

# ORG<sup>nic</sup> ActionPlan

## **ORGAP WP 3**

### **Task 3.2**

#### **Revised version**

Christian Eichert, Stephan Dabbert  
University of Hohenheim

<b>1. Introduction .....</b>	<b>1</b>
<b>3. Methodology.....</b>	<b>1</b>
2.1. <i>Selection of common evaluation standards and set of criteria for the meta-evaluation .....</i>	<i>3</i>
2.2. <i>Description of the empirical process .....</i>	<i>3</i>
a.) <i>Criteria used for the assessment of the evaluation studies .....</i>	<i>3</i>
b.) <i>Analysis of the evaluations` context and recommendations .....</i>	<i>4</i>
c.) <i>Comparison of the studies corresponding to the criteria .....</i>	<i>4</i>
<b>3. First results of the meta-evaluation.....</b>	<b>5</b>
3.1 <i>Description of the evaluation studies and their underlying programmes.....</i>	<i>5</i>
3.1.1 <i>Title and underlying programmes of the Dutch study.....</i>	<i>5</i>
3.1.2 <i>Title and underlying programmes of the Danish study .....</i>	<i>6</i>
3.1.3 <i>Title and underlying programmes of the German study .....</i>	<i>6</i>
3.1.4 <i>Purpose, key evaluative question and main methods used in the Dutch study .....</i>	<i>6</i>
3.1.5 <i>Purpose, key evaluative question and main methods used in the Danish study .....</i>	<i>7</i>
3.1.6 <i>Purpose, key evaluative question and main methods used in the German study .....</i>	<i>7</i>
3.2 <i>Brief narrative description of the Dutch evaluation.....</i>	<i>8</i>
3.3 <i>Brief narrative description of the Danish evaluation .....</i>	<i>8</i>
3.4 <i>Brief narrative description of the German evaluation.....</i>	<i>9</i>
3.5 <i>Evaluation according to the Criteria and Assessment of the Operability of the used Set of Criteria.....</i>	<i>10</i>
3.5.1 <i>Evaluation according to the criteria (and the resulting interpretation according to Widmer).....</i>	<i>10</i>
3.6 <i>Strengths and weaknesses of the Dutch study.....</i>	<i>14</i>
3.7 <i>Strengths and weaknesses of the Danish study.....</i>	<i>14</i>
3.8 <i>Strengths and weaknesses of the German study .....</i>	<i>14</i>
3.9 <i>Comparative Synthesis and Recommendations on the Design of Future Evaluations.....</i>	<i>14</i>
3.9.1 <i>Comparison of the studies:.....</i>	<i>14</i>
<b>4. Specification of the question of ORGAP Task 3.3 .....</b>	<b>15</b>
<b>5. Final note on ORGAP WP 3 .....</b>	<b>16</b>
<b>6. Bibliography.....</b>	<b>18</b>
<b>7. Annexes.....</b>	<b>19</b>

## 1. Introduction

**Task 3.2** is divided into two sub-tasks. One focuses on methodology, and the other on content. The first sub-task is the meta-evaluation study; it aims to conduct an in depth analysis of existing evaluation studies, carried out at national level, in order to gain information about the evaluation methodologies used in these studies. Collateral to this, there will be further examination of the feasibility and appropriateness of evaluation studies in the field of organic action plans. This meta-evaluation study and the further examination of the evaluations' contexts will help "to determine the potential contribution of the results and available data to the application of the ORGAPET toolbox at the European level".

One can see this as a self-contained milestone (cf. a.) which will be further analyzed and interpreted in the next sub-task. The work will be completed by October 2006.

In a second step (cf. b.) (coupled to Task 3.3, finished in month 20; end of December 2006), the project team will draw preliminary conclusions concerning issues affecting the success or failure of these plans as input into WP5.

The selected countries for the meta-evaluation are the following: DE, DK, NL and UK. Accordingly, one can refer to corresponding evaluation studies undertaken in these countries. It is further planned to call in expert assessment to judge the impact of the conducted evaluations on the appropriate action plan (e.g. adaptation of the action plan concept as a result of the evaluations' results).

In summary, one can conclude that Task 3.2 contributes to a methodological learning process (in the field of evaluation), as well as helps to optimize the ORGAPET toolbox. In addition, it will provide information on a content level about the success and failure of Organic Action plans.

### *Further context*

In **Task 3.3** (coupled to Task 3.2b, finished in month 20; end of December 2006) the results of the foregoing tasks will be taken up for analyzing the relevant context of the evaluation studies and for giving first recommendations.

**Task 3.4** is concerned with the identification of areas of conflict and/or synergy between the objectives of national and EU action plans and their significance for the implementation of the EU Action Plan at national level.

## 2. Methodology

This chapter specifies the methodology used for task 3.2. First, the meta-evaluation analyses the chosen studies according to a defined set of criteria, and then draws conclusions about each evaluation methodology by comparing the studies to each other in order to gather information on each particular methodology.

As a result of the meta-evaluation, one can differentiate according to

- the scientific quality,
- the relevance for the practice,
- the utilisation and
- the effects of a study or a set of studies (c.f. Widmer 1996, p. 4).

In order to guarantee scientifically justifiable results, it is necessary to get an overview of each complete evaluation study. Thus, the studies had to be available in either English or German.

As described in the Technical Annex there will be 4 studies which were considered:

As regards UK and Denmark, the topic of the meta-evaluation work is an action plan<sup>1</sup>. In the German study, the evaluation work done on the “Federal organic farming Scheme” is included, even though the FOFS is not a “pure” action plan. The FOFS can be seen as a preliminary stage on the way to a German Organic Action plan<sup>2</sup>.

For Denmark, the plan is to include the evaluation work done on the “First law to support organic farming”<sup>3</sup> (from the year 1987, with evaluation work finished 1992). As reason for this decision, one can refer to the fact that the two Danish action plans were not evaluated as self-contained units; *“The Danish statusnotater” are political documents describing the state of art with regards to implementation of the recommendations (anbefalinger) from the action plan seen from the view of the minister and are as such not traditional evaluations (Mette Meldgaard)*. The “First law to support organic farming” can be seen as a forefather of the two Danish action plans, including all relevant points and references. The Danish study was included in the meta-evaluation process to develop a deeper understanding on evaluation methodology in the context of organic farming policy (in more general terms).

At the moment (September 2006) it is not yet clear whether the British study can be used for the meta-evaluation work. The British evaluation work covers an organic action plan, but the methodology used can be seen only as rudimentarily<sup>4</sup> useful for the considered meta-evaluation methodology. Consequently, the study will build (together with the Danish study) a framework for deeper understanding of the political field.

---

<sup>1</sup> “.....strategic actions developed and undertaken in close partnership between the public and the private sector, including consumers, farmers, producers, retailers, non governmental organisations (NGO's), researchers and other important stakeholders.” Danish Ministry for Food, Agriculture and Fisheries (2001).

