

*German Evaluation Society
Wolfgang Beywl (ed.)*

**Selected Comments to the
Standards for Evaluation
of the German Evaluation Society
– English Edition -**

Edited by: German Evaluation Society Deutsche Gesellschaft für Evaluation e.V.
(DeGEval) / Wolfgang Beywl

Contact: DeGEval
Bücheler Weg 27
D-53347 Alfter

Tel: +49 228-96 49 99 09
Fax: +49 228-96 49 99 20
URL: <http://www.degeval.de>
Email: standards@degeval.de

Editor's note

This translation comprises the most essential parts of the booklet, which has been published by the German Evaluation Society (Deutsche Gesellschaft für Evaluation - DeGEval) together with the Standards for Evaluation, ratified in 2001. The complete booklet in German language can be ordered by the Internet (<http://www.degeval.de/standards/>).

As chair of the DeGEval's standards commission and member of the executive board I have accepted the responsibility for the selection carried out.

The translation was made possible by support of the European Centre for the Development of Vocational Training (CEDEFOP) in Thessaloniki, Greece. This is connected with the study 'Standards for Evaluation Practices - On the Way to Develop Standards for Evaluation in Vocational Education and Training contexts' I worked out together with my colleague Sandra Speer. The study will be published in 2004 as contribution to the third CEDEFOP-Report on vocational education and training research in Europe: 'Research on evaluation and impact of vocational education and training'.

I am very thankful to Pascal Descy and Manfred Tessaring from CEDEFOP for their support and to David Crabbe for the exact and linguistical sensitive translation from the German.

Wolfgang Beywl, Cologne and Berne, April 2003.

© DeGEval 2003.

List of Contents

- Introduction..... 4**
 - 1. Standard objectives..... 4
 - 2. Definition and forms of evaluation 5
 - 3. Application of the Standards 8
 - 4. Evolution of the Standards 8
 - 5. Further Action..... 9
- Explanatory notes to the Evaluation Standards 11**
- Bibliography..... 27**

Introduction

This paper consists of an introduction and the Evaluation Standards together with explanatory notes.

The introduction first presents the objectives of the Evaluation Standards. It then defines evaluation and summarises its forms. A brief section sketches how the standards shall be applied. The final part traces the evolution of the standards and shows the ideal direction for their future development.

Evaluations shall^(*) demonstrate the following four basic attributes: utility, feasibility, propriety and accuracy. The 25 standards are divided into these four categories.

1. Standard objectives

Evaluation has advanced considerably in Germany in recent years. Its profile is increasing in most areas of society, politics, business and research ⁽¹⁾.

This has prompted the German Evaluation Society (DeGEval) to formulate the Evaluation Standards. The 25 standards have a title and a description. The description includes up to three statements of intent. They target evaluators as well as individuals and organisations who commission evaluations and evaluand stakeholders. The standards are intended to assure and promote evaluation quality. They shall foster dialogue and provide a specific frame of reference for discussing the quality of professional evaluations. They are also designed to offer orientation for evaluation planning and implementation, to form a basis for initial and continuing training in this field and for the evaluation of evaluations (meta-evaluation) and to make professional practice more transparent for a wider public. The standards can assist evaluators when communicating with clients, addressees and a broad spectrum of stakeholders before, during and after specific evaluations.

This version of the standards deliberately avoids laying any binding foundations for accrediting and certifying individuals and organisations offering or conducting evaluations, or for further training in this sphere.

* I deliberately choose the verb 'shall' to translate the German 'sollen', which is mostly used in the standards statements. The indicative 'shall' underlines that it is obligatory to strive in evaluation planning and implementation for the conditions described by the standards statements. As an alternative the conditional 'should' would be too non-committal – it is *not* a free choice to take into account a standard or not. The verb 'must' on the other side is too directive and would put in charge an unconditional demand, which is not adequate for most of the standards. There are some standards in the group "accuracy" which *must* be adhered to; in these cases this verb is used in the translation. It is really enlightening to reflect on a professional translation of the standards text into another language and to be confronted with new requirements to revise the text in its original language (additional remark by W.B.).

1 Hellstern/Wollmann (1994) and Stockmann (2000) provide overviews on different periods. V. Wedel (1998) gives the Federal Court of Audit's perspective. See Bussmann/Klöti/Knoepfel 1997 on Switzerland. Grohmann (1997) offers a comprehensive review of developments in one particular evaluation field.

2. Definition and forms of evaluation

Evaluation is the systematic investigation of an evaluand's worth or merit. Evaluands include programmes, studies, products, schemes, services, organisations, policies, technologies and research projects. The results, conclusions and recommendations shall derive from comprehensible, empirical qualitative and/or quantitative data.

The great variety of evaluations, fields of application, tasks and fundamental concepts has generated numerous definitions, which differ in individual respects. Use of the term 'evaluation' is increasing in day-to-day speech in German, although frequently not in this technical sense, but simply to express that something has been tested, assessed or judged (Kromrey 2001). Professional evaluation strives for comprehensive assessment of the evaluand to ensure as precise a definition as possible of its merit and/or worth⁽²⁾. Systematic description of the evaluand paves the way for its assessment. Evaluators themselves, those commissioning the evaluation, those responsible for the evaluand (e.g. management or employees in organisations or programmes) or other stakeholders can perform this task. A joint assessment from various groups is also feasible.

Evaluation is systematic, i.e. an evaluation takes place for specific purposes⁽³⁾, and the intended purposes ideally prompt its implementation. Evaluators shall make a reasonable effort to determine the necessary information and data and lay solid foundations on which assessments can build.

Evaluation planning and compilation and processing of data shall give appropriate consideration to the relevant expertise and academic research findings in the field concerned.

Evaluation is supported by databases and employs a wide range of empirical and scientific methods, particularly the quantitative and qualitative methods of empirical social research.

Evaluation purposes, applied methods, factual basis, assessment and assessment foundations shall be comprehensible and capable of undergoing critical appraisal. Evaluation principles and steps shall be documented. Reports on the evaluation and its findings shall be supplied either in writing or in another appropriate form. Evaluations shall conform to the specialist standards described below. Evaluators shall subject their work to expert criticism.

Evaluations may focus on various aspects and hence have different purposes. 'Formative evaluation', which accompanies the shaping of the evaluand, is concerned primarily with encouraging improvement. It helps decision-makers and participants to improve the evaluand and its worth and to deploy resources optimally.

