



Astrid Kuffner

EVALUATION OF SUSTAINABILITY- SUSTAINABLE EVALUATION? LESSONS LEARNED FROM A SURVEY OF EVALUATION REPORTS

Task and Objectives

Project description

This analysis marked the first step of a comprehensive research project on the evaluation of sustainability in the Research Focus "Managing Sustainability" at the Vienna University of Economics and Business Administration. It was initiated by the German Evaluation Society (DeGEval) with its working group on environmental evaluation. Its main objective was to perform a survey allowing conclusions on the current state of evaluation in this sector in the German- speaking region.

Methodology

Research

The author went through a list of selected evaluators of the DeGEval¹, who were most helpful by providing their own materials as well as new leads through the "snowball system". The author offered the contacted evaluators a broad definition of evaluation and sustainability, which naturally resulted in a variety of studies provided.

After a period of collection from winter 1999 to spring 2000 the sample of reports amounted to 43. The collection provided the first data base for the evaluation market (customers as well as suppliers) in this special field of application for Germany, Switzerland and Austria. The approach might have resulted in a potential sampling bias, but was the most apt given the green-field nature of sustainability-evaluation.

Analysis

To analyse the reports the author developed a system for classification which consisted of two elements. First she defined common traits of the reports by reading and re-reading them several times (qualitative element). Then the author developed a grid of appraisal factors based on descriptive statistics, sustainability criteria, a checklist for sustainability projects² and selected Program Evaluation Standards³ (quantitative element). Out of scope for the analysis were reports which did not even treat one column of the triad of sustainability, Structural Funds Evaluations⁴ and examinations of methods without an example or use in practice.

It should be noted, that the two elements of the analysis were not meant to exclude some of the reports in the first place, but to iteratively define types of evaluation reports. Both the fields of evaluation and sustainability still lack precise definitions- as do the reports. Hence, the author took an explorative approach to the problem as outlined above.

Given a strict time budget and methodological constraints, the analysis focused on measuring levels of analysis rather than the intensity of the analysis- for the given areas of in-depth appraisal, and the sustainability criteria.

¹ Special thanks to Wolfgang Meyer of the DeGEval (working group for environmental evaluation).

² The checklist for sustainability projects of the so called "NachhaltigkeitsTATENBank", served as a basis (a project of the Austrian Federal Ministry of Environment in cooperation with the Interdisciplinary Institute for Environmental Economics).

³ The Program Evaluation Standards of the Joint Committee on Standards for Educational Evaluation, compare SANDERS, James R. (Editor): Handbuch der Evaluationsstandards- Die Standards des Joint Committee on Standards for Educational Evaluation, Leske + Budrich, Opladen 1999

⁴ because they represent a type of evaluation which is already regulated by the European Union in interval, methodology and form and would have disturbed the explorative character of the analysis.



As an example for these “levels of evaluation” might serve the degree of in- depth analysis, which was categorised the following way:

- description: description of measures taken for documentation purposes
- efficiency evaluation: follows the subjective valuation of participants in a project- comparison of target and actual state
- outcome and/or effectiveness evaluation: tries to establish cause-effect chains, measures outcomes quantitatively or qualitatively, against a given reference value
- success factors: tries to open the “black box” of impacts even further by defining prerequisites which helped or inhibited the success of a project

Results

Descriptive statistics

43 reports were collected and analysed in this survey. The results show trends in this specialised field, and allow statements on the market situation for environmental and sustainability evaluations.⁵

A general increase of environmental and sustainability evaluation in the German speaking region from 1994 until 1999 can be deducted.

The major part of the evaluations are concomitant (65%), on the regional level (54%) and performed externally (86%). Politics are preferably evaluated in Switzerland, otherwise programmes and projects dominate as evaluanda (both 42%). 30% of evaluations are formative, 30% summative and 40% are both. The type of intervention dominantly evaluated are subsidies (33%).

Almost 50% of the evaluation data were collected and analysed with quantitative methodology (46,5%) (also compare table 1).

table 1: Descriptive Statistics (n=43), marked with * multiple choice possible

n= 43	percentage	evaluanda	percentage
timing			
ex-ante	7%	policies & regulations	16%
concomitant	65%	program	42%
ex-post	28%	project	42%
supplier		level of analysis*	
internal	14%	description	12%
external	86%	valuation (efficiency)	93%
		cause effect (effectiveness)	40%
		success factors	30%
country		methodology used	
Germany	49%	qualitative	21%
Switzerland	14%	quantitative	46%
Austria	37%	both	33%
regional level		additional ecologic*	7, %
national	25%	additional technical*	2%
regional	54%	use	percentage
local	21%	formative	30%
		summative	30%
		both	40%

⁵ For full details of the results also view: KUFFNER, Astrid: Evaluation von Nachhaltigkeitsaspekten- nachhaltige Evaluation?- Eine vergleichende Analyse von deutschsprachigen Evaluationsberichten, die sich mit nachhaltiger Entwicklung beschäftigen, Universität Wien, Wien 2000



Sustainability check

In general a long-term perspective is not a criterion for performing a good evaluation – an evaluation is normally done to give support choosing between alternatives (before or during a project) or looking at it afterwards. The projection of future effects is a well-known methodological problem and in addition normally not in scope of the ordering party. Only one report out of 43 treated possible effects after the intended end of the program.

We can safely say that about one third of the reports treated all aspects of the sustainability triad, although only one gave them approximately equal weight.

