Thomas Widmer

EVALUATION STANDARDS IN THE CONTEXT OF SUSTAINABILITY

My contribution consists of three parts:

1. First, I would like to introduce evaluation standards as a tool for quality assurance and quality control in the field of evaluation.

This will be the longest part. So don't worry about it taking more than half of my time.

2. The second part of my presentation deals with the specific challenges of evaluation in the context of sustainability.

3. Finally, in part three, I will conclude my talk with some propositions that follow from parts one and two.

1. Evaluation Standards

In order to understand the thinking that underlies evaluation standards, it's a good idea to keep in mind what we do when we make a judgment or an assessment. Expressed simply, we relate a given object to a specific criterion. Obviously, defining the appropriate criterion is a crucial task in evaluation. If our goal is to judge the quality of an evaluation, we must choose the appropriate criterion to make this judgment. Because evaluation quality is a multidimensional concept, we need not only one, but several evaluation criteria that represent the expectations that evaluations must fulfill.

1.1. Reference Levels for Evaluation Quality

One can now differentiate the demands made on evaluation according to two different assessment criteria, namely internal and external, and use them to assess the quality of an evaluation. Table 1 provides an overview of the respective reference levels with their corresponding assessment bases:

Table 1: Reference Levels (Overview)

Level	Object	Basis for Internal Assessment Criteria	Basis for External Assessment Criteria
First Order Constructs	Program, Project, Measure, etc.	Objectives of the Program, Project, etc.	Social Significance of the Program
Second Order Constructs	Evaluation	Objectives of the Evaluation	Evaluation Theory and Methodology
Third Order Constructs	Metaevaluation	Objectives of the Meta-evaluation	….

An evaluator can thus assess a program (or another evaluandum; a first order construct) based on the program objectives (internal assessment criteria). But an evaluator can also assess a program based on the social significance of the program (external assessment criteria). The same reasoning can be applied to an evaluation as a second order construct, where the relevant assessment criteria would therefore first be the objectives of the evaluation (internal assessment criteria) and second general evaluation theory and methodology (external assessment criteria).

By analogy, this argument can be pursued at the third (or subsequent) order, and if one undertakes an assessment (or evaluation) of an evaluation, then one speaks of a *metaevaluation*. But in answer to the question who evaluates a metaevaluation, I agree with Michael Scriven who has written that: "no infinite regress is generated because investigation shows it usually doesn't pay after the first metalevel on most projects and after the second on any" (1991:230).

At this point, since I have used the term metaevaluation, let me say something about wording in order to avoid confusion.

The term *metaevaluation*, noted above, needs elucidation as well as demarcation from other approaches such as *evaluation syntheses* or *meta-analysis*.[1] The latter two instruments are derived from the substantive contents or results of evaluations; the basis for the investigation is the thematic evidence the evaluations themselves can provide. A *meta-analysis* requires a sufficient number of existing evaluation studies to permit a quantitative answer to be given in answer to a specific question. This quantitative focus typically makes a *meta-analysis* more narrowly oriented than an *evaluation synthesis*. An *evaluation synthesis*, though it requires a thematic clustering among the evaluation studies upon which it is based, is, in its qualitative approach, far less narrowly circumscribed.

---

[1] On this, and in following, see the classic texts of Scriven (1969), Stufflebeam (1974) and Cook/Gruder (1978), and, more recently, Stufflebeam (2000).

Table 2: Overview of Third Order Evaluation Instruments

Instrument	Description
Evaluation Synthesis	Content Synthesis of Various Evaluation Studies (largely qualitative) (global evaluation/cross-sectional analysis)
Meta-Analysis	Quantitative Integration of the Results of Various Evaluation Studies (research synthesis)
Metaevaluation	Evaluation of Evaluation(s): Systematic Assessment of the Quality of one or more Evaluation Studies

A *metaevaluation*, by contrast, has a fundamentally different goal, as it is an evaluation of one or more evaluations that intends to systematically establish the worth and merit of evaluation(s). In other words, a *metaevaluation* assesses the quality of an evaluation and for that it needs assessment criteria. As previously noted, both evaluation objectives (internal assessment criteria), and evaluation theory and methodology (external assessment criteria) need to be taken into account for this, and evaluation Standards allow for a more precise formulation of the external assessment criteria. As is true for evaluations themselves, metaevaluations can be fashioned in quite different ways: as self-evaluations or heteronomous evaluations, executed internally or externally, or they can fulfill formative or summative functions.

