated and adapted Rossi/Freeman's "Evaluation: A systematic approach".

Which looks like an impressive list here, was not exciting from the German perspective: literally a handful of books specifically related to evaluation in more than 30 years!

Ties to the U.S. evaluation community have generally been strong. Also Koch/Wittmann (1990) and most recently Stockmann (2000, 2004a) incorporated chapters on U.S. developments and by U.S. authors in their handbooks, Beywl's work (1988) drew on U.S. methodological developments in evaluation. Maybe Wottawa and Thierau's book from 1990 (in the meantime second edition 1998 and third 2003) could be considered a turning point towards a more German-centered evaluation-related body of literature. For example, next to introductions of evaluation concepts stemming largely from U.S. writings, they tried to focus on specific German developments, including, e.g., related concepts of quality assurance like quality management systems. Though, it may be difficult to pinpoint to a certain book since a general development has taken place.

Current Status

Starting from the early 1990s and certainly since the mid and late 1990s the relative scarcity of writings about evaluation in German language simply belongs to the past. Book publications and articles have popped up from all kinds of disciplines, within Germany and the German-speaking countries Austria and Switzerland (German-speaking part), respectively. Fortunately, there is not any “border mentality”—in the contrary: many intellectual and personal exchanges occur between German-language academics, journals and at conferences facilitated by the common language. Noteworthy in this regard is, e.g., that so far evaluators from Austria simply join the German Evaluation Society ([DeGEval](#)), while the current DeGEval president is Austrian, and the last annual conference in

November 2004 took place in Vienna. Currently there are considerations underway towards changing the society's name reflecting the German and Austrian membership (but that decision has not been made at the time of this article). Moreover, within the coming years a joint conference of the DeGEval with the Swiss Evaluation Society (SEVAL) is planned. Also, the DeGEval entertains ties with the European Evaluation Society (EES).

Probably the most significant turning point towards the establishment and professionalization of evaluation in Germany meant the foundation of the DeGEval in 1997. Since then, the society has continued to grow and spurred the intensity of intellectual discourse on evaluation-related topics. To date the DeGEval has about 370 individual and more than 50 institutional members, these numbers increasing steadily since its foundation. Unlike, e.g., the American Evaluation Association (AEA) or the UK Evaluation Society (UKES), the DeGEval does not have any regional chapters or regional networks.

The DeGEval's internal structure compares to the AEA's formation in TIGs (Topical Interest Groups). The number of TIGSs (in German called "Arbeitskreis"—"working circle" or "working group") within the DeGEval has also increased steadily and amounts to 14 as to date. Like in the AEA, the TIGs are mostly centered on a certain field of application, e.g., evaluation in schools, evaluation of developmental aid, environmental evaluation, evaluation in the field of human services, and so on. TIGs are created and may break up again once a task is done and depending on the actuality of a certain topic. For example, in the beginning of the DeGEval a working group was formed to develop German "Standards for Evaluation". Once accomplished, this group was terminated. As there is a first revision process of the standards in the making, again such a task force has been formed. A few TIGs deal with aspects of broader interest that

mainly concern matters of an evaluation profession, e.g., one TIG prepared the DeGEval's "Recommendations for education and training in evaluation" to specify evaluator competencies necessary for sound evaluation practice.

Currently, the 14 TIGs are dealing with:

- 1) Training and education in evaluation
- 2) Vocational education
- 3) Developmental aid
- 4) Research, technology and innovation
- 5) Health sector
- 6) Higher education
- 7) Media
- 8) Schools
- 9) Human services
- 10) Urban and regional development
- 11) Structural funds
- 12) Environment
- 13) Public administration
- 14) Corporate sector (in preparation)

Comparable to the AEA, each DeGEval TIG has a chair and vice-chair. The TIGs are largely autonomous in their activities. Some TIGs exist since the foundation of the DeGEval and not only sponsor meetings during the annual conference but also organize meetings (like small conferences or workshops regarding a certain topic)

throughout the year. At the two to three days long annual conferences the TIGs sponsor sessions. So far, the chair and vice-chair or TIG members look for appropriate and interesting presenters and invite them. Also, presenters can directly address the TIG and offer a topic they would like to present on. Since there is not a general call-for-papers (only by certain TIGs, if they opt for one), on the one hand, it opens the opportunity for people to be invited who normally would not answer to a call-for-papers by themselves (e.g., when they work at government agencies, foundations, corporations or other institutions with only loose ties to the academic world); on the other hand, it somewhat limits the range of presenters to the perspective of the TIG. The annual conferences sponsored by the DeGEval take place since its establishment in 1997, prior to the conference professional development workshops are offered. The conferences are held in fall (mostly October; this year it will take place at the University of Duisburg-Essen, located in the Ruhr valley, from Oct. 12-14) and in German language. However, visitors with, let's say, a working level of German language proficiency should feel welcome to attend, since there would be also ways to communicate in other languages, foremost English. Even though rarely, but there have been already occasional presentations in English.