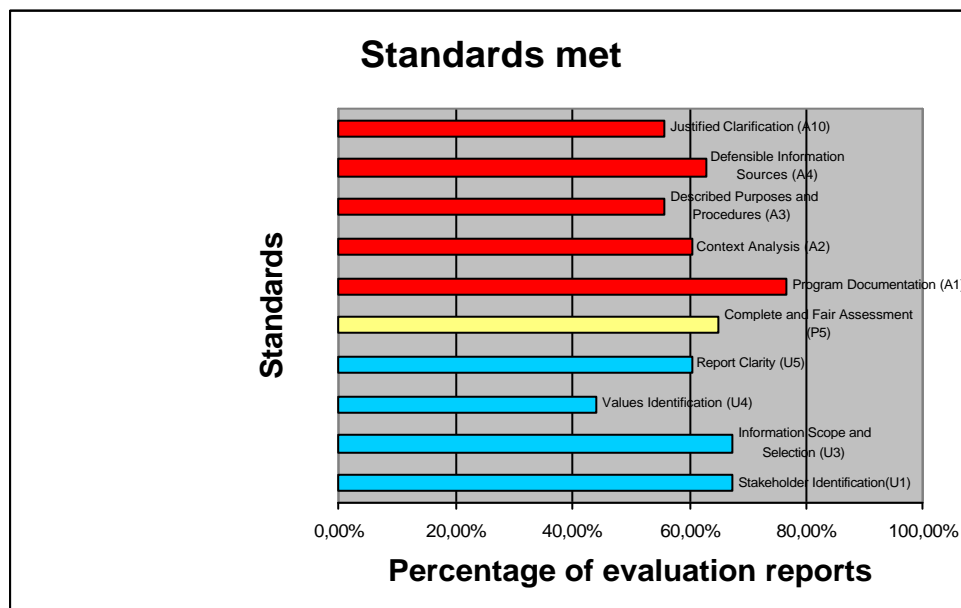
If one defines participation as the questioning of participants in a program/project this criterion is fulfilled by 65% of the reports due to the questionnaires used. If one considers developing the design and/or discussing the results with those, who are stakeholders of the evaluandum, the rate converges towards 0%.

Quality Standards

The Program Evaluation Standards have originally been developed to evaluate evaluations (metaevaluation). This was not the purpose of the study at hand. Nonetheless the Standards have been the only accredited way to rate the quality of evaluations - at that time⁶. The Standards used for this analysis were selected as to apply to evaluation reports only. That is why the category “Feasibility Standards” has not been applied at all: one can not draw conclusions by reading the evaluation report only.

The author showed how the reports would have performed if accredited Standards were used (for the performance and the standards used also view graph 1).

graph 1: The percentage of reports which met the selected quality standards.



⁶ Today, the DeGEval (10/2001) and the SEVAL (Swiss Evaluation Society, 12/2000) have developed standards for evaluation.



Lessons learned

Remarks

The following collection of experiences reflects the view of the author. It contains good ones as well as bad ones. The list is certainly far from complete, but may be of help to the Young Researchers who will be working in a similar field after EASY- ECO I.

Trash and Flash Experiences

- The collection of reports is surely a good way to get in touch with evaluators in this field. You can define hot spots in the market- be it in the field of application (e.g. environmental management and energy sector), the region, or the person of the evaluator. Collecting itself has an multiplier effect: the more persons you contact the more persons contact you. Spread the rumor and make use of it.
- It is common sense that there are different uses of evaluations e.g. affirmation, legitimacy support or severance for programs, establishing a new consulting branch, etc. The question is not, whether evaluations are used, but what for. When assessing an evaluation it is very important to keep the quality of the content, the line of arguments and the analytic insight in mind, not personal judgement. The accredited Program Evaluation Standards were the perfect means for this purpose.
- It will be very easy to define a “who is who” of the evaluation market. The evaluators with reputation, who evaluated big projects with a lot of subsidies behind them will be easy to find. The challenge lies in finding evaluators, who do good work, but are not well-known - yet. Keep your eyes open for institutions who promote small projects, local governments, and University Institutes. The motivation for choosing one evaluator over another are variegated and there are many evaluations in the field of subsidies.
- Although the public sector is probably the richest source of evaluations, it is not the only one. When looking for evaluations, one sometimes has to behave like a truffle swine, going where it apparently leads astray. Follow up every organisation that gives subsidies e.g. government, NGO, foundations, societies, associations, unions, banks, institutes, or clubs. The tighter the link to firms/economy and the more experimental the approach (test runs) the higher the potential. But still, highly interesting evaluations may be done internally and often be confidential.
- The importance of an evaluation is not automatically linked to its quality, but to the importance of the evaluandum. The perspectives of the researcher and the consumer could sometimes not differ more. Keeping in mind that every evaluation is used for a certain purpose and trying to understand the perspective of the ordering party might help.
- Although the objects of evaluation differ, the methods do not. This makes it easier to classify evaluations and compare them. Depending on the field of application, additional methods from the core discipline of the evaluandum are used. After several evaluations with multiple choice questionnaires you may appreciate the contrasts provided by an interdisciplinary approach.
- Precise definitions are not always given- not for evaluation and not for the evaluandum. A mix of methods may be of help in this case. Reading evaluation reports several times helps you to develop “Gestaltwahrnehmung”- a sense for important common traits and properties- commonness itself. Once you own this special sense you will find it easier to label the evaluations and put them in order.