2 The terms 'merit' and 'worth' feature in the Joint Committee standards definition. They stem from a distinction made by Guba/Lincoln (1981), where 'merit' describes the intrinsic quality of an evaluand (e.g. concept stringency and technical safeguarding), and 'worth' reflects user value in specific situations at certain times. In this sense 'merit' is temporally and spatially unaffected, whereas 'worth' depends on the situation. Consequently 'worth' embodies multiple (even competing) applications for different users.

3 This text calls the intended uses of evaluations and their findings 'evaluation purposes'. The Commission has adopted the expression 'evaluation purposes' deliberately to distinguish them clearly from 'objectives'. These are typically formulated in the evaluand's field (e.g. as programme goals) and provide important points of reference and steering factors for 'goal-based' evaluations. A typical programme goal would be to trigger certain desirable conditions in a *programme's* target groups. The purpose of an *evaluation*, in contrast, could be to improve an existing programme, for example. This terminological distinction shall facilitate clear communication regarding evaluation purposes and objectives in the evaluand's field.

‘Summative evaluation’, on the other hand, constitutes a summarising appraisal of an evaluand. Its purpose is often to facilitate the making of fundamental decisions on the evaluand. Another potential evaluation purpose besides making improvements and paving the way for decisions is to glean findings without any direct application intent (cf. Standard U2). Both formative and summative evaluation can serve this end (⁴).

Evaluations are performed in highly diverse fields. The various focuses of the DeGEval topical interest groups reflect this diversity (⁵). The financial, temporal and spatial scope of evaluation projects and the number of people involved, whether evaluators or addressees, can vary widely. For example, interviewing a few individuals can provide a valuable and sufficient basis for assessments, improvements and decisions. However, a different case may require more sustained planning and a much more complicated organisation and programme development process. Both examples can conform to the Evaluation Standards.

Irrespective of differences, all evaluations involve (⁶):

- clients (⁷)
- sponsors (⁸)
- users, addressees, stakeholders (⁹)
- evaluation implementation teams (¹⁰)
- evaluation purposes (¹¹)
- evaluands
- evaluation locations
- values on which the evaluation and evaluand are based
- evaluation questions
- evaluation plans
- methods of inquiry (for surveys, processing and interpretation)
- assessments and findings, generally in written form.

There are many professional evaluation approaches. These have been reappraised synoptically in comparative studies (Wulf 1972, Beywl 1988, Owen/Rogers 1999, Uhl 1999, Kromrey

4 Cf. Scriven (1991) for more on the terms ‘formative’ and ‘summative’ and on other topics of specialist evaluation terminology.

5 When the final version of these standards was published in summer 2001, these were evaluation of higher education institutions; schools; environmental bodies; vocational and in-company training; development policy; research, technology and innovation policy; media; social services; health, administration and evaluation; structural policy. This is intended to be an illustrative list rather than an exclusive one. See also <http://www.degeval.de/arbeitskreise.htm>.

6 Wottawa/Thierau (1998) examine some of these aspects in the chapter ‘Design aspects of evaluative studies’, pp. 55-66.

7 The evaluation team can commission the evaluation themselves, notably in the case of self-evaluation.

8 These can, but do not have to be the clients.

9 Evaluation ‘users’ are mentioned first. People who use the evaluation and its findings as the evaluation plan foresees are the ‘addressees’. Not all addressees actually use the data. Finally, stakeholders are addressed. There are two types of stakeholders. ‘Active participants’ include individuals, groups and organisations which play an active role in connection with the evaluand, e.g. programme sponsors, programme managers and paid or voluntary players within the evaluand’s context. ‘Affected parties’ generally have little influence. They often belong to a programme’s target groups, for example, and can be excluded from a programme’s benefits or disadvantaged by it, sometimes without even knowing of its existence. Participatory evaluation approaches in particular may try to transform affected parties into active participants. The distinction between the former and the latter is blurred. The groups specified often overlap. Workers from organisations under evaluation are generally involved or affected, but may simultaneously be addressees who also profit from it. A counter-example is a foundation which sponsors evaluations and receives the final report, but which has no direct interest in the evaluand and therefore is neither an active participant nor an affected party. The same applies when the addressee is the general public or specialists. These designations are analytical distinctions. They indicate whether individuals and organisations are addressed primarily with reference to the evaluand or to the evaluation. They shall encourage those responsible for the evaluation to clarify whom the various evaluation activities address and what role they have.

10 In many cases teams consisting of two or more people are responsible for an evaluation. Sometimes one evaluator shallers all the responsibility. The following text includes these individuals when referring to evaluation teams.

11 cf. footnote 3!

2000, Uhl 2000, Stufflebeam 2001). They can be distinguished according to the following criteria:

- the principles of recognition theory the evaluation demonstrates (critical-rational, hypothesis-testing, pragmatic, constructivist, etc.)
- the participation concepts relating to evaluation and its utilisation (evaluation approaches with competing 'legal teams', parliament-based procedures tailored to legal decision-makers, methods strongly or moderately geared to foster participation, fundamentally democratic approaches, etc.)
- the (dominant) reference disciplines involved (economics, sociology, political science, educational science, psychology, ethnology, engineering, etc.)
- the intended purposes of the evaluation (e.g. preparing decisions on the evaluand, improving organisations or programmes, supporting learning and reflection processes, broadening knowledge in a particular field)
- what the evaluation shall achieve (preformative – developing, formative – shaping, summative – appraising)
- the evaluation's central steering factors (e.g. the explicit purpose of the evaluation, merits for evaluand stakeholders, objectives of the evaluated programme, *ad-hoc* or systematically derived hypotheses, stakeholder interests, cost-benefit ratios)
- the stage of evaluand development (development and routine phase) the evaluation targets (proactive, clarifying, interactive, documenting, assessing impact / *ex ante*, ongoing, *ex post*)
- the evaluand's scope and complexity
- which dimension of the evaluand receives priority (context, structure, concept, input, process or impact)
- whether the evaluation is conducted externally, i.e. by a commissioned outside evaluation team, or by the people in charge of the evaluand (self-evaluation), or internally by a member of the organisation with no responsibility for the evaluand.

Both the list of distinctions and the characteristics mentioned under the individual dimensions are incomplete. As a rule individual evaluations are mixed forms which combine the different characteristics. It is important that the concrete evaluation format is optimally adapted to the evaluation purpose, the evaluand's characteristics, the specific conditions and options and the available resources (time and money) and that it is planned and tailored with this in mind.