Despite the DeGEval's attempts to encompass various disciplines (like the AEA is devoted to “evaluation in all its forms”) and the growing interest in the DeGEval's annual conferences, there are still “parallel universes” where evaluation is dealt with, mainly dominated by the traditional disciplines like education, psychology, and sociology. Their professional associations do pick up evaluation topics, e.g., in the form of own TIGs or working groups, but linkages to the DeGEval are still rather weak and sporadic. Mainly such connections exist in the way that people attend the conferences and engage in both the DeGEval and another association—

like U.S. evaluators may attend both the AERA (American Educational Research Association) as well as the AEA meeting, e.g.. The next years will show how much overlap and integration will be possible to overcome the “disciplinary segmentation” (Stockmann, 2004b, p. 35) in order to develop a common understanding on core elements of evaluation and synthesize scientific debates.

Apart from the foundation of the DeGEval, other developments foster the professionalization of evaluation in Germany. One of the society's founding members, Wolfgang Beywl, set up and administers the German-language mailing list called “forum-evaluation”, which engages several hundred enlisted members in discussions about concepts and ideas, exchanges of references, announcements of events, calls-for-papers, and the like. Similarly beneficial to the field is the first German-language “Journal for Evaluation” (*Zeitschrift für Evaluation—ZfEv*²). In its third volume (2005) it incorporates articles on theory, methods and practice of evaluation, book reviews, updates on activities of the DeGEval, and other pertinent information of interest to evaluators, sponsors, and anybody else concerned. International readers: the journal does include English-language abstracts!

Not only have evaluators and others interested in evaluation-related issues found their forums. In addition and appreciably so, as of now there are two German-language postgraduate, one of them master-level, degree programs for evaluation set in place: one in Berne/Switzerland³, up and running since 2001, the second one in Saarbrücken⁴, which had its first cohort fall 2004. In addition, during the last

² <http://www.zfev.de/>

³ <http://www.evaluationsstudium.ch/>

⁴ <http://www.htw-saarland.de/evaluation/>

years, professor positions within departments of social sciences and education have been set in place with an emphasis on evaluation, so more and more students will be trained more formally in techniques, methods and context factors concerning evaluation.

The first major product of the DeGEval has been to adopt and as professional association responsible to pass the “Standards for Evaluation”, also called the DeGEval-Standards (Deutsche Gesellschaft für Evaluation, 2002). The proximity to the wording used in the “Joint Committee Standards for Educational Evaluation” is no surprise, since the former are closely related to the latter. Like the Swiss Evaluation Society SEVAL had issued its respective evaluation standards in 2001 after a review and revision process, so did the DeGEval finalize its review process in 2001/2002 and prepared a brochure listing and explaining the standards to their members and others being interested. A new review process of the existing standards, based on a survey among DeGEval's members, is currently underway. By and large there have been only slight differences between the Joint Committee, SEVAL and DeGEval Standards, so the latter are based to a large extent on the work and experience of the Joint Committee.

In addition to the Evaluation Standards, in 2004 the DeGEval's TIG “Education and Training in Evaluation” also issued “Recommendations for education and training in evaluation—Required competencies for evaluators” (Deutsche Gesellschaft für Evaluation, 2004). These influenced already the existing academic training programs mentioned above. The latest major recommendation passed by the DeGEval is the adoption of the Evaluation Standards to the special form of self-evaluation (Deutsche Gesellschaft für Evaluation, 2005).

Specifics in Germany

In one way, for German evaluators the situation has probably been much like in the U.S. in the 1960s when the profession just started and evolved—had to find its way. But since the English language is only a relatively slight barrier to many, those interested in evaluation indulge in the English-language literature on evaluation and pick up the ideas. As a consequence, an interesting mixture of concepts rooted in American culture and approaches stemming from German traditions of social science, policy-analysis (and every other respective field) merge and emerge.