<sup>2</sup> Isermeyer et al. (2001)

<sup>3</sup> Out of the Danish questionnaire, filled-in by Mette Meldgaard (in 2005 member of the Danish research team), ORGAP task 3.1: “Public support of organic farming dates back to 1987 where the first law to support organic farming was implemented including definition of organic farming, area support and support for development projects. The evaluation of this law was made by a private consulting firm and it included specific recommendations for future development areas, which serve as model for the two Action Plans. The same consulting firm was appointed by the Minister to carry out practical tasks in preparing the two action plans and involve the Organic Food Council.”

<sup>4</sup> “Action plan to develop organic food and farming in England – Two years on” enfoldes recommendations on the future of the action plan. Further on there is a short list of recommended criteria for the (future) evaluation of the action plan.

## ***2.1. Selection of common evaluation standards and set of criteria for the meta-evaluation***

The studies were built on different evaluation methodologies adapted to the topic of policy and programme evaluation. In order to comply with the requirements for a certain comparability of the findings, it is necessary to choose a neutral grid for testing each single study.

One of the most common evaluation standards in Europe includes the “Standards for Evaluation” of the “Gesellschaft für Evaluation” (DeGEval; German evaluation society). This evaluation set is fundamentally based on the evaluation standards of the US-“Joint Committee on Standards for Educational Evaluation”. The standard set was partly adapted to the central European cultural area.

Because of its well-balanced design allowing for a neutral applicability, it seems useful to conduct the planned meta-evaluation for the ORGAP project by applying the above mentioned set of standards (German Evaluation Society 2001).

## ***2.2. Description of the empirical process***

### ***a.) Criteria used for the assessment of the evaluation studies***

Each of the studies was analysed with respect to the chosen set of standards (DEGEVAL standards). These standards were divided into four main categories:

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

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

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

The **Accuracy** Standards are intended to ensure that the evaluation produces and discloses valid and useful information and findings pertaining to the evaluation questions (cp. German Evaluation Society 2001).

The outcome will present an insight into the design and methodological configuration of a study (“How is the evaluation study built-on?”), as well as into the classification or the indexing of the standards (“Do the evaluation study meet the criteria?”) (cf. Widmer 1996, p. 5). Furthermore, it will verify whether the chosen criteria for the meta-evaluation set were appropriate.

One point apparent in Widmer (1996) is that of the (consequent) impacts of evaluation studies. Widmer argues that it is advisable to include “impact driven criteria” for the assessment of each evaluation work. His argumentation shows that the underlying set of criteria built up in North America (by the “Joint Committee on Standards for Educational Evaluation”) is of little value in terms of integrative aspects. The (auxiliary) criteria can be seen as follows:

Deliverables and reports shall be available to all interested parties to guarantee use of the results; evaluations shall be planned, conducted and described in a way to encourage the stakeholders in following the evaluation work and thus raise the likelihood of subsequent use of the findings.

For the ORGAP meta-evaluation, special attention will be paid to this aspect of stakeholder integration in conducting the further work.

### ***b.) Analysis of the evaluations` context and recommendations***

As a next step, the insights of the meta-evaluation of each study were set into context. Therefore, a specific explanation of the (political) environment of the action plan and its development process, as well as a description of the evaluation study's environment was considered. Included herein will be considerations on a classification of several **cultural peculiarities** (the political and cultural context of the evaluation) in the meaning of evaluation approaches (possible variations in approach can be evoked due to the prevailing political system; e.g. federalism vs. centralism). These insights can help in defining and categorizing different instruments of the ORGAPET toolbox.

Also, relevance for the evaluations` context is seen in the problem of **data availability**. The analysis of the evaluations` context will comment on that point.

A third specific focus area is the analysis of how **stakeholders** were **involved** in the design, as well as in the evaluation process. This aspect will be addressed in Task 3.3.

In the end, a description of the conclusions and the argumentation line was presented, documented and explained. In some cases it may be desirable to include the view of the evaluation team of the primary study (by means of a statement on the conclusions).

The analysis of the evaluations` context is mainly done in combination with Task 3.3 (consequentially to the initial run during the meta-evaluation work; UWA and UHO in collaboration).

### ***c.) Comparison of the studies corresponding to the criteria***

Following the above, the subsequent step reflects on a comparative synthesis (matching) of the three (four) studies with the aim to

- get an insight into the **adequacy** of the chosen **set of criteria** for assessing the certain topic,
- obtain **conclusions** about appropriate **modifications**, and finally
- give **recommendations** about the advancement of the **evaluation practice** in the range of organic farming support policies (cf. Widmer 1996, p. 5 et seq.).

To make this point clear: The main aim of this step is to get an insight into the specific approaches and methods used in the three (four) studies. The derived knowledge will be used in the application of the ORGAPET toolbox.

Figure 1 can be seen as a survey or detailed work plan on the meta-evaluation work steps in Task 3.2. In particular the points 6 – 11 show how the single steps are interlocked:

Figure 1: Structure for Identifying Alternative Metaevaluation Procedures

Tasks	
1.	Determine and arrange to interact with the meta-evaluation's stakeholders
2.	Staff the meta-evaluation team with one or more qualified evaluators
3.	Define the meta-evaluation questions
4.	Agree on standards, principles, and/or criteria to judge the evaluation system or particular evaluation
5.	Develop the memorandum of agreement or contract to govern the meta-evaluation
6.	Collect and review pertinent available information
7.	Collect new information as needed, including, for example, on-site interviews, observations and surveys
8.	Analyse the qualitative and quantitative information
9.	Judge the evaluation's adherence to appropriate standards, principles, and/or criteria.

10. Convey the meta-evaluation findings through reports, correspondence, oral presentations, etc.
11. As needed and feasible, help the client and other stakeholders interpret and apply the findings.

Source: Stufflebeam (2001b)

### **3. First results of the meta-evaluation**

It is clear that the researched evaluation studies are not perfect. In doing this meta-evaluation, the question is whether the evaluations achieve their purpose at an acceptable qualitative level.

In general, the in-hand meta-evaluation pursues the approach of programme theory. Accordingly, the methodology of the researched evaluation study (as a construction of second order) builds upon the programme theory for research on the meta-level. (Widmer, page 9).

As the meta-evaluation within the ORGAP project is planned as a desk study, the presented results could not be used as an overall picture of the researched evaluation work. Some of the valuations were not possible due to the limited data and document insights. Regrettably, several interesting points concerning, for instance, many questions on reliability, financing, treating and interim reports could not be evaluated.

In other areas, the in-hand meta-evaluation can help illuminate methods and details for improving evaluation methodology in the field of organic farming support schemes. One important improvement is the analysis of stakeholder integration in the planning, conducting and assessment of evaluations.