Let me now return to the issue of assessment criteria and reference level: The relevant point here is that if we want to investigate the quality of an evaluation, we essentially have two possible assessment criteria, the internal and the external. Because it is fundamentally impossible to generalize about objectives that are set for and in specific evaluations, we instead focus on the external assessment criteria for evaluations, that is, on evaluation theory and methodology.

1.2. Evaluation Standards

1.2.1. The Origin of the Standards

As part of the professionalization of evaluation, various parallel initiatives were launched in the U.S. during the late 1970s to define a set of criteria applicable to the many dimensions of evaluation quality.[2] The intent was to establish criteria capable of assessing the quality of an evaluation with respect to a general theory and methodology of evaluation – in effect to craft standards that would be external assessment criteria. Depending upon author, such parallel definitional efforts were sometimes based on thematic areas of evaluation, and sometimes on varying cognitive or epistemological basic premises. Various

---

[2] On the process of professionalization, see the two five-step models of Harold Wilensky (1964), and Dreyfus/Dreyfus (1986) as well as the discussion and literature noted in Altschuld (1999a: 483-6), who refers directly to evaluation.

lists of suggested criteria were published at the time, many marked by more or less limited epistemological orientations.

The publication in 1981 of the "Standards for Evaluation of Educational Programs, Projects and Materials" by the "Joint Committee on Standards for Educational Evaluation" (henceforth "JC Standards"), was, by contrast, characterized by a comparatively open methodological orientation. This was fortunate, since the U.S. evaluation community was embroiled in a long and heated controversy during the 1980s over the question whether qualitative methods - and not just classical, quasi-experimental orientations or understandings of evaluation - were acceptable. Due to its methodological and epistemological openness, the JC Standards anticipated the resolution of this controversy, namely that the use (or existence) of alternative methodologies is (generally) acceptable.

The JC Standards are directed primarily at those who deal with education evaluation, and most of the organizations represented in the "Joint Committee" come from education. While the Standards have been widely adopted in American evaluation practice, it has increasingly been in realms and topic areas far removed from their originally intended application to educational evaluation. This development is reflected not only in the new name for the Standards themselves (reissued in a revised version in 1994 as "The Program Evaluation Standards" with the original connection to education only visible in the subtitle "How to Assess Evaluations of Educational Programs"; see Joint Committee 1994) but also in the fact that the organizations represented in the Joint Committee go well beyond those only concerned specifically with education (see the overview in Widmer/Beywl 2000: 250). Thus the name change only completes what, in my view, has already long changed in practice, namely that the Standards are widely employed outside of educational contexts.[3]

The JC Standards were not given much attention in German-speaking Europe until the mid-1990s (Widmer 1996a-h, Widmer/Rothmayr/Serdült 1996; for a relatively early exception, see Beywl 1988: 113-23), despite the gradual increase in the use of evaluations during the 1980s. To ease access to the Standards for German speakers, a translation was finally published in 1999 (Joint Committee 1999), and the fact that it soon had to be reprinted in a second edition testifies to how quickly these Standards were taken note of and used (Joint Committee 2000; see also Beywl/Taut 2000).

In its efforts to professionalize evaluation in Switzerland, the Swiss Evaluation Society SEVAL gave the task of developing evaluation standards appropriate to Switzerland to one of its Working Groups.[4] This Working Group, after carefully considering a wide variety of existing models and standards (see Beywl/Widmer 2000), decided to develop country-appropriate evaluation standards, using the basic structure of the JC Standards but revising and adapting specific individual Standards.

A first version of this effort was greeted with great interest in the spring of 1999 by the SEVAL General Meeting, and after the Working Group reworked and submitted a revised version, the SEVAL General Meeting in the spring of 2001 approved what are now known as the SEVAL Standards (Widmer/Landert/Bachmann 2000). In so doing, SEVAL acted as a pioneer, becoming the first European evaluation society to establish its own evaluation standards. As Table 3 indicates, Switzerland is not an international exception; rather, it is

---

[3] The former chairman of the Joint Committee has recently assessed a whole series of general evaluation models (Stufflebeam 2001).

[4] This Working Group included members from quite different realms, including private sector evaluation firms, state administrative offices, and universities.

one example of the many efforts being made by national and regional evaluation societies to improve quality.