Self-evaluations are also possible. They are characterised by the fact that evaluand decision-makers conduct and often commission them themselves (Heiner 1998, von Spiegel 1993, Allgäuer 1997, Bühren/Killus/Müller 1998). They often enlist the help of external or internal evaluation teams. The standards described below only apply to self-evaluations to a limited extent, particularly since the latter often feature different role and interest constellations. DeGEval will tackle the issue of standards which also or specifically apply to self-evaluations in the course of advancing and differentiating the Evaluation Standards (Müller-Kohlenberg/Beywl 2001).

3. Application of the Standards

The DeGEval standards are designed to uphold and raise the quality of evaluations. They formulate key points which evaluators shall respect and goals they shall pursue. They are intended to provide a frame of reference for conducting and assessing evaluations. How they are implemented is a deciding factor. It cannot take place schematically. The quality standards are not intended to devalue evaluations which do not meet a particular standard in a certain way. Some standards will not be applicable to certain evaluations. Brief grounds for not applying a standard shall be provided in such a case.

When assessing the quality of evaluations, we shall ask whether the applicable standards were considered when planning and conducting the evaluation and whether enough care was taken to observe them within the bounds of the relevant conditions.

The evaluation will often need to weigh up the advantages and disadvantages of various alternatives. It is not always possible to give all standards equal consideration. From time to time requirements will contradict each other⁽¹²⁾. The evaluation team and all participants have the job of finding an appropriate solution which takes account of the purposes and context of the evaluation in hand.

The standards apply to evaluations in general, but not to the evaluation of individuals, for example in performance judgement processes or employee assessment (staff evaluation is excluded).

4. Evolution of the Standards

In 2000 the German Evaluation Society surveyed its members to obtain a clear picture of opinions on producing evaluation standards and guidelines for evaluators. The vast majority advocated prioritising the formulation of evaluation standards. Many members also wanted or favoured guidelines comparable with those of the American Evaluation Association (AEA 1995) or the Canadian Evaluation Society (no year specified), which focus on evaluator activities, behaviour and competence. The DeGEval finally passed the decision to formulate evaluation standards at its 2000 annual general meeting in Berlin. It called on the board to draft a work programme and form a commission.

The decision of the 2000 annual general meeting and the work programme incorporates the task of yoking the DeGEval standards with those of the Joint Committee on Standards for Educational Evaluation (JC Standards 2000/1994) and the related Swiss Evaluation Society standards (SEVAL Standards 2001). The DeGEval standards thus build on over 20 years of experience. The meeting deliberately rejected a complete reformulation and restructuring of the standards to facilitate international exchange and to profit from the existing material and specialist publications from the U.S. and Swiss evaluation fields (ibid. Joint Committee 2000).

A commission drafted the proposal for the DeGEval's Evaluation Standards. It consisted of seven evaluators, six men and one woman, and two male client representatives. The nine commission members come from various fields of application and academic disciplines. The draft underwent several revisions and was examined by 13 expert commentators. The

12 See Hager/Patry/Brezing (2000) on such dilemmas.

DeGEval board studied the final version before passing it on to the annual general meeting. They advised the AGM to approve the resolution. Before this meeting all DeGEval members received a copy of the Evaluation Standards for perusal so that they could form an opinion and make an informed decision on whether to adopt the resolution. Following the AGM's unanimous decision on 4 October 2001, the board was called upon to publicise the Evaluation Standards and the commission's annotations and introduce them into expert circles and public debate.

5. Further Action

Following the annual general meeting resolution, the DeGEval will continue to promote widespread discussion of the standards. Coordination with scientific societies, vocational and specialist associations from a broad spectrum of disciplines and other groups interested in evaluation is also essential.

Some evaluation fields have specific regulations which apply both internationally and in the German-speaking world, irrespective of the standards. They include the DAC Principles for Evaluation of Development Assistance (OECD 1998) and the Public Management Service Best Practice Guidelines for Evaluation (PUMA OECD Guidelines 1998a). The MEANS Handbook No. 5 – Quality Assessment of Evaluation Reports, European Union – Regional Policy and Cohesion (1996) is based on the original U.S. version of the Joint Committee Standards (2000). The recommendations for drug and drug addiction evaluation (EMCDDA 1998), PR evaluation (DPRG 2000) and youth support evaluation (Beywl et al. 2001) were created for specific fields.

The DeGEval and its topical interest groups are called to discuss how the Evaluation Standards relate to these and other key regulations and issue statements from an expert perspective. Revision of the standards shall take these existing and emerging regulations into account. The DeGEval special interest groups are highly instrumental in harmonising the standards with those from specific policy areas.

Ideally a revised version of the Evaluation Standards shall contain detailed explanations and advice on applying the individual standards, e.g. in the form of leading questions. It shall draw attention to potential misuse, and present and discuss case studies which demonstrate clearly how each standard shall be applied. Finally, it shall incorporate a glossary defining key terms and explaining terminology ⁽¹³⁾.

Comments and opinions on the standards are always welcome.
Please send them to the German Evaluation Society office,
preferably by email (**standards@degeval.de**).
You will find the current postal address under Contact on the DeGEval website
(**<http://www.degeval.de>**).

13 The Joint Committee (2000) publication can serve as a model.

This shall result in an extensive document for controlling and estimating evaluation quality in the foreseeable future. The DeGEval board will continue to encourage both members and non-members interested in evaluation to study and comment on the standards. As soon as sufficient proposals and comments are available, the board will set a standards revision procedure in motion. This shall be concluded by the end of 2004.

Members of the DeGEval's standards commission

Dr. Wolfgang Beywl (chair)	Dieter Brauns	Dr. Hansjörg Drewello
Dr. Andreas Hellmann	Thomas Kuby	Sabine Müller
Dr. Alfred Uhl	Gerald Wagner	Hein Winnubst*

** Until August 2001*

Expert commentators

Dr. Alois Basler	Karin Haubrich	Dr. Christian Lüders
Alexandra Caspari	Thomas Hochgesang	Dirk Scheffler
Dr. Karin Fischer-Bluhm	Ursula Hütte	Prof. Dr. Reinhard Stockmann
Werner Fuchs	Marc Jelitto	Gerlinde Struhkamp
	Dr. Kathleen Toepel	

* *bis August 2001

Explanatory notes to the Evaluation Standards

Utility	12
U1 Stakeholder Identification	12
U2 Clarification of the Purposes of the Evaluation.....	13
U3 Evaluator Credibility and Competence	13
U4 Information Scope and Selection.....	13
U5 Transparency of Values.....	14
U6 Report Comprehensiveness and Clarity	14
U7 Evaluation Timeliness	15
U8 Evaluation Utilisation and Use	15
Feasibility	17
F1 Appropriate Procedures.....	17
F2 Diplomatic Conduct	17
F3 Evaluation Efficiency.....	18
Propriety	19
P1 Formal Agreement	19
P2 Protection of Individual Rights	19
P3 Complete and Fair Investigation	20
P4 Unbiased Conduct and Reporting	20
P5 Disclosure of Findings	21
Accuracy.....	22
A1 Description of the Evaluand	22
A2 Context Analysis	22
A3 Described Purposes and Procedures	23
A4 Disclosure of Information Sources	23
A5 Valid and Reliable Information	23
A6 Systematic Data Review.....	24
A7 Analysis of Qualitative and Quantitative Information	25
A8 Justified Conclusions.....	25
A9 Meta-Evaluation	26

Utility

The Utility Standards are intended to ensure that an evaluation is guided by both the clarified purposes of the evaluation and the information needs of its intended users.