This meta-evaluation will verify whether the already done evaluations could help improve their political field and/or society. According to Fitzpatrick et al. (1996), it is important that an evaluation find its audience in order to exploit its worth. Consequently, an evaluation shall assure

- relevance for the decision makers and/or other stakeholders,
- involvement of the users in the planning and reporting stages,
- reputation/credibility of the evaluator,
- qualitatively adequate communication of the findings and results (in terms of time, frequency and method) and
- procedures that assist in the use or recommendation of (future) action.

In the first subchapter, the studies from Denmark, The Netherlands and Germany and their underlying programmes will be briefly presented. Subsequently, there is an assessment of the studies according to the used set of criteria. The chapter will be concluded with an overall assessment comparing the three studies to each other and the inferred recommendations.

#### ***3.1 Description of the evaluation studies and their underlying programmes***

##### ***3.1.1 Title and underlying programmes of the Dutch study***

The Dutch study is concerned with a political document which highlights the possible future design of the Dutch organic farming sector. The study is named “Biologisch meer gangbaar” Evaluatie-onderzoek Nota Biologische Landbouw 2001-2004 (“Organic more mainstream” Evaluation research Policy document Organic Agriculture 2001-2004).

### ***3.1.2 Title and underlying programmes of the Danish study***

The Danish study evaluates a legal act which rules the support of development projects and the conversion support for the Danish organic sector. The study is named “Evaluation of the Organic Farming Act no. 363 of June 10, 1987”.

### ***3.1.3 Title and underlying programmes of the German study***

The German study is named “Final report of the evaluation of the federal organic farming scheme” and accordingly evaluates the federal organic farming scheme (“Bundesprogramm Oekologischer Landbau”), a program supporting the development of the German organic sector in a broad approach and can partly be seen as a “action plan” (e.g. the formulation of a certain objective concerning organically managed area).

### ***3.1.4 Purpose, key evaluative question and main methods used in the Dutch study***

#### **Purposes of the Dutch evaluation study:**

- Verify whether the societal aims of the policy document ‘Een biologische markt te winnen’ are realized
- Provide justification to the parliament: Justification for costs (spent on the described programmes supporting organic farming); “How well has the money been used?”
- Learn from experience (for future policies and support schemes)

#### **Key evaluative questions of the Dutch evaluation study:**

Final evaluation of the policy document Organic Agriculture 2001-2004; i.e. what went well, what should be continued, what should change, concerning:

- professionalizing the demand oriented supply-supply-chains,
- optimizing the transparency and closing of the supply-supply-chains,
- knowledge development and dissemination,
- stimulating of organic primary production,
- other measures (e.g. fiscal measures)

Research questions were formulated by using a **causality diagram**. Per policy spearhead, two causality schemes have been defined: one in accordance with the policy document; one in accordance with the actual policy implementation. As a result, four main evaluation questions have been formulated:

1. **The degree of achieving the targets:**
  - a.) Growth of the area: Will the target of 10% organic farming area in 2010 be realized? Will the intended growth of the organic area be realized?
  - b.) Consumer expenditure: Will the Dutch consumer, at the end of 2010, belong to the top users of organic products in Europe?
  - c.) Competitive strength: Will the sector compete internationally by the end of 2010?
2. **Effectiveness of the policy:** Has the policy contributed enough to the growth of the organic sector and the growth of consumer expenditure in organic products? (Followed by sub-questions)
3. **Effectiveness of the implementation:** Could the implementation of the policy (product and services) of the national government have been done with lesser means, without damaging the quality?



4. **Assessment of “Platform Biologica” (market support stage):** To what degree has “Biologica” contributed to the objectives of the policy document (ex post)? What will “Biologica” do if the yearly fixed contribution of LNV is removed?

**Main methods of the Dutch evaluation study:**

- Existing research results and data on organic farming in the Netherlands;
- Collection of extra information (large amount of interviews and written information, provided by project leaders/coordinators);
- Quantitative information: Collection of information on project level (visible outputs, results, effects of that output, causality);
- Qualitative information: interviews to research difficulties, critical factors, general picture, ideas, relative effectiveness, appreciation, etc.

### ***3.1.5 Purpose, key evaluative question and main methods used in the Danish study***

**Purpose(s) of the Danish evaluation study:**

- Assess whether the Organic Farming Act has promoted the development of organic farming
- Serve as a strategic basis for decision making in relation to future initiatives within the organic field

**Key evaluative questions of the Danish evaluation study:**

- What effects do the development projects initiated under the Act have?
- What importance does the support for conversion have for the development of organic farming?
- Assessment of the administration of the development programmes and the interplay between the Directorate for Farming, the Council for Organic Farming and the programme stakeholders.
- Assessment of the labelling system for organic products.
- Assessment of the functioning of the whole organic food chain (including monitoring and marketing).

**Main methods of the Danish evaluation study:**

- Phone interviews,
- face-to-face interviews and
- document analysis.

### ***3.1.6 Purpose, key evaluative question and main methods used in the German study***

**Purpose(s) of the German evaluation study:**

- The main purpose of the evaluation of the Federal Organic Farming Scheme (FOFS) is to analyse and review the conduction and impacts of selected programme measures and to review the programme and process management as a whole. Out of that analyse the evaluation gives recommendations for amelioration of the execution and impacts of the future FOFS.

**Key evaluative questions of the German evaluation study:**

- The evaluation design was build-on in order to give tool for a direct input during the implementation phase of the programme measures. Accordingly, one main question was to assess the achievement of objectives of the measures. Furthermore the evaluation assesses the effectiveness and efficiency of the FOFS measures in order to detect potentially unexploited capabilities.

**Main methods of the German evaluation study:**

- Combination of qualitative and quantitative methods
- Half standardised interviews (via phone, postal, on-line, face-to-face)

### ***3.2 Brief narrative description of the Dutch evaluation***

After a midterm evaluation in 2002, the study constitutes the final evaluation of the policy document Organic Agriculture 2001-2004. The study was finalised in July 2004. The evaluation study started with a clearly structured plan. The main topic (documenting policy on organic farming) is structured in comprehensible subchapters, which are assessed separately.

The Dutch study very strictly follows the prescribed plan and the formal requirements. It provides supporting documents and sources in an accurate manner. The statements made are neutral, well founded and justified, the evaluation team appears clearly distanced from the evaluation topic and accordingly can be seen as neutral and fair in their judgements. External factors (e.g. the development of the organic market in other EU countries) and unwanted side effects of the researched support schemes were observed and assessed. In the annex, the evaluation team presents a detailed overview on the evaluation questions.

Of notable value in the study is the detailed analysis of stakeholder integration in the political process; the Dutch study provides a very detailed assessment of how stakeholders were integrated via the existing or newly established different bodies or boards (e.g. the new established “Task force” building a network and institutionalizing the dialogue between policy, market actors and other interested stakeholders).

The Dutch evaluation report is available as an internet document to all interested parties and subsequently, guarantees the use of its results. The report is written in an open and impartial way which allows and encourages all interested stakeholders to draw their own conclusions. Concerning the aspect of stakeholder integration, one can judge the report as valuable.

