

# Evaluation

<http://evi.sagepub.com/>

---

## Peer Evaluation and its Blurred Boundaries: Results from a Meta-evaluation in Initial Vocational Education and Training

Sandra Speer

*Evaluation* 2010 16: 413

DOI: 10.1177/1356389010382265

The online version of this article can be found at:

<http://evi.sagepub.com/content/16/4/413>

---

Published by:



<http://www.sagepublications.com>

On behalf of:

THE  
TAVISTOCK  
INSTITUTE®

The Tavistock Institute

**Additional services and information for *Evaluation* can be found at:**

**Email Alerts:** <http://evi.sagepub.com/cgi/alerts>

**Subscriptions:** <http://evi.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.com/journalsPermissions.nav>

**Citations:** <http://evi.sagepub.com/content/16/4/413.refs.html>



# Peer Evaluation and its Blurred Boundaries: Results from a Meta-evaluation in Initial Vocational Education and Training

Evaluation

16(4) 413–430

© The Author(s) 2010

Reprints and permission: sagepub.

co.uk/journalsPermissions.nav

DOI: 10.1177/1356389010382265

<http://evi.sagepub.com>



**Sandra Speer**

Independent Evaluator, Germany

## Abstract

Peer evaluation has proved to be a popular approach with both peers and the evaluated, but there has been considerable variation in the ways in which peer evaluations have been implemented. There are different forms, purposes and ways of organizing peer evaluation. Peer evaluations are also not uniform and links to other evaluations can be made. After providing a general overview of different kinds of peer evaluation, this article focuses on the results from a meta-evaluation of peer evaluation in Initial Vocational Education and Training (IVET). This meta-evaluation was partly inspired by existing evaluation standards and covered fifteen peer-review pilots in eight European countries. First, the scope of the peer team and the underlying evaluative framework are discussed. Second, the nuances of implementation and the conflicts that exist are highlighted. The article concludes that, under certain conditions, peer evaluation and self-evaluation provide a powerful tandem to produce empirically grounded, context-sensitive information to support school improvement.

## Keywords

meta-evaluation; peer evaluation; peer review; self-evaluation; vocational education and training

## Introduction

Peer review may well be one of the oldest forms of formal evaluation.

Peer review arouses very diverse emotions, beliefs, and ambitions. It angers, it reassures, it intimidates, it tramples egos, and it puffs them up. For some, peer review demonstrates the vacuousness and unreliability of social science; for others, the substance and reliability of social science. (Starbuck, 2003: 348)

---

### Corresponding author:

Sandra Speer, Untere Matthias-Claudius-Str. 11, D-65185 Wiesbaden, Germany

Email: [sandra\\_speer@t-online.de](mailto:sandra_speer@t-online.de)

Especially in the context of scientific publications, peer review has been intensively discussed.<sup>1</sup> Peer evaluation is, however, not limited to the evaluation of publications. It is an area that is expanding and increasingly peer evaluation is being used with other evaluation objects and in other policy fields, especially in the field of education.

The term 'peer evaluation' has not been consistently defined and there is no single method but rather a family of methods. The terms 'peer review' and 'peer evaluation' are often used synonymously, as they are in this text, but the former is more widespread. In general, peer review is structured feedback between colleagues or