Table 3: The Development of Standards and Guiding Principles (selection) [5]

Organization	Guidelines and Standards	Source
African Evaluation Society	African Evaluation Guidelines (adapted from the Program Evaluation Standards of the Joint Committee)	www.geocities.com/afreval
American Evaluation Association (AEA)	Guiding Principles for Evaluators and Program Evaluation Standards (Joint Committee)	www.eval.org (see AEA 1995)
Australasian Evaluation Society (AES)	Guidelines on Ethical Conduct of Evaluation and Program Evaluation Standards (Joint Committee)	www.aes.asn.au
Canadian Evaluation Society (CES)	CES Guideline for Ethical Conduct	www.evaluationcanada.ca
German Evaluation Society (DeGeval)	Evaluation Standards (DeGeval-Standards) (adapted from the Program Evaluation Standards (Joint Committee) and the SEVAL-Standards)	www.degeval.de (see Beywl 2000 and Beywl/Taut 2000:367)
French Evaluation Society	Evaluation standards in preparation	www.sfe.asso.fr (see Perret/Barbier 2000)
Italian Evaluation Society	Guiding Principles	www.valutazioneitaliana.it

1.2.2. The SEVAL Standards: Structure and Content

The SEVAL Standards are based on the premise that an evaluation should at once be useful, feasible, proper, and accurate so as to fulfill the demands placed on it: good evaluations should therefore demonstrate all these characteristics. To make these category characteristics more tangible, the SEVAL Standards are subdivided into 27 individual Standards that fall into one of the four larger categories (Widmer/Landert/Bachmann 2000):

> **N Utility**
> *The utility standards are to guarantee that an evaluation is oriented to the information needs of the intended evaluation users.*

---

[5] For further comparisons between quality guidelines and standards that are employed internationally, see Widmer/Beywl 2000.

N1   Stakeholder Identification
Those participating in, and those persons affected by, an evaluation are identified in order that their interests and needs can be taken into account.

N2   Clarifying Evaluation Objectives
All persons who take part in an evaluation are concerned that the objectives of the evaluation are made clear to all stakeholders.

N3   Credibility
Those who conduct evaluations are both competent and trustworthy; this helps ensure that the evaluation results are accorded a maximum amount of acceptance and credibility.

N4   Scope and Selection of Information
The scope and selection of the information collected permits pertinent questions to be asked of the evaluandum while simultaneously taking the interests and needs of the party contracting the evaluation, as well as other stakeholders, into account.

N5   Transparency in Assessment
The underlying reasoning and point of view used to arrive at the interpretation of the results are described in such a manner that the bases for the value judgments are clear.

N6   Comprehensiveness and Clarity of the Report
Evaluation reports describe the evaluandum, including its context, objectives, the questions asked, the procedure(s) used, and the results of the evaluation. This ensures that the pertinent information is made available and can be readily understood.

N7   Timeliness in Reporting
Significant interim results and (the) final report(s) are made available to the intended users in such a fashion that they can be used in a timely manner.

N8 Evaluation Impact
The planning, execution, and presentation of an evaluation encourages stakeholders both to follow the process of evaluation and utilize the evaluation.

D   Feasibility
The feasibility standards ensure that an evaluation is conducted in a realistic, considered, diplomatic and cost effective manner.

D1   Practical Procedures
Evaluation procedures are designed in such a manner that the necessary information is collected without disrupting the evaluandum or the evaluation.

D2   Political Viability and Support
Evaluations take the various positions of different interests into consideration in their planning and execution in order to win their cooperation and discourage attempts by one or another group to limit evaluation activities or distort or misuse the results.

D3   Cost Efficiency
Evaluations produce information of a value justifying the financial resources expended.

K   Propriety
The propriety standards ensure that an evaluation is conducted in a legal and ethnical manner and that the welfare of stakeholders is given appropriate attention.

K1   Formal Agreements

The obligations of the parties to an evaluation (specifying what, how, by whom and when what is to be done) are set forth in written agreements that obligate the parties to fulfill all the conditions of the agreements, or if not, to renegotiate these agreements.

K2   Protecting Individual Rights

Evaluations are planned and executed in such a manner as to protect and respect human rights and welfare.

K3   Human Interaction

Evaluations are laid out in such a manner that contacts between participants are characterized by mutual respect.

K4   Complete and Fair Assessment

In their assessment and depiction of strengths and weaknesses in the evaluandum, evaluations are both complete and fair, such that strengths can be built on and problem areas addressed.

K5   Disclosure of Findings

The contracting parties to an evaluation ensure that the evaluation results are made available to the affected persons as well as to all those who have a legitimate claim to receive these results.