One can say that the study follows the intended evaluation cycle in a systematic manner, starting with a status quo analysis, followed by the fixing of targets, fixing of measures and criteria for the assessment, a mid term evaluation (not included in this document, but existing), a final census and recommendations how to design future policies.

### ***3.3 Brief narrative description of the Danish evaluation***

The evaluation study starts with a sound status-quo analysis of the Danish organic farming sector. To this end, the evaluation team has collected an abundance of information via **phone interviews**,

- a.) with farmers receiving conversion support
- b.) project leaders (participation in the development programme) and
- c.) conventional farmers.

The overall coverage of these interviews can be considered to be very representative (for farmers receiving conversion support 17% coverage of the total number, for project leaders 91% coverage of the total number). The **face-to-face interviews** were conducted with

- a.) members of the Council for Organic Farming,
- b.) representatives from the Directorate for Farming,
- c.) representatives from the Directorate for Plants and the Veterinary Directorate,
- d.) representatives from organisations and associations with interest in organic farming, and

e.) researchers  
and, in addition, a **document analysis** (e.g. summaries from meetings of the Council for Organic Farming) was done.

The received basic data on the Organic Farming Act was further analysed relating to three different aspects:

1. Management of the Organic Farming Act (assessment of the efficiency of the involved actors and institutions)
2. Quantitative effects of the Organic Farming Act (assessment via turnover, export, investments, and employment)
3. Qualitative effects of the Organic Farming Act (assessment of the influence on the professionalization, organisational development, information and counselling, research, labelling and monitoring, reduction of barriers)

As a result, the overall strategic perspective of the Organic Farming Act on production and market perspective is described. The assessment of each subchapter is made in a neutral and logical manner. The final recommendations are listed in a separate (overview) chapter for ease in understanding.

As a final result, the in-hand study was conducted systematically along a prescribed path and well describes the status quo of the topic, the intended targets, the applied measures and criteria, as well as gives a number of recommendations for the future design of the researched policy field.

### ***3.4 Brief narrative description of the German evaluation***

The method and goal of the German evaluation study was to get an insight into the functioning of the process management of a new established programme. Accordingly the study can be seen as an on-going evaluation. As a consequence a final résumé on the quality and achievement of objectives of every single measure is not the intention and not possible with the chosen evaluation design.

The evaluation is divided into two sub-studies: The topic “consumer information” was evaluated separately by another company. All results are included and presented in one final study.

The used methods are mainly standardised surveys of indicators in the direct sphere of influence of the researched measures, partly supplemented by some further qualitative data. Accordingly the resulting statements move directly into the assessment of the short-term programme effects.

Relating to the circumstances of the evaluation (e.g. time pressure, facts of a new established programme with accordingly new involved programme officers) the study was done in a courageous manner. The evaluation team surveyed a plenty of (relevant) information and data on the topic and interpreted them very well. The evaluation design is build-up precise and adequate according to the underlying topic. One can say that the most of the achieved statements attend to be valid and comprehensible.

As a final result one can attest the evaluation team a good job and the consistent tracing of their well-build evaluation design.

### 3.5 Evaluation according to the Criteria and Assessment of the Operability of the used Set of Criteria

#### 3.5.1 Evaluation according to the criteria (and the resulting interpretation according to Widmer)

The set of criteria used is adapted from DEGEVAL (2003), Stufflebeam (1999) and Stufflebeam (2001a). A detailed description of the applied set of criteria can be found in the annex. The final assessment will be made when the meta-evaluation of the German study is finalised. For the moment their will be a short data presentation in form of tables and a critique of each criteria group relating to both studies in hand.

#### Assessment codes for the study:

☐ / 1. row ☐ = No

☒ / 2. row ☒ = Yes

≈ / 3. row = No answer / No data available to evaluate

### Utility

	Denmark			The Netherlands			Germany			overall		
	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈
U1 Stakeholder Identification (max. 10 Points)		5	5	6	4		1	8	1	7	17	6
U2 Clarification of the Purposes of the Evaluation (max. 3 Points)		3			3		1	2		1	8	0
U3 Evaluator Credibility and Competence (max. 10 Points)	1	6	3	1	5	4		8	2	2	19	9
U4 Information Scope and Selection (max. 10 Points)	1	7	2		9	1		10		1	26	3
U5 Transparency of Values (max. 13 Points)	6	7		1	12		3	10		10	29	0
U6 Report Comprehensiveness and Clarity (max. 14 Points)	3	11		1	13		1	13		5	37	0
U7 Evaluation Timeliness (max. 10 Points)		1	9	1	2	7		6	4	1	9	20
U8 Evaluation Utilisation and Use (max. 13 Points)	1	7	5	2	9	2	1	10	2	4	26	9
<b>Total</b>	<b>12</b>	<b>47</b>	<b>24</b>	<b>12</b>	<b>57</b>	<b>14</b>	<b>7</b>	<b>67</b>	<b>9</b>	<b>31</b>	<b>171</b>	<b>47</b>
<b>%</b>	<b>14,5</b>	<b>56,6</b>	<b>28,9</b>	<b>14,5</b>	<b>68,7</b>	<b>16,9</b>	<b>8,4</b>	<b>80,7</b>	<b>10,8</b>	<b>12,4</b>	<b>68,7</b>	<b>18,9</b>

<b>Appraisable (in %)</b>	<b>71,1</b>	<b>83,1</b>	<b>89,2</b>	<b>81,1</b>
---------------------------	-------------	-------------	-------------	-------------

Altogether more than two third of the sub criteria of both studies can be classified according to the criterion “utility”. The Dutch (with 83% classification) and the German study (with 89% classification) reach a very good result and allow a well-founded interpretation.

The Danish study reaches a relatively (in relation to the 71% quantifiable single criterion) high rating according to the criterion group “utility”. Especially the sub criterion “comprehensiveness and clarity of the report” can be stated as adequate and facilitate the use of the study for interested stakeholders.

The Dutch and the German study have a relatively higher achievement of objectives according to the criterion “utility”. Especially the part “transparency of values” can be rated as very good. That means that the evaluation team reveals their used raster and criteria for assessing the different parts of the researched programme.

## Feasibility

	Denmark			The Netherlands			Germany			overall		
	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈
F1 Appropriate Procedures (max. 11 Points)	1	5	5		7	4	1	6	4	2	18	13
F2 Diplomatic Conduct (max. 3 Points)		3		1	2			3		1	8	0
F3 Evaluation Efficiency (max. 13 Points)	1	7	5	1	9	3	1	10	2	3	26	10
<b>Total</b>	<b>2</b>	<b>15</b>	<b>10</b>	<b>2</b>	<b>18</b>	<b>7</b>	<b>2</b>	<b>19</b>	<b>6</b>	<b>6</b>	<b>52</b>	<b>23</b>
<b>%</b>	<b>7,4</b>	<b>55,6</b>	<b>37,0</b>	<b>7,4</b>	<b>66,7</b>	<b>25,9</b>	<b>7,4</b>	<b>70,4</b>	<b>22,2</b>	<b>7,4</b>	<b>64,2</b>	<b>28,4</b>
<b>Appraisable (in %)</b>	<b>63,0</b>			<b>74,1</b>			<b>77,8</b>			<b>71,6</b>		

The criterion “feasibility” achieves all in all between 55 and 70% of the prescribed objectives. According to a classification feasible for only 63 to 77% this can be seen as a very good grade.

The Danish study can be quoted as highly “diplomatic conducted”. The Dutch study reaches its highest assessable appraisal according to the single criterion “efficiency”. The German study is also quoted as being highly “effective”.

## Propriety

	Denmark			The Netherlands			Germany			overall		
	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈
P1 Formal Agreement (max. 11 Points)			11	2	3	6	1	10		3	13	17
P2 Protection of Individual Rights (max. 12 Points)	1	9	2		10	2		11	1	1	30	5
P3 Complete and Fair Investigation (max. 10 Points)	2	6	2	4	5	1		10		6	21	3
P4 Unbiased Conduct and Reporting (max. 2 Points)		2		1	1		1	1		2	4	0
P5 Disclosure of Findings (max. 11 Points)	2	8	1	4	7		2	7	2	8	22	3
<b>Total</b>	<b>5</b>	<b>25</b>	<b>16</b>	<b>11</b>	<b>26</b>	<b>9</b>	<b>4</b>	<b>39</b>	<b>3</b>	<b>20</b>	<b>90</b>	<b>28</b>
<b>%</b>	<b>10,9</b>	<b>54,3</b>	<b>34,8</b>	<b>23,9</b>	<b>56,5</b>	<b>19,6</b>	<b>8,7</b>	<b>84,8</b>	<b>6,5</b>	<b>14,7</b>	<b>66,2</b>	<b>20,6</b>
<b>Appraisable (in %)</b>	<b>65,2</b>			<b>80,4</b>			<b>93,5</b>			<b>80,9</b>		

In overall the Danish study is assessable for 65%, whereas the Dutch and the German study are again ahead with its 80% respectively 93% of assessable sub criteria.

The reporting and conduction of all three studies appears unbiased and fair to all involved stakeholders. All in all the German study reaches the relatively highest quoting according to the criterion “propriety”.

## Accuracy

	Denmark			The Netherlands			Germany			overall		
	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈
A1 Description of the Evaluand (max. 11 Points)	3	6	2	4	6	1	1	8	2	8	20	5
A2 Context Analysis (max. 11 Points)	3	8		5	6		2	9		10	23	0
A3 Described Purposes and Procedures (max. 12 Points)	9	2	1	7	4	1	1	8	3	17	14	5
A4 Disclosure of Information Sources (max. 11 Points)	6	5		3	6	2	1	10		10	21	2
A5 Valid and Reliable Information (max. 22 Points)	12	9	1	12	8	2	6	12	4	30	29	7
A6 Systematic Data Review (max. 1 Points)			1			1			1	0	0	3

A7 Analysis of Qualitative and Quantitative Information (max. 20 Points)	12	7	1	15	5		10	9	1	37	21	2
A8 Justified Conclusions (max. 11 Points)	8	3		4	7		1	10		13	20	0
A9 Meta-Evaluation (max. 11 Points)	10		1	10		1	10		1	30	0	3
<b>Total</b>	<b>63</b>	<b>40</b>	<b>7</b>	<b>60</b>	<b>42</b>	<b>8</b>	<b>32</b>	<b>66</b>	<b>12</b>	<b>155</b>	<b>148</b>	<b>27</b>
<b>%</b>	<b>57,3</b>	<b>36,4</b>	<b>6,4</b>	<b>54,5</b>	<b>38,2</b>	<b>7,3</b>	<b>29,1</b>	<b>60,0</b>	<b>10,9</b>	<b>47,0</b>	<b>44,8</b>	<b>8,2</b>
<b>Appraisable (in %)</b>	<b>93,6</b>			<b>92,7</b>			<b>89,1</b>			<b>91,8</b>		

Even if the criterion of “accuracy” reaches with 89% to 94% the highest worth concerning the possibility to classify the sub criteria, both studies failed the goal of being accurate partly. The sub criterion “description of purposes and procedures” shows that both studies miss to correctly list the relevant documentation in its final report. The same assessment can be made for the sub criterion of the “analysis of qualitative and quantitative information” and “meta-evaluation”.

The worth of (rateable) failed sub criteria exceeds for both studies the worth of the reached sub criteria.

## Overall assessment

	Denmark			The Netherlands			Germany			overall		
	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈
<b>Total</b>	<b>82</b>	<b>127</b>	<b>57</b>	<b>85</b>	<b>143</b>	<b>38</b>	<b>45</b>	<b>191</b>	<b>30</b>	<b>212</b>	<b>461</b>	<b>125</b>
<b>%</b>	<b>30,8</b>	<b>47,7</b>	<b>21,4</b>	<b>32,0</b>	<b>53,8</b>	<b>14,3</b>	<b>16,9</b>	<b>71,8</b>	<b>11,3</b>	<b>26,6</b>	<b>57,8</b>	<b>15,7</b>
<b>Appraisable (in %)</b>	<b>78,6</b>			<b>85,7</b>			<b>88,7</b>			<b>84,3</b>		

The overview shows that 78% to 89% of the sub criteria were ranked (either positive or negative). That means a relatively high informational value according to the criteria for both studies.

The Dutch and the German studies have a relatively higher achievement of objectives according to all criteria.

### ***3.6 Strengths and weaknesses of the Dutch study***

#### **(Apparent) strengths:**

The apparent strengths of the study are the feedback process with the steering committee scheduled in the evaluation design, as well as the substantiated data collection (where relevant and essential). Also, there is a clear differentiation between depiction of facts and the parts of the study allowing interpretation, such as those interpreted by one of the stakeholders. The evaluation team has chosen a competent and realistic approach in preparing the plan and time schedule of the evaluation study.

#### **Apparent weaknesses:**

An apparent weakness of the study is the fact that too few data and statistical sources are named or consulted in interpreting the different programme parts.

### ***3.7 Strengths and weaknesses of the Danish study***

#### **(Apparent) strengths:**

The apparent strengths of the study are a well-founded analysis of the Danish organic sector, a detailed analysis of the current market situation and an assessment of future opportunities for the Danish sector to expand (e.g. potential for Danish exports). A multitude of interviews with high coverage and accordingly good statistical value (eased by small size of the researched sector) have been provided.

The recommendations concerning the improvement of the administration and monitoring function is a worthwhile aspect not seen in comparable studies. When further analysed, these recommendations can be exemplarily useful for ORGAPET with regard to complex competency in the organic sector at the European level.

#### **(Apparent) weaknesses:**

The quantitative data are, at least in part, not well presented and analysed in the final document. In the explanation of the data, it would be desirable that the validity, reliability and consequently informational value of the study, be clearly demonstrated.

### ***3.8 Strengths and weaknesses of the German study***

One of the strengths of the German study is the fact, that the whole evaluation process was guided by a steering committee. The first results were discussed with that board and with the concerned stakeholders during a meeting (Reflexionsworkshop”) in order to ensure a feedback process and to avoid a misinterpreting of the drawn conclusions.

The used methods are relevant and all data well interpreted. The evaluation team traced systematically a consistent and well-build evaluation design.

### ***3.9 Comparative Synthesis and Recommendations on the Design of Future Evaluations***

#### ***3.9.1 Comparison of the studies:***

Bei einem Vergleich der drei Studien muss deutlich gemacht werden, dass es sich um relativ unterschiedliche Voraussetzungen handelt. Die D Studie stellt klar die Erhebung von Prozessdaten in den Vordergrund und kann folglich als on-going oder Begleitevaluation aufgefasst werden.



Regarding the evaluation methodology, the discussed studies have not generated many new insights. Their overall worth is as a broad compilation and analysis of the existing data on the researched topic. Starting with a well-founded status quo analysis, this was adequate/ample to answer the most important evaluation questions of the programmes. For future evaluations, one would recommend a systematic accumulation of the relevant data at the European level to simplify and standardise the approach. It is conceivable to approach this by using the first results of other EU-funded projects, for instance EISfOM. Out of these existing databases on organic farming, the ORGAP project could create a certain evaluation grid for improving and standardising the applied methodologies Europe-wide.

In general, it can be asserted that the evaluation methodology and professionalism has improved since the beginning of the 1990ies. This statement can be supported by comparing the Danish (year 1992) and the Dutch (year 2004) evaluation studies, among which a degree of improvement concerning the used methodology can be seen. For the reader and user of the Dutch evaluation study, it is much easier to gain an impression of the analysed topic, the study is organised more systematically, and it is easy to interpret of the statements made in relation to their environment.

In conclusion, one must address the worth of the studies for their political field; all studies contain a well founded analysis of the underlying political field (support of organic farming in national context) in form of status quo analyses. The studies achieved their purpose in presenting the relevant politicians or actors a sound basis for judgement and will likely lead to improvement of the researched programme or policy.

#### **Problems of data availability:**

The meta-evaluation makes clear that it is necessary to have a well-founded and standardized data base on the political field of organic farming. The researched studies have had to spend a considerable degree of their resources on data collection for gaining an insight into the topic. In the case of the Danish study, it is not fair to assess too harshly, because the study dates back to the year 1992 and does not reflect the current database. In either case, an EU-wide systematic evaluation approach needs an EU-wide systematic database. Accordingly, the recommendations of the ORGAP project will be linked to the existing data on organic farming, such as, for instance, results of the EISfOM project (see *Willer and Lux 2004*). Departing from this point, ORGAPET can reveal ways to use this database for evaluating the political field of organic farming.

## **4. Specification of the question of ORGAP Task 3.3**

As mentioned in *Analysis of the evaluations` context and recommendations*, there is, to some degree, a coupling of Task 3.2 and Task 3.3. That linkage has relevance for the subjects of

- stakeholder integration,
- data availability,
- explication of cultural peculiarities and
- context interpretation.

Having been stated in the abovementioned section, they will not be repeated here.

In the matter of stakeholder integration, one can add the aspect of the stakeholders' role in the evaluation process; the ORGAP concept sees them as part of a capacity building process instead of actors in an accountability process<sup>5</sup>. This integrative aspect explicitly incorporates the efforts in ORGAP to integrate the evaluation section (primary study level, cf. *include the sight of the evaluation team*).

Moreover, the integration of such a “softer” evaluation facets can include learning oriented forms of evaluation<sup>6</sup>: In consequence of the more “technical expertise” in Task 3.2 a “good practice example” can be selected and analysed in Task 3.3. Wollmann (2005) describes strengths and prospects of this methodology as follows: *“While (rigorous) evaluation aims at giving a comprehensive picture of ‘what has happened’ in the policy field and project under scrutiny, encompassing successful as well as unsuccessful courses of events, the best practice approach tends to pick up and ‘tell success stories’ of reform policies and projects, with the analytical intention of identifying the factors that explain the ‘success’, and with the ‘applied’ (learning and ‘pedagogic’) purpose to foster ‘lesson drawing’ from such experience in the intra-national as well as in the inter- and trans-national contexts. These obtained factors have to be related to the political and cultural context of the evaluation.*

Based on the described stakeholder oriented processes, Task 3.3 can constitute further recommendations on evaluation methodology in terms of action plans.

As a milestone in the stakeholder integration process, there will be group interviews in the 8 case study countries. Their use is to identify aspects of the operability and appropriateness of ORGAPET in execution at the national level (scope and relevance of indicators, compatibility with existing procedures, programmes, etc.).

## 5. Final note on ORGAP WP 3

In summary, one can describe Task 3.2 as a support or feed-in of recommendations for reflecting on a new policy cycle concerned with organic farming and action plan policies. This can be done at either the European or national level. For clarification of this point,

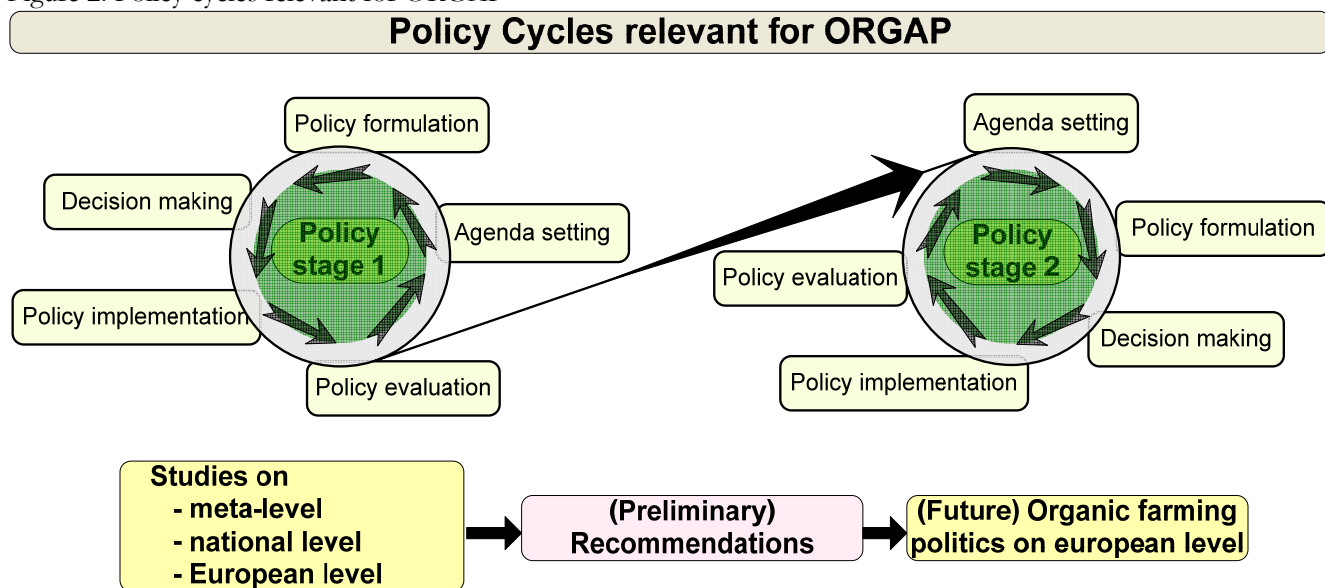
Figure 2 describes the interlocking of the mentioned policies. Wollmann (2005) describes this process as follows: *“Being carried out after the “termination” of a policy of a program, ex-post evaluation may not only have the task of informing the policy-makers (and the general political public) about the assets and liabilities of the past policy intervention, but may (and ideally should) be fed, within the concept of the “policy cycle”, into the preparation and formulation of next policy “round”.*“

---

<sup>5</sup> Referring to Herrmann and Hoefer (1999) one can cite their understanding of *integrative evaluation*: “evaluation is a social, not a metrological process”

<sup>6</sup> cf. Wollmann (2005)

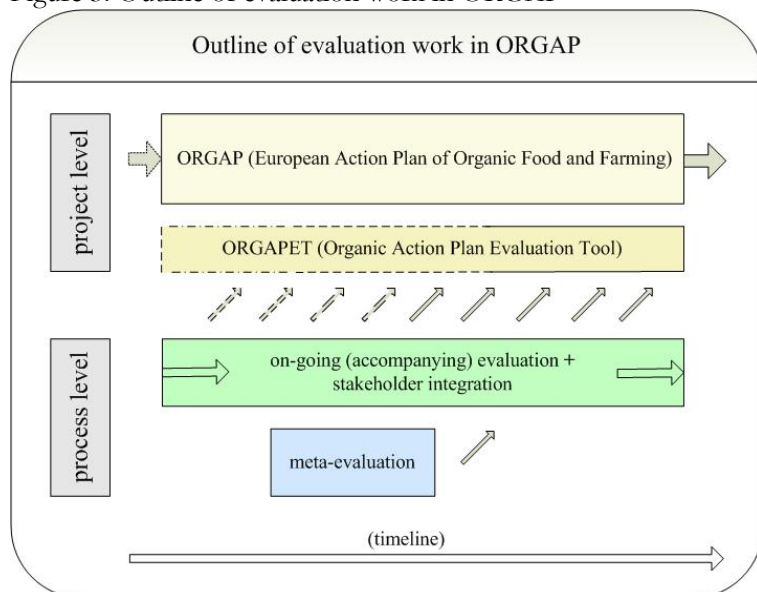
Figure 2: Policy cycles relevant for ORGAP



Source: own chart

In Figure 3 the whole outline of the ORGAP project is visualized. The figure shows the role of the “self-contained” meta-evaluation work and the location of the integration and recommendation aspects in the project plan.

Figure 3: Outline of evaluation work in ORGAP



Source: own chart, based on Eichert (2004)

## 6. Bibliography

- Danish Ministry for Food, Agriculture and Fisheries (2001); The Copenhagen declaration. Organic Food and Farming. Towards Partnership and Action. Proceedings of the Conference held 10./11.5.2001 in Copenhagen, Copenhagen
- DEGEVAL (2003); Selected Comments to the Standards for Evaluation of the German Evaluation Society – English Edition - edited by Beywl, Wolfgang.
- Eichert, Christian (2004); Das Projekt ORGAP. Lecture hold at the Ressourcenökonomisches Seminar, University of Hohenheim. Unpublished.
- Fitzpatrick, Jody L., James R. Sanders and Blaine R. Worthen (1996); Program evaluation: alternative approaches and practical guidelines. 2nd Ed. New York, NY: Longman.
- German Evaluation Society (DeGEval) (2001); DeGEval-Standards: Summary of evaluation standards. <http://www.degeval.de/calimero/tools/proxy.php?id=72>
- Herrmann, Joachim, Christoph Höfer (1999): Evaluation in der Schule – Unterrichtsevaluation. Berichte und Materialien aus der Praxis. Gütersloh.
- Isermeyer, Folkhard, Hiltrud Nieberg, Stephan Dabbert, Jürgen Heß, Thomas Dosch and Felix Prinz zu Löwenstein (2001); Bundesprogramm Ökologischer Landbau. Entwurf der vom BMVEL beauftragten Projektgruppe. Braunschweig
- Stufflebeam, Daniel L. (1999); Program Evaluations Metaevaluation Checklist - based on The Program Evaluation Standards
- Stufflebeam, Daniel L. (2001a); Guiding Principles Checklist for Evaluating Evaluations in consideration of The Guiding Principles for Evaluators
- Stufflebeam, Daniel L. (2001b); The metaevaluation imperative. American Journal of Evaluation, 22 (2), 183-210.
- Widmer, Thomas (1996); Meta-Evaluation: Kriterien zur Bewertung von Evaluationen. Bern: Haupt.
- Willer, Helga and Susanne Lux (2004); Europäische Konferenz Statistische Informationssysteme zum Öko-Landbau [European Conference on Statistical Information Systems for Organic Farming]. *Ökologie & Landbau* 2004(3):pp. 36-37
- Wollmann, Hellmut (2005); Evaluation and evaluation research. Working Paper 54/2005. Universidad Autónoma de Madrid, Facultad de Derecho.

## 7. Annexes

1. Raster and overview
2. Meta-Evaluation Form Denmark
3. Meta-Evaluation Form The Netherlands
4. Meta-Evaluation Form Germany

## Annex 1: Meta-Evaluation Form<sup>7</sup> - Raster and overview on the results

### Assessment codes for the study:

☐ / 1. row ☐ = No

☒ / 2. row ☒ = Yes

≈ / 3. row = No answer / No data available to evaluate

### Utility

	Denmark			The Netherlands			Germany			overall		
	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈
U1 Stakeholder Identification (max. 10 Points)		5	5	6	4		1	8	1	7	17	6
U2 Clarification of the Purposes of the Evaluation (max. 3 Points)		3			3		1	2		1	8	0
U3 Evaluator Credibility and Competence (max. 10 Points)	1	6	3	1	5	4		8	2	2	19	9
U4 Information Scope and Selection (max. 10 Points)	1	7	2		9	1		10		1	26	3
U5 Transparency of Values (max. 13 Points)	6	7		1	12		3	10		10	29	0
U6 Report Comprehensiveness and Clarity (max. 14 Points)	3	11		1	13		1	13		5	37	0
U7 Evaluation Timeliness (max. 10 Points)		1	9	1	2	7		6	4	1	9	20
U8 Evaluation Utilisation and Use (max. 13 Points)	1	7	5	2	9	2	1	10	2	4	26	9
<b>Total</b>	<b>12</b>	<b>47</b>	<b>24</b>	<b>12</b>	<b>57</b>	<b>14</b>	<b>7</b>	<b>67</b>	<b>9</b>	<b>31</b>	<b>171</b>	<b>47</b>
<b>%</b>	<b>14,5</b>	<b>56,6</b>	<b>28,9</b>	<b>14,5</b>	<b>68,7</b>	<b>16,9</b>	<b>8,4</b>	<b>80,7</b>	<b>10,8</b>	<b>12,4</b>	<b>68,7</b>	<b>18,9</b>
<b>Appraisable (in %)</b>	<b>71,1</b>			<b>83,1</b>			<b>89,2</b>			<b>81,1</b>		

<sup>7</sup> (adapted from DEGEVAL (2003); Selected Comments to the Standards for Evaluation of the German Evaluation Society – English Edition - edited by Beywl, Wolfgang, Stufflebeam, Daniel L. (1999); Program Evaluations Metaevaluation Checklist - based on The Program Evaluation Standards and Stufflebeam, Daniel L. (2001); Guiding Principles Checklist for Evaluating Evaluations in consideration of The Guiding Principles for Evaluators)

## Feasibility

	Denmark			The Netherlands			Germany			overall		
	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈
F1 Appropriate Procedures (max. 11 Points)	1	5	5		7	4	1	6	4	2	18	13
F2 Diplomatic Conduct (max. 3 Points)		3		1	2			3		1	8	0
F3 Evaluation Efficiency (max. 13 Points)	1	7	5	1	9	3	1	10	2	3	26	10
<b>Total</b>	<b>2</b>	<b>15</b>	<b>10</b>	<b>2</b>	<b>18</b>	<b>7</b>	<b>2</b>	<b>19</b>	<b>6</b>	<b>6</b>	<b>52</b>	<b>23</b>
<b>%</b>	<b>7,4</b>	<b>55,6</b>	<b>37,0</b>	<b>7,4</b>	<b>66,7</b>	<b>25,9</b>	<b>7,4</b>	<b>70,4</b>	<b>22,2</b>	<b>7,4</b>	<b>64,2</b>	<b>28,4</b>
<b>Appraisable (in %)</b>	<b>63,0</b>			<b>74,1</b>			<b>77,8</b>			<b>71,6</b>		

## Propriety

	Denmark			The Netherlands			Germany			overall		
	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈
P1 Formal Agreement (max. 11 Points)			11	2	3	6	1	10		3	13	17
P2 Protection of Individual Rights (max. 12 Points)	1	9	2		10	2		11	1	1	30	5
P3 Complete and Fair Investigation (max. 10 Points)	2	6	2	4	5	1		10		6	21	3
P4 Unbiased Conduct and Reporting (max. 2 Points)		2		1	1		1	1		2	4	0
P5 Disclosure of Findings (max. 11 Points)	2	8	1	4	7		2	7	2	8	22	3
<b>Total</b>	<b>5</b>	<b>25</b>	<b>16</b>	<b>11</b>	<b>26</b>	<b>9</b>	<b>4</b>	<b>39</b>	<b>3</b>	<b>20</b>	<b>90</b>	<b>28</b>
<b>%</b>	<b>10,9</b>	<b>54,3</b>	<b>34,8</b>	<b>23,9</b>	<b>56,5</b>	<b>19,6</b>	<b>8,7</b>	<b>84,8</b>	<b>6,5</b>	<b>14,7</b>	<b>66,2</b>	<b>20,6</b>
<b>Appraisable (in %)</b>	<b>65,2</b>			<b>80,4</b>			<b>93,5</b>			<b>80,9</b>		

## Accuracy

	Denmark			The Netherlands			Germany			overall		
	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈
A1 Description of the Evaluand (max. 11 Points)	3	6	2	4	6	1	1	8	2	8	20	5
A2 Context Analysis (max. 11 Points)	3	8		5	6		2	9		10	23	0
A3 Described Purposes and Procedures (max. 12 Points)	9	2	1	7	4	1	1	8	3	17	14	5
A4 Disclosure of Information Sources (max. 11 Points)	6	5		3	6	2	1	10		10	21	2
A5 Valid and Reliable Information (max. 22 Points)	12	9	1	12	8	2	6	12	4	30	29	7
A6 Systematic Data Review (max. 1 Points)			1			1			1	0	0	3
A7 Analysis of Qualitative and Quantitative Information (max. 21 Points)	12	7	1	15	5		10	9	1	37	21	2
A8 Justified Conclusions (max. 11 Points)	8	3		4	7		1	10		13	20	0
A9 Meta-Evaluation (max. 11 Points)	10		1	10		1	10		1	30	0	3
<b>Total</b>	<b>63</b>	<b>40</b>	<b>7</b>	<b>60</b>	<b>42</b>	<b>8</b>	<b>32</b>	<b>66</b>	<b>12</b>	<b>155</b>	<b>148</b>	<b>27</b>
<b>%</b>	<b>57,3</b>	<b>36,4</b>	<b>6,4</b>	<b>54,5</b>	<b>38,2</b>	<b>7,3</b>	<b>29,1</b>	<b>60,0</b>	<b>10,9</b>	<b>47,0</b>	<b>44,8</b>	<b>8,2</b>
<b>Appraisable (in %)</b>	<b>93,6</b>			<b>92,7</b>			<b>89,1</b>			<b>91,8</b>		

## Overall assessment

	Denmark			The Netherlands			Germany			overall		
	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈	<input type="checkbox"/>	<input checked="" type="checkbox"/>	≈
<b>Total</b>	<b>82</b>	<b>127</b>	<b>57</b>	<b>85</b>	<b>143</b>	<b>38</b>	<b>45</b>	<b>191</b>	<b>30</b>	<b>212</b>	<b>461</b>	<b>125</b>
<b>%</b>	<b>30,8</b>	<b>47,7</b>	<b>21,4</b>	<b>32,0</b>	<b>53,8</b>	<b>14,3</b>	<b>16,9</b>	<b>71,8</b>	<b>11,3</b>	<b>26,6</b>	<b>57,8</b>	<b>15,7</b>
<b>Appraisable (in %)</b>	<b>78,6</b>			<b>85,7</b>			<b>88,7</b>			<b>84,3</b>		



## **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 eV. Arbeitsgemeinschaft Sozialwissenschaftlicher Institute eV. Berufsverband Deutscher Markt- und Sozialforscher eV. 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. Bern (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 ([www.evaluationcanada.ca](http://www.evaluationcanada.ca)).
- Cronbach, L.J. et al. (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.
- Hellstem, 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.planunganalyse.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 Humandienstleistungen. 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. 105-1 31.
- Lamnek, G. (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.