For example, in the fields of social work and human services, e.g., concerning child and youth services, the approach of “self-evaluation” has become prominent. In fact, it proved to be a “gate-opener” during the mid and late 1990s in introducing evaluation to the field, not only for approaches of self-evaluation but also other “traditional” forms of external evaluation. The DeGEval responded to this in adjusting the DeGEval Evaluation Standards to applications of self-evaluation. A respective paper explaining the specifics of self-evaluation and how the Evaluation Standards respond to them was adopted by the DeGEval members at the general assembly during last year's annual conference.

More as a side note: An interesting discussion, indebted to the terminological differentiation that's made possible by the German language, sparks from time to time, e.g., in the mailing list “forum-evaluation”. There is a dispute regarding the differences between various forms of evaluation that could be distinguished by the attributes of “internal” (German: “intern”) and “external” (“extern”) as well as “self-” (“selbst”) and—well, here there's the German term “fremd”, meaning literally “strange” or more metaphorically “outside”, which is hard to translate into

English. Thomas Widmer, Swiss evaluation researcher, suggested to translate it as “heteronomous” (evaluation), thus, an evaluation in which the evaluatees are not in charge of the evaluation, i.e., have a say in the conduct of the evaluation. In contrast, in a self-evaluation they are in charge of both the evaluand and the evaluation (so a prominent, yet not undisputed definition). Also other attempts have been made to provide a German-language glossary of evaluation terminology (even referring to the English corresponding term, if applicable)⁵, in its make-up very similar to The Evaluation Center's glossary project⁶.

Another topic that has been prevalent and mixed with evaluation debates in some sectors concerns approaches of “quality management” (“Qualitätsmanagement”), e.g., according to the approaches of Total Quality Management (TQM), the European Foundation for Quality Management's model EFQM, or the International Standard Organization's (ISO) norms (being transferred into German language and context by the German Institute for Norming—“Deutsches Institut für Normung”, DIN). A debate that has largely been absent from the evaluation community in the U.S., as far as my observation goes. In Germany, however, in some sectors there is a prevalence of quality management terminology whereas in others one of evaluation. And since terminology carries concepts, it has not been easy to pull the two strands apart. Several authors have worked out differences and similarities between these two approaches (Wottawa/Thierau, 1998, pp. 43-45; Beywl, 2001; Stockmann, 2002), but as of now it more looks like another “parallel universe”, with a conceptual conciliation still to be worked out.

⁵ <http://www.univation.org/glossar/index.php>

⁶ <http://ec.wmich.edu/Glossary/glossaryList.htm>

These are but two examples from my work context. Others surely could add more, e.g., concerning the fields of developmental aid, European structural funds, evaluation of sustainable development and so on—a list too long to be presented here.

Such disputes over various forms of evaluation and the assisting terminology are not yet settled. It's been taken serious what Michael Scriven suggested in his editorial in JMDE Num. 1: that “one must treat the definition of key existing concepts as an extremely serious matter, not a matter of casual linguistic convenience [...]. Conceptual schemes, and the definitions that go with them, are powerful instruments of analysis and hence persuasive support for particular interpretations, not minor precursors to it [...].” (2004, pp. 15-16). Indeed, there is this seriousness of—constructive—debate in the German-language evaluation community.

About the Author

After an academic degree in Germany the author received her M.A. in ‘Evaluation Studies’ from the University of Minnesota, USA, in 2002. Since then she works in Germany again, currently concerned with evaluation in the field of children and youth programs. Her general interests are evaluation of human services and the theoretical foundations of evaluation. Correspondence to: Gerlinde Struhkamp, German Youth Institute, Nockherstr. 2, D-81541 Munich, Germany. Tel.: 0049-89-62306-340, E-mail: struhkamp@dji.de or Gerlinde.Struhkamp@gmx.de. The author wishes to thank both Sandra Speer and Karin Haubrich for their thoughtful comments on an earlier version.