K6   Declaring Conflicts of Interest

Conflicts of interest are addressed openly and honestly in a manner that they will least compromise evaluation processes and conclusions.

G   Accuracy

The accuracy standards ensure that an evaluation produces and disseminates valid and usable information.

G1   Documentation of the Evaluandum

The evaluandum is clearly and precisely described and documented so that it can be unambiguously identified.

G2   Analysis of Context

The influences of the context on the evaluandum are identified.

G3   Description of Objectives and Procedures

The objectives, questions, and procedures used in and by the evaluation are described and documented with sufficient precision that they are identifiable and can be judged.

G4   Trustworthy Information Sources

The sources of information used in an evaluation are sufficiently precisely described that the appropriateness of the information can be assessed.

G5   Valid and Reliable Information

The procedures for collecting information are so chosen or developed, and then employed, that the validity and reliability of the interpretation arrived at is ensured for the given purpose.

G6   Systematic Information Checks

The information that is collected, analyzed, and presented in an evaluation is systematically checked for errors.

G7   Qualitative and Quantitative Analysis

Qualitative and quantitative information in an evaluation are systematically and appropriately analyzed so that the questions posed in the evaluation are actually answered.

G8   Substantiated Conclusions

The conclusions reached by an evaluation are explicitly substantiated in a manner that stakeholders can follow and that permits them to assess the conclusions.

G9   Objective Reporting

Reporting is free from distortion through personal feelings or preferences for one or another party to the evaluation; evaluation reports fairly reproduce the results.

G10 Metaevaluation

The evaluation itself will be evaluated on the basis of existing (or other important) standards so that the execution can be done accordingly and so that stakeholders can in the end assess the strengths and weaknesses of an evaluation.

Beyond this list, the SEVAL Standards also include a general introduction (with information about the objectives, scope, and addressees of the Standards), further explication of each individual Standard, and various supporting materials (functional overview, creation process, etc).[6]

As can readily be seen, the number of individual Standards in each category varies; feasibility, for example, has only three while accuracy has ten individual Standards. This should not be misunderstood as implying a *weighting* among the four categories. Rather, there is a deliberate wish to avoid weighting either categories or individual Standards, because the significance of any specific Standard can only be determined in each individual case. A generalized weighting that made claims to be valid across cases thus would be inappropriate.

The question of *weighting* is particularly important because the SEVAL Standards contain demands that are, at least in part, at odds with one another. In practice, in fact, evaluation is often confronted with the question which Standard should enjoy prominence of place. But this lack of internal consistency in the SEVAL Standards should not by any means be seen as weakness, since it reflects the (previously described) tension-ridden and conflictual arena in which evaluations move.

Put differently, the SEVAL Standards are maximum demands - not minimal standards of what is an absolute must but rather a statement of what a good quality evaluation should ideally try to achieve. In practice, it will also rarely ever be possible to completely fulfill every one of the 27 individual Standards. Nevertheless, all participants, and not just the evaluator, should try to take the SEVAL Standards into account as much as possible. With the SEVAL Standards, evaluation has an instrument available to it that describes, in a precise and sufficiently differentiated manner, what the demands of quality are.

---

[6] The complete SEVAL Standards can be found, in German and French, on the SEVAL homepage (http://www.seval.ch/); Italian and English versions are in preparation.

2. Evaluation with sustainability

Some weeks ago, when I told a few colleagues of mine that I would be participating in a conference on the Evaluation **of** Sustainability, they asked me: "Sustainability of what?" I couldn't give them an answer then, but perhaps this conference will help me to do so.

Sustainability is not an object of evaluation. Sustainability is a quality, a capacity – an evaluation criterion. So it makes no sense to evaluate sustainability itself. But it does make sense to evaluate something, an object, a program or a public policy, using sustainability as an evaluation criterion.

Sustainability is a non-trivial concept, which – in contradiction with what the title of this conference suggests – is not so easy to deal with empirically than most other evaluation criteria. The concept of sustainability is multidimensional. Indeed, we face at least three dimensions of assessment: the economic, social and ecological dimensions. Some scholars add a fourth dimension, namely the quality of the political process (Thierstein 1998, Widmer/Schenkel/Hirschi 2000). It is thus unavoidable to use a multicriteria approach for assessment.