U1 Stakeholder Identification

Persons or groups involved in or affected by the evaluand shall be identified, so that their interests can be clarified and taken into consideration when designing the evaluation.

The following questions can help determine which players an evaluation and its evaluand shall take into consideration:

- Who decides the evaluand's future?
- Who is responsible for the evaluand's conception and design?
- Who is involved in implementing the evaluand under investigation?
- Who are the evaluand's direct and indirect focuses (target groups and their social environment)?

The individuals, groups and organisations thus identified are labelled 'stakeholders'. Factors such as whom the evaluand might disadvantage and who could be excluded from a programme shall also be determined. As far as finances and time allow, the identified stakeholders shall be involved in planning and conducting the evaluation in line with the evaluation purposes (Heiner 1998).

It is particularly crucial to ascertain how much information addressees of the evaluation and its findings require, to consult them when clarifying purposes and questions and to tailor the evaluation accordingly. Other interested parties shall also be considered (e.g. decision-makers planning similar projects, specialists and the general public).

An evaluation which stakeholders help shape and which is tailored to their information requirements has optimal chances of winning addressee acceptance and actually being used.

U2 Clarification of the Purposes of the Evaluation

The purposes of the evaluation shall be stated clearly, so that the stakeholders can provide relevant comments on these purposes, and so that the evaluation team knows exactly what it is expected to do.

The evaluation purposes shall be discussed at an early stage. In the course of the evaluation they shall be negotiated and put in writing. Stakeholders shall receive as much information as possible so that they can participate in planning. A clear purpose facilitates the task of the person or institution commissioning the evaluation and of the evaluation team because it provides a sound basis for distributing the findings and encouraging people to use them.

The ideal typical distinction between three main evaluation purposes is helpful for clarification (Patton 1997, p. 79):

- providing information suitable for guiding the step-by-step design of the evaluand, e.g. for optimising concepts and processes;
- providing information suitable for guiding a fundamental decision on the evaluand, e.g. introduction, continuation, expansion or cessation of a programme;
- providing findings to fuel public, scientific or political debate.

The first two key purposes focus on instrumental use, whereas gleaned findings concerns conceptual use (Cronbach et al. 1980, p. 112 ff.). Simultaneous and equal pursuit of several key purposes can impair the use of the evaluation. It is therefore often advisable to set one clear priority and process various key purposes in separate phases or to divide the work between different teams.

U3 Evaluator Credibility and Competence

The persons conducting an evaluation shall be trustworthy as well as methodologically and professionally competent, so that the evaluation findings achieve maximum credibility and acceptance.

Evaluator credibility affects evaluation implementation and impact considerably. Evaluators shall demonstrate the following qualities if the various affected groups are to find them credible: professional and methodological competence, integrity, independence and social and communication skills. These qualities shall play a role in evaluator selection. It is often helpful to form an evaluation team to ensure that the necessary skills are available.

U4 Information Scope and Selection

The scope and selection of the collected information shall make it possible to answer relevant questions about the evaluand and, at the same time, consider the information needs of the client and other stakeholders.

Evaluation study planning shall consider what information is essential to answer the questions and what may be interesting and desirable but is irrelevant to the key topics. Care shall be taken to assign the existing data-collection resources according to priorities in addressing the core questions and the information needs of the most important potential evaluation users (Federal Health Office 1997).

U5 Transparency of Values

The perspectives and assumptions of the stakeholders that serve as a basis for the evaluation and the interpretation of the evaluation findings shall be described in a way that clarifies their underlying values.

Stakeholder values play a major role in several evaluation phases, right from establishing evaluation purposes and formulating questions. It is vital for the evaluation to identify and focus on these to encourage stakeholders to play an active part in it and to utilise its findings (Wottawa/Thierau 1998). Interpretation of the collected information and findings in the final phase is one of the most important and critical parts of the evaluation process. Societal values (norms) necessarily play a major role in this. The underlying values shall be as transparent as possible so that interpretation is convincing, comprehensible and assessable.

U6 Report Comprehensiveness and Clarity

Evaluation reports shall provide all relevant information and be easily comprehensible.

Successful communication of evaluation findings demands comprehensiveness and clarity in the written report (or other form of feedback). Addressees shall be able to understand the language, and it shall define important terms and use them consistently. Compiling the most important findings – e.g. in an introductory summary, a table or diagrams, can improve comprehension of the report significantly.

The design and nature of reports shall be geared to the receptiveness of the evaluation addressees. A detailed final report is not the best way of transmitting information for every target group. Sometimes presentations, workshops or similar communication forms are more appropriate, depending on the situation and the intended audience.

U7 Evaluation Timeliness

The evaluation shall be initiated and completed in a timely fashion so that its findings can inform pending decision and improvement processes.

The date by which evaluation preparations are finished, invitations for tenders issued, if necessary, and initial activities in the evaluation sphere commissioned and commenced shall allow for completion of the necessary tasks by the time the evaluation findings are required. The schedule shall allocate a realistic period for essential coordination processes, data compilation preparations, monitoring, information processing and interpretation. It shall include enough leeway to cope with unforeseen events.

Users shall be informed of significant interim results and final reports so that they can apply them in good time. Experience has shown that an evaluation study has more impact when its progress is coordinated with planned decision-making and/or improvement processes. Otherwise it loses much of its effectiveness. It is important to remember that many cases (e.g. a commission from public administration) shall allow considerable processing time, as the evaluation report shall be handled internally (e.g. countersigning procedures, statements), before decisions can be made and steps initiated. Submitting interim results and reports during the course of investigations is advisable for many evaluation studies, especially when the information could influence addressees' next move. We shall consider such feedback patterns when planning the evaluation and allocate appropriate resources to this end.