References

- Beywl, W. (1988): Zur Weiterentwicklung der Evaluationsmethodologie. Grundlegung, Konzeption und Anwendung eines Modells der responsiven Evaluation. [Development in evaluation methodology. Basis, conception and application of the model of responsive evaluation] Frankfurt a.M./Bern/New York u.a.: Peter Lang.
- Beywl, W. (2001): Evaluation und Qualitätsmanagement. Systemische Verfahren zur Entwicklung von Qualität im Bildungswesen. [Evaluation and quality management. Systemic procedures to develop quality in education] In: Bundesministerium für Bildung, Wissenschaft und Kultur/Bundesinstitut für Erwachsenenbildung St. Wolfgang (Hg.): *Konzepte der Qualität in der Erwachsenenbildung* [Concepts of quality in adult education], Materialien zur Erwachsenenbildung Nr. 2/2001, Aufsätze und Protokoll im Rahmen der Werkstatt am Bundesinstitut für Erwachsenenbildung St. Wolfgang vom 2. bis 3. Oktober 2000. Verfügbar unter http://wwwapp.bmbwk.gv.at/medien/6048_PDFzuPubID88.pdf [20.04.2005], pp. 7-17.
- Deutsche Gesellschaft für Evaluation (DeGEval) (2002): *Standards für Evaluation*. [Standards for evaluation] Köln: Deutsche Gesellschaft für Evaluation.
- Deutsche Gesellschaft für Evaluation (DeGEval) (2004): *Empfehlungen für die Aus- und Weiterbildung in der Evaluation. Anforderungsprofile an Evaluatorinnen und Evaluatoren*. [Recommendations for education and training in evaluation. Required competencies for evaluators] Alfter: Deutsche Gesellschaft für Evaluation.

Deutsche Gesellschaft für Evaluation (DeGEval) (2005): *Empfehlungen zur Anwendung der Evaluationsstandards der DeGEval im Handlungsfeld der Selbstevaluation*. [Recommendations for application of the DeGEval evaluation standards to self-evaluation] Verfügbar unter <http://www.degeval.de/calimero/tools/proxy.php?id=139> [20.04.2005].

Hellstern, G.-M./Wollmann, H. (Hg.) (1984): *Handbuch zur Evaluierungsforschung Bd. 1*. [Handbook of evaluation research] Opladen: Westdeutscher Verlag.

Koch, U./ Wittmann, W. W. (Hg.) (1990): *Evaluationsforschung. Bewertungsgrundlage von Sozial- und Gesundheitsprogrammen*. [Evaluation research. Basis for assessment of social and health programs] Berlin/Heidelberg/New York: Springer.

Rossi, P. H./Freeman, H. E./Hofmann, G. (1988): *Programm-Evaluation. Einführung in die Methoden angewandter Sozialforschung*. [Program evaluation. Introduction to the methods of applied social science research] Stuttgart: Enke.

Scriven, M. (2004): Editorial: The Fiefdom Problem. In: *Journal of Multidisciplinary Evaluation*, No. 1 (Oct. 2004), pp. 11-18.

Stockmann, R. (Hg.) (2000): *Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder*. [Evaluation research. Foundations and selected fields] Opladen: Leske + Budrich.

Stockmann, R. (2002): Qualitätsmanagement und Evaluation – Konkurrierende oder sich ergänzende Konzepte?. [Quality management and evaluation –

- competing or complementary concepts?] In: *Zeitschrift für Evaluation*, 2/2002, pp. 209-243.
- Stockmann, R. (Hg.) (2004a): *Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder*. 2. Auflage. [2nd edition] Opladen: Leske + Budrich.
- Stockmann, R. (Hg.) (2004b): Evaluation in Deutschland. In: *Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder*. 2. Auflage. [2nd edition] Opladen: Leske + Budrich, pp. 13-43.
- Weiss, C. H. (1974): *Evaluierungsforschung. Methoden zur Einschätzung von sozialen Reformprogrammen*. [Evaluation research. Methods to assess social reform programs] Opladen: Westdeutscher Verlag.
- Wollmann, H. (1997): Evaluation in Germany. In: *European Evaluation Society, Newsletter* (3), pp. 4-5.
- Wottawa, H./Thierau, H. (1990): *Lehrbuch Evaluation*. [Textbook evaluation] Bern/Göttingen/Toronto: Huber.
- Wottawa, H./Thierau, H. (1998): *Lehrbuch Evaluation*. 2. vollst. überarb. Auflage. [2nd fully revised edition] Bern/Göttingen/Toronto: Huber.
- Wottawa, H./Thierau, H. (2003): *Lehrbuch Evaluation*. 3. korr. Auflage. [3rd corrected edition] Bern/Göttingen/Toronto: Huber.
- Wulf, C. (Hg.) (1972): *Evaluation. Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen*. [Evaluation. Description and Assessment of instruction, curricula and schooling trials] München: Piper.