But this is not the only problem we are confronted with when trying to evaluate the quality of sustainability. The core problem lies elsewhere. The concept of sustainability has other characteristics that make it difficult to use as a measure to assess a policy, a program, a project or any other object. In particular, the following three aspects make the evaluator's life difficult:

1. the long-term perspective of the concept of sustainability,

2. uncertainty as a crucial component of the concept, as well as

3. the fact that the concept of sustainability is holistic, thus refusing to restrict itself to one specific aspect.

The problem with the *long-term perspective* of the concept lies in the difficulties connected with investigating future situations empirically. How do you measure the future development or the future state of affairs? It is certainly possible to develop expectations about the way in which current behavior will impact on the future. But it is impossible to avoid making assumptions when trying to forecast the future, assumptions that are not empirically testable and that may be wrong.

The second problem - *uncertainty* - bothers the evaluator even more than the first one does. People familiar with risk assessment or with a basic understanding of random theory will follow me on this one.   The main question is the following: how do you assess the significance for the future of knowledge about risks elaborated in the present? Estimating the incidence of a specific risk is a hard task in itself. Assessing its relevance for the present is difficult. Assessing its relevance for the future is almost impossible.

Third, the *holistic* nature of sustainability, referred to in German as "Ganzheitlichkeit", forces the evaluator to give an overall, comprehensive judgment. A sound empirical investigation, however, needs definitions to focus on. Investigating the whole world is not possible in one evaluation.

Many discussions on the subject of sustainability systematically ignore these three aspects. This is the case especially when people discuss indicators (or the measurement) of sustainability. Even in the field of evaluation this is normally the case. In my view, this

reduces the complexity of the concept of sustainability too much. I am tempted to call this practice a corruption of the underlying assumptions of sustainability.

I conclude this discussion about the evaluation *with* (and not *of*!) sustainability with the statement that to take sustainability seriously means to avoid using it as an overarching evaluation criterion. Would that mean that we have to stop using sustainability as an evaluation criterion in general?

3. What conclusions can we draw from all of this?

If you have a positive image of both evaluation and sustainability – and as attendees of this meeting, I suppose you do! – , how do you manage the dilemmas outlined? Let me make some propositions on how to approach the problem:

- The fist proposition would be that we should be more honest and acknowledge the restrictions we are confronted with when using the concept of sustainability in evaluation. This would include the more sensitive use of the term sustainability, even though the concept it is now quite trendy.

- As I already mentioned, we need an object to evaluate. Sustainability is not an object. Therefore, let's not try to evaluate sustainability. Let's evaluate specific objects with sustainability as an evaluation criterion.

In this context, I would suggest having a look at the SEVAL Standards. They might help solve some of the problems.

- Evaluation Standards teach us something about the multicriteria problem I mentioned. The Standards formulate 27 criteria to judge evaluation quality. The Standards reflect the different demands that the evaluation is confronted with. They sometimes conflict; they sometimes are not strictly separated from one other; some are highly interdependent. The situation is quite similar for the multifaceted concept of sustainability. But in the SEVAL standards, and this is my point, no weighting of the criteria exists. In the case of sustainability, this would mean that we should avoid calculating aggregated values, consisting of values of indicators of the different dimensions of sustainability. We should restrict ourselves to a more modest level of judgment, separated for the different dimensions of sustainability.

- A second proposition could thus be to use the SEVAL Standards as a tool to assess the quality of evaluations in the context of sustainability. With the Standards, we are able to distinguish the better from the worse and we are in a well-founded position to judge the quality of the results produced by the different studies. This would be a contribution to sustainability in itself.

Evaluation Standards cannot make miracles happen and – I must be modest on this point too – they can't solve all the existing problems. But in my view, this is also a benefit. If evaluation standards gave us all the answers to all of our problems, then there would be no need to continue the EASY-ECO workshops.

Thomas Widmer, Department of Political Science, University of Zurich, Switzerland

References

ALTSCHULD, James W. (1999): The Certification of Evaluators. In: *American Journal of Evaluation* 20(3): 481-93.

AMERICAN EVALUATION ASSOCIATION, TASK FORCE ON GUIDING PRINCIPLES FOR EVALUATORS (1995): Guiding Principles for Evaluators. *New Directions for Program Evaluation* 66: 19-26.

BEYWL, Wolfgang (1988): Zur Weiterentwicklung der Evaluationsmethodologie. Frankfurt: Lang.

BEYWL, Wolfgang (2000): Standards for Evaluation: On the Way to Guiding Principles in German Evaluation. *Evaluation Center Occasional Papers Series*, Vol. 17. Kalamazoo, Western Michigan University: 60-5.