U8 Evaluation Utilisation and Use

The evaluation shall be planned, conducted, and reported in ways that encourage attentive follow-through by stakeholders and utilisation of the evaluation findings.

Even results acquired by very professional means may often not be used at all or used quite differently than the evaluation agreements planned. High evaluation feasibility, propriety and accuracy and the above-mentioned aspects of the seven utility standards are important prerequisites for any intended application. Implementing conclusions or recommendations depends particularly on whether addressees expect the evaluation to benefit them and their study. Often different addressees also have different utilisation expectations which may be contradictory or even mutually exclusive. An important prerequisite for promoting the agreed or intended utilisation and hence the use of evaluations is the appropriate involvement of the various addressees in planning and implementing the evaluation and preparing findings. Continuous and transparent feedback during the course of the investigation also has a positive

effect (Stockbauer 2000). This fosters process use during the evaluation, regardless of the degree to which the findings are used (Patton 1998).

Feasibility

The Feasibility Standards are intended to ensure that an evaluation is planned and conducted in a realistic, thoughtful, diplomatic and cost-effective manner.

F1 Appropriate Procedures

Evaluation procedures, including information collection procedures, shall be chosen so that the burden placed on the evaluand or the stakeholders is appropriate in comparison to the expected benefits of the evaluation.

Evaluation processes shall meet scientific merit criteria while not unnecessarily burdening or imposing on the evaluand or stakeholders. The most relevant methods from a scientific point of view are often unsuitable because they are too time-consuming or costly or ethically unacceptable for the situation concerned (Uhl 2000). The evaluation team shall clarify advantages and disadvantages and justify the relevance of the chosen procedure.

Diplomatic Conduct

The evaluation shall be planned and conducted so that it achieves maximal acceptance by the different stakeholders with regard to the evaluation process and findings.

Ensuring balanced consideration of all stakeholder interests when implementing the evaluation fosters acceptance, approval and cooperation among the various parties. This shall eliminate the potential for any one of these groups to try to restrict evaluation activities or distort or abuse the findings. The evaluation team shall proceed diplomatically to avoid such an outcome.

Political viability and willingness to compromise are vital if evaluation findings in political and organisational contexts are to find subsequent use (Wottawa/Thierau 1998, Hager/Patry/Brezing 2000, Beywl 2001, Faßmann 2001). However, politicians often commission evaluations to shift responsibility for difficult decisions or to belatedly legitimise steps already taken. In such cases the evaluation team shall promote dialogue between decision-makers and other stakeholders.

If the team succeeds in generating a cooperative attitude among the individual interest groups, it will positively affect their willingness to participate in the evaluation process, provide information, accept the findings and use them where appropriate.

F3 Evaluation Efficiency

The relation between cost and benefit of the evaluation shall be appropriate.

When commencing and concluding an evaluation, it is often hard to make precise statements on a project's cost and benefit (¹⁴). However, these difficulties must not lead to a total disregard of cost and benefit considerations. The decision on whether or not to conduct an evaluation shall certainly involve a cost-benefit assessment. Evaluation planning shall present a clear estimate of the predicted time and cost and the expected advantages.

The total evaluation costs comprise the monetary value of all necessary resources, e.g. evaluator remuneration, stakeholder time investment, travel expenses and materials. Other costs also arise, including those incurred by third parties. Many costs are not quantifiable.

Quantifying the benefits is even more difficult. As a rule we can only estimate them. Advantages can result from direct and indirect, intended and unintended effects of the evaluation.

14 Cf. the fundamental criticism of cost-benefit analyses from an economic perspective in Rürup/Hansmeyer (1984, p. 107ff).

Propriety

The Propriety Standards are intended to ensure that in the course of the evaluation all stakeholders are treated with respect and fairness.

P1 Formal Agreement

Obligations of the formal parties to an evaluation (what is to be done, how, by whom, when) shall be agreed to in writing, so that these parties are obligated to adhere to all conditions of the agreement or to renegotiate it.

Formal agreements on the evaluation shall be founded on mutual respect and trust and settled at least in the following areas: finances, timing, methodology and participants. Decision-makers' and participants' rights and obligations particularly shall be specified as precisely as possible. A formal, written agreement reduces the probability of misunderstandings between the contracting parties and helps resolve them if they occur.

We shall remember that rigid stipulations for question detail, methods and procedures can easily become straitjackets which obstruct the gathering of findings in the long term. If changes prove necessary over time, it is possible to justify deviations and renegotiate conditions, but this is easier if the evaluation team informs all parties to the agreement of the potential need for adaptation at the start.

P2 Protection of Individual Rights

The evaluation shall be designed and conducted in a way that protects the welfare, dignity and rights of all stakeholders.

Evaluators shall ensure that they do not encroach on the dignity and self-respect of the people with whom they interact in the course of the evaluation⁽¹⁵⁾. If an evaluation exposes weaknesses, it can seriously undermine the protection of interests. All interests shall be weighed up carefully in these cases. This is particularly crucial when evaluations unmask legal offences.

15 For these standards cf. AERA/APA/NCME (1999), German Society for Educational Science 1997, IHK/ESOMAR (no date specified), the Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V. (German Market Research Society) et al. (2001) and the Internet (<http://www.datenschutz-berlin.de>).

P3 Complete and Fair Investigation

The evaluation shall undertake a complete and fair examination and description of strengths and weaknesses of the evaluand so that strengths can be built upon and problem areas addressed.

It goes without saying that the evaluand's weaknesses shall be assessed and presented in detail, otherwise they cannot be eliminated. However, it is also important to record strengths. Both are necessary to assess the evaluand comprehensively and to improve it.

An evaluand's strengths and weaknesses are often closely related. We shall remember that correcting weaknesses can sometimes undermine existing strengths. However, existing strengths may also counteract weaknesses. Often both the development of strengths and the intended elimination of weaknesses underlie an evaluation. Furthermore, stakeholders are usually more willing to cooperate and accept the evaluation if it also documents strengths.

Assessment and presentation of strengths and weaknesses shall be as comprehensive as possible to highlight all significant aspects. They shall treat all stakeholders fairly. Clients and other active participants shall not attempt to influence the evaluation and the report. Evaluators are responsible for ensuring a complete, balanced and fair evaluation.

P4 Unbiased Conduct and Reporting

The evaluation shall take into account the different views of the stakeholders concerning the evaluand and the evaluation findings. Like the entire evaluation process, the evaluation report shall evidence the impartial position of the evaluation team. Value judgements shall be made as unemotionally as possible.

Various perspectives mould the evaluation context. Stakeholders often have divergent views on the evaluand. There is a danger that a particular group could take over or exploit evaluations.

Evaluations shall avoid adopting one specific perspective. They shall strive to treat all pertinent interests fairly. The evaluation team shall therefore be as unbiased as possible. It shall also avoid too close a relationship with either clients or those responsible for the evaluand. Evaluators' relationships to evaluation clients and other relevant groups shall be clarified from the outset. This includes decisions on publishing evaluation reports (Müller-Kohlenberg / Münstermann 2000).

P5 Disclosure of Findings

As far as possible, all stakeholders shall have access to the evaluation findings.

In principle everyone participating in or affected by an evaluation shall have access to the findings and the report. All parties (e.g. evaluation team, clients, addressees, employees from the evaluated organisation) are responsible for ensuring this. As the group of stakeholders is often very large, it may be necessary to publicise the report.

Disclosure of evaluation findings shall enhance the evaluation's benefit. In some cases, however, complete disclosure could reduce an evaluation's worth, e.g. if rival firms could use the findings to the detriment of an evaluated company, if stakeholders' willingness to participate in evaluating and implementing findings could diminish due to their knowledge of subsequent publication (of weaknesses) or if there is a risk that publication could compromise information providers. In such cases it is the players' task to find a common solution and agree on limited disclosure if necessary. Any restrictions on disclosure shall be justified.

Publication of evaluation findings, their nature and their scope shall be agreed at the beginning of an evaluation and documented in contract form. If this is not possible, it shall be agreed at the outset who shall disclose the findings, how, according to which criteria and when (e.g. at the end of an evaluation study).

Accuracy

The Accuracy Standards are intended to ensure that an evaluation produces and discloses valid and useful information and findings pertaining to the evaluation questions.

A1 Description of the Evaluand

The evaluand shall be described and documented clearly and accurately so that it can be unequivocally identified.

The evaluand, whether it is an initiative, a programme or an organisation, shall be described precisely. We must remember that the evaluand can take different forms depending on various temporal and spatial contexts. The description of the evaluand shall clearly show what is under investigation. This allows addressees to compare it with other evaluands. A precise evaluand description helps identify connections between the evaluand and its impact and may reveal previously disregarded side-effects. We shall observe the degree of discrepancy between the evaluand's original form and its actual implementation.

A2 Context Analysis

The context of the evaluand shall be examined and analysed in sufficient detail.

The evaluand's context is provided by its environment. It includes the social and political climate, the characteristics and interests of the chief stakeholders and the economic situation. Analysis of this environment is vital for gleaning information on conditions which could affect the evaluand and the evaluation results. A sound context analysis also permits assessment of the transferability of evaluation results to other contexts. We shall avoid defining the context too rigidly. However, we shall also beware of analysing the context in too much detail, as this can detract from processing the other evaluation stages.

A3 Described Purposes and Procedures

Object, purposes, questions and procedures of an evaluation, including the applied methods, shall be accurately documented and described so that they can be identified and assessed.

To ensure transparency, the evaluation purposes shall be documented in the course of the process and communicated to addressees clearly in the report. The same applies to the questions under consideration, the chosen procedures, the methods and the reasoning behind the decisions underlying the evaluation plan. The schedule and any deviations shall be recorded in writing.

The description of evaluation purposes and questions shall take contrasting viewpoints into consideration. Documentation of the process incorporates a detailed description of the organisational precautions, investigations, including any sample-taking, data processing and analysis, interpretation and finally, reporting. We shall remember that the procedure can change in the course of the evaluation and that the actual process can thus differ from the plan. Any deviations and the reasons for them shall be explained clearly.

A4 Disclosure of Information Sources

The information sources used in the course of the evaluation shall be documented in appropriate detail so that the reliability and adequacy of the information can be assessed.

Clear description of the information sources used allows addressees to form their own opinion on their quality. Evaluation information sources include individuals and groups, documents, audiovisual materials and statistics. Consultation of various sources permits comparison of the data retrieved. Missing or insufficient data on an information source can cast doubt on an evaluation's credibility. Description of the sources must go hand in hand with assessment and evaluation of the information they contain (German Research Committee – DFG - 1999).

A5 Valid and Reliable Information

The data collection procedures shall be chosen and developed and then applied in a way that ensures the reliability and validity of the data with regard to answering the evaluation questions. The technical criteria shall be based on the standards of quantitative and qualitative social research.

For empirical analysis of the evaluand, the data collection tools applied and the data gathered must comply with certain merit criteria to minimise the effect of erroneous sources on the collection process.

The central merit criteria for quantitative methods state that measurements must take place independently of the person using the instrument (objectivity). The tools shall retrieve consistent, reproducible and reliable information and be as resistant as possible to interruptions and random errors (unsystematic error sources) (reliability and measurement precision). Tools must actually record the characteristics and behavioural patterns they claim to measure (validity). Objectivity and reliability are necessary but insufficient conditions for valid measurements (AERA/APA/NCME 1999, Dieckmann 1995, Bortz/Döring 1995, Häcker et al. 1998).

Often evaluations cannot have recourse to standardised, quality-controlled quantitative tools with known reliability and validity parameters. Qualitative access is often necessary and is more appropriate when applied to evaluands and questions. To ensure the reliability and validity of non-standardised instruments, qualitative data and subjective interpretations, specific merit criteria for qualitative research have been developed (e.g. 'Process Documentation', 'Interpretation Support Argumentation', 'Evaluand Proximity' and 'Triangulation' according to Mayring 1999).

Validity concerns not only the quality of data collection tools and data, but also the merit and credibility of the conclusions drawn from the investigation. In this sense we can assess validity only within the particular evaluation context and for the specific objectives and questions of the empirical compilation. Validation is the compilation of evidence which supports the data-based interpretation. Several quantitative and/or qualitative processes for gathering data shall therefore be applied. Conclusions from all procedures adopted shall be validated individually and in combination. The scientific merit criteria presented here form the basis for decisions on selecting, developing and implementing methods and shall be specified and expanded according to the demands of evaluation practice (Hager/Patry/Brezing 2000).

A6 Systematic Data Review

The data collected, analysed and presented in the course of the evaluation shall be systematically examined for possible errors.

Collecting, processing, assessing and interpreting information and presenting findings creates a wealth of potential pitfalls. These can be methodological errors in the narrowest sense (examples from quantitative research are: incorrect measurements, errors in the investigation design due to insufficient monitoring of interference variables, distortion of the investigation situation due to evaluation team influence, errors in sample surveys, disregard for the application criteria on statistical procedures, etc.). They may also involve a lack of care (confusion of survey ID numbers, documentation and audiovisual recordings, incorrect entry of data for processing, switched labels in presentation of findings, etc.).

It is therefore crucial to design the evaluation process so that potential pitfalls can be identified at an early stage and errors avoided or corrected as far as possible. Systematic training of participants and systematic monitoring and precision examinations (plausibility tests, parallel processing, communicative validation, etc.) can achieve this. The evaluation report must prompt clear discussion of possible errors and their consequences. Mistakes lead to erroneous interpretations and can invalidate the whole evaluation.

A7 Analysis of Qualitative and Quantitative Information

Qualitative and quantitative information shall be analysed in an appropriate, systematic way so that the evaluation questions can be effectively answered.

The data analysis process sorts, summarises and assesses the quantitative and qualitative data gathered during evaluations. This forms the basis for interpretations and conclusions in the process of answering the evaluation questions. Selection of appropriate survey and analysis procedures shall be based on the evaluation questions, the current level of information on the evaluand and context variables in the evaluation field. Evaluator preferences shall play no role in this decision. The information gathered shall be assessed systematically. Independent merit criteria and regulations shall apply to qualitative procedures (e.g. Mayring 1999, Flick 2001, Lamneck 1995, Miles/Huberman 1994) and quantitative procedures (e.g. Bortz/Döring 1995, Bortz 1999, Dieckmann 1995, Kromrey 1998). The applicable regulations and their methodological foundations shall be made available for consultation, e.g. via references to relevant literature. Choice and application of procedures shall be transparent and comprehensible so that selection decisions and findings can undergo critical appraisal. Benchmarks and formulae shall be explained in a way that everyone can understand. The value and limitations of the methods must be stated explicitly.

A8 Justified Conclusions

The conclusions reached in the evaluation shall be explicitly justified so that the audiences can assess them.

Evaluation conclusions shall be substantiated and presented clearly with the fundamental suppositions and the procedures applied. The scope of the conclusions shall also be emphasised. Where necessary, reports shall incorporate a discussion of alternative interpretations and the reasons why these were rejected. Adherence to these standards allows users of the findings to estimate the value of the conclusions. It also reinforces the conclusions' credibility.

A9 Meta-Evaluation

The evaluation shall be documented and archived appropriately so that a meta-evaluation can be undertaken.

Unsuccessful evaluations can prompt erroneous decisions. Evaluations can also provoke sharp and sometimes unjustified criticism. The evaluation itself shall be evaluated to avoid such situations. These meta-evaluations are devised to assess and improve the quality of evaluation processes and findings. The standards presented here can be applied to meta-evaluations (Widmer 1996). Either the evaluation team (internal) or outsiders (external) can conduct a systematic meta-evaluation. A comprehensive and in-depth meta-evaluation is only viable in certain cases, but those responsible for the evaluation shall conduct a brief self-evaluation as a matter of course. Regularly conducting meta-evaluations will increase the credibility of individual evaluations and raise the status of evaluating as a profession.

The evaluation team shall document and archive the evaluand, the key purposes, steps, methods, data and findings of an evaluation study to permit meta-evaluations, evaluation syntheses and meta-analyses. This fosters scientific progress and knowledge accumulation in the relevant evaluand field and for models and methods of evaluation research.

Any scientific publications on evaluation studies and their procedures, problems and findings are also welcome. They promote evaluation progress and the dissemination of knowledge in the evaluand field, boost quality development and encourage evaluation distribution and acceptance.

Bibliography

- AEA** American Evaluation Association (1995): Guiding Principles for Evaluators (published in any issue of the American Journal for Evaluation).
- AERA/APA/NCME** American Educational Research Association/American Psychological Association/National Council on Measurement in Education (1999): The Standards for Educational and Psychological Testing. Washington.
- Allgäuer, R.** (1997): Evaluation macht uns stark! Frankfurt.
- Arbeitskreis** Deutscher Markt- und Sozialforschungsinstitute e.V./ Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e.V./ Berufsverband Deutscher Markt- und Sozialforscher e.V./ Deutsche Gesellschaft für Online-Forschung e.V. (D.G.O.F.). (2001): Standards zur Qualitätssicherung für Online-Befragungen. Frankfurt (<http://www.adm-ev.de>).
- Beywl, W.** (1988): Zur Weiterentwicklung der Evaluationsmethodologie. Grundlegung, Konzeption und Anwendung eines Modells der responsiven Evaluation. Frankfurt (Reprint, Cologne 1999).
- Beywl, W.** (2001): Konfliktfähigkeit der Evaluation und die "Standards für Evaluation"; in: Sozialwissenschaften und Berufspraxis Nr. 2/2001, pp. 151-164.
- Beywl, W. u. a.** (2001) (ed.): Evaluation im Alltag. Münster.
- Bortz, J.** (1999): Statistik für Sozialwissenschaftler. Berlin.
- Bortz, J./Döring, N.** (1995): Forschungsmethoden und Evaluation. Berlin.
- Buhren, C. G./Killus, D./Müller, S.** (1998): Wege und Methoden der Selbstevaluation. Ein praktischer Leitfaden für Schulen. Dortmund.
- Bundesamt für Gesundheit (BAG)** (1997) (Hg.): Leitfaden für die Planung von Projekt- und Programmevaluationen. Berne (Bundesamt für Gesundheit, Fachbereich Evaluation).
- Bussmann, W./Klöti, U./Knoepfel P.** (1997): Einführung in die Politikevaluation. Zurich.
- Canadian Evaluation Society** (not specified): CES Guidelines for Ethical Conduct. Ottawa (<http://www.evaluationcanada.ca>).
- Cronbach, L.J. u.a.** (1980): Toward Reform of Program Evaluation. San Francisco.
- Deutsche Gesellschaft für Erziehungswissenschaft** (1997): Standards erziehungswissenschaftlicher Forschung; in: Friebertshäuser, B./Prengel A. (ed.): Handbuch qualitative Forschungsmethoden in der Erziehungswissenschaft. Munich. pp. 858-863.
- Deutsche Norm DIN 33430.** Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen. Draft. October 2000. Berlin (www.mybeuth.de).
- DFG - Deutsche Forschungsgemeinschaft/ Kaase, M.** (1999) (ed.): Qualitätskriterien der Umfrageforschung – Denkschrift. Berlin.
- Diekmann, A.** (1995): Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen. Reinbek.
- DPRG** (Deutsche Public-Relations-Gesellschaft), GPRA (Gesellschaft Public Relations Agenturen (2000): PR-Evaluation. Messen – Analysieren – Bewerten. Bonn.
- EMCDDA** (1998): Guidelines for the Evaluation of Drug Prevention: A Manual for Programme-Planners and Evaluators, Manuals, European Monitoring Centre for Drugs and Drug Addiction, Lisbon.

- European Union - Regional Policy and Cohesion** (1996): MEANS-Handbook Nr. 5. Quality Assessment of Evaluation Reports: A Framework. Brussels.
- Faßmann, H.** (2001): Probleme der Evaluation von Programmen im Bereich der Rehabilitation; in: Sozialwissenschaften und Berufspraxis Nr. 2/2001, S.133-149.
- Flick, U.** (2001): Qualitative Forschung. Theorie, Methoden, Anwendung in Psychologie und Sozialwissenschaften. Reinbek.
- Grohmann, R.** (1997): Das Problem der Evaluation in der Sozialpädagogik. Bezugspunkte zur Weiterentwicklung der evaluationstheoretischen Reflexion. Frankfurt.
- Guba, E.G./Lincoln, Y.S.** (1981): Effective Evaluation. San Francisco.
- Hager, W./Patry J.L./Brezing, H.** (2000): Handbuch Evaluation psychologischer Interventionsmaßnahmen. Standards und Kriterien. Berne.
- Heiner, M.** (1988) (ed.): Selbstevaluation in der Sozialen Arbeit. Freiburg.
- Heiner, M.** (1998) (ed.): Experimentierende Evaluation. Ansätze zur Entwicklung lernender Organisationen. Weinheim.
- Hellstern, M./Wollmann H.** (1984) (ed.): Handbuch der Evaluierungsforschung. Opladen.
- IHK/ESOMAR** (not specified) (Internationale Handelskammer/Europäische Gesellschaft für Meinungs- und Marketing-Forschung): Internationaler Kodex für die Internationaler Kodex für die Praxis der Markt- und Sozialforscher. Amsterdam (<http://www.planung-analyse.de>).
- Joint Committee on Standards for Educational Evaluation** (2000): Handbuch der Evaluationsstandards, 2. Aufl., Opladen. (Original: The Program Evaluation Standards, 1994. 2. Auflage. Thousand Oaks).
- Kromrey, H.** (1998): Empirische Sozialforschung. Opladen.
- Kromrey, H.** (2000): Die Bewertung von Humandienstleitungen. Fallstricke bei der Implementations- und Wirkungsforschung sowie methodische Alternativen; in: Müller-Kohlenberg / Münstermann (2000), pp. 19-55.
- Kromrey, H.** (2001): Evaluation – ein vielschichtiges Konzept; in: Sozialwissenschaften und Berufspraxis Nr. 2/2001, pp. 105-131.
- Lamnek; S.** (1995): Qualitative Sozialforschung, Bd. 1 u. 2.. 3. Auflage. Munich.
- Mayring, P.** (1999): Einführung in die qualitative Sozialforschung. 4. Auflage. Weinheim.
- Miles, M.B./Huberman, A.M.** (1994): Qualitative data analysis: An expanded source book. Thousand Oaks.
- Müller-Kohlenberg, H. /Münstermann, K.** (2000) (ed.): Qualität von Humandienstleistungen. Evaluation und Qualitätsmanagement in Sozialer Arbeit und Gesundheitswesen. Opladen.
- Müller-Kohlenberg, H./Beywl, W.** (2001): Der Arbeitskreis Soziale Dienstleistungen. Entwurf eines Arbeitsplans, in: Deutsche Gesellschaft für Evaluation e.V. (ed.): Evaluation - Reformmotor oder Reformbremse. Cologne 2001.
- OECD** (1998): Review of the DAC Principles for Evaluation of Development Assistance. Paris.
- OECD** (1998a): Public Management Service Best Practice Guidelines for Evaluation. Paris.
- Owen, J. M./Rogers, P. J.** (1999): Program evaluation: Forms and approaches, Thousand Oaks.
- Patton, M.Q.** (1998): Die Entdeckung des Prozessnutzens. Erwünschtes und unerwünschtes Lernen durch Evaluation; in: Heiner (1998), pp. 55-66.
- Patton, M.Q.** (1997): Utilization-Focused Evaluation. 3. Auflage. Thousand Oaks.

- Rürup, B./Hansmeyer, K.-H.** (1984): Staatswirtschaftliche Planungsinstrumente. 3.Auflage. Dusseldorf.
- Scriven, M.** (1991): Evaluation Thesaurus. 4. Auflage. Newbury Park.
- SEVAL** (2001) Evaluation Standards of the Swiss Evaluation Society (SEVAL-Standards. <http://www.seval.ch>.
- Spiegel, H. v.** (1993) (ed.): Aus Erfahrung lernen. Qualifizierung durch Selbstevaluation. Münster.
- Stockbauer, U.** (2000): Was macht Evaluationen nützlich? Überblick zum Forschungsstand – Ergebnisse von Fallstudien; in: Müller-Kohlenberg/Münstermann (2000), pp. 121-128.
- Stockmann, R.** (2000) (ed.): Evaluationsforschung – Grundlagen und ausgewählte Forschungsfelder. Opladen.
- Stufflebeam, D.L.** (2001): Evaluation Models, New Directions for Evaluation, No. 89, Spring. San Francisco.
- Uhl, A.** (1999): Evaluation. in: Stimmer, F. (ed.): Suchtlexikon. Munich.
- Uhl, A.** (2000): The Limits of Evaluation; in: Neaman, R./Nilson, M./Solberg, U.(Hg.): Evaluation - A Key Tool for Improving Drug Prevention. EMCDDA Scientific Monograph Series, No 5. Lisbon.
- Wedel, H.v.** (1998): Erfolgskontrolle finanzwirksamer Maßnahmen in der öffentlichen Verwaltung (Schriftenreihe der Bundesbeauftragten für die Wirtschaftlichkeit in der Verwaltung, Bd. 2). Stuttgart.
- Widmer, T.** (1996): Meta-Evaluation. Kriterien zur Bewertung von Evaluationen. Berne.
- Widmer, T./Landert, C./Bachmann, N.** (2000): Evaluations-Standards der Schweizerischen Evaluationsgesellschaft (SEVAL-Standards). Berne/Geneva.
- Wottawa, H./Thierau, H.** (1998): Lehrbuch Evaluation, 2. Auflage. Berne.
- Wulf, C.** (1972) (ed.): Evaluation. Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen. Munich.