BEYWL, Wolfgang/TAUT, Sandy (2000): Standards: Aktuelle Strategie zur Qualitätsentwicklung in der Evaluation. *Vierteljahreshefte zur Wirtschaftsforschung* 3/2000: 358-70.

BEYWL, Wolfgang/WIDMER, Thomas (2000): Die 'Standards' im Vergleich mit weiteren Regelwerken zur Qualität fachlicher Leistungserstellung. In: Joint Committee on Standards for Educational Evaluation (Hrsg.): Handbuch der Evaluationsstandards. Opladen: Leske+Budrich: 259-95.

COOK, Thomas D./GRUDER, C.L. (1978): Metaevaluation Research. *Evaluation Quarterly*, 2: 5-51.

DREYFUS, Hubert L./DREYFUS, Stuart E. (1986): Mind over Machine. New York: Free Press.

JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION (1981): Standards for Evaluations of Educational Programs, Projects, and Materials. New York: McGraw-Hill.

JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION (1994): The Program Evaluation Standards. Newbury Park: Sage.

JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION (Hrsg.) (1999): Handbuch der Evaluationsstandards. Opladen: Leske + Budrich.

JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION (Hrsg.) (2000): Handbuch der Evaluationsstandards. 2. Auflage. Opladen: Leske + Budrich.

PERRET, Bernard/BARBIER, Jean-Claude (2000): Ethical Guidelines, Process and Product Quality Standards, What For?. Paper presented at the EES Conference, Lausanne, October, 12th – 14th 2000.

SCRIVEN, Michael (1969): An Introduction to Meta-Evaluation. *Educational Product Report* 2(5): 36-8.

SCRIVEN, Michael (1991): Evaluation Thesaurus. Fourth Edition. Newbury Park: Sage.

STUFFLEBEAM, Daniel (1974): Meta-Evaluation. Evaluation Center Occasional Paper Series, Vol. 3. Kalamazoo, Western Michigan University.

STUFFLEBEAM, Daniel L. (2000): The Methodology of Metaevaluation. In: Stufflebeam, Daniel L./Madaus, George F./Kellaghan, Thomas (eds.): Evaluation models. Boston: Kluwer: 457-71.

STUFFLEBEAM, Daniel L. (2001): Evaluation Models. *New Directions for Evaluation*, 89.

THIERSTEIN, Alain (1998): Raumordnung und nachhaltige Entwicklung: Handlungsansätze für eine nachhaltige Raumentwicklung in der Schweiz.. Hrsg. vom Bundesamt für Raumplanung. Bern: EDMZ.

WIDMER, Thomas (1996a): Meta-Evaluation: Kriterien zur Bewertung von Evaluationen. Bern: Haupt.

WIDMER, Thomas (1996b-h): Fallstudien zur Meta-Evaluation. 7 vols. Reihe 'Schlussberichte' des NFP 27. Bern: Schweizerischer Nationalfonds.

WIDMER, Thomas (2000): Evaluating Evaluations: Does the Swiss Practice Live up to The Program Evaluation Standards? *Evaluation Center Occasional Papers Series*, Vol. 17. Kalamazoo, Western Michigan University: 67-80.

WIDMER, Thomas/BEYWL, Wolfgang (2000): Die Übertragbarkeit der Evaluationsstandards auf unterschiedliche Anwendungsfelder. In: Joint Committee on Standards for Educational Evaluation (Hrsg.): Handbuch der Evaluationsstandards. Opladen: Leske+Budrich: 243-57.

WIDMER, Thomas/LANDERT, Charles/BACHMANN, Nicole (2000): Evaluations-Standards der Schweizerischen Evaluationsgesellschaft (SEVAL-Standards). Bern / Genève: SEVAL.

WIDMER, Thomas/ROTHMAYR, Christine/SERDÜLT, Uwe (1996): Kurz und gut? Qualität und Effizienz von Kurzevaluationen. Zürich: Rüegger.

WIDMER, Thomas/SCHENKEL, Walter/HIRSCHI, Christian (2000): Akzeptanz einer nachhaltigen Verkehrspolitik im politischen Prozess. Deutschland, Niederlande und Schweiz im Vergleich. Nationales Forschungsprogramm 'Verkehr und Umwelt' (NFP 41), Bericht D13. Bern: BBL/EDMZ.

WILENSKY, Harold L. (1964): The Professionalization of Everyone. *American Journal of Sociology* 70(2): 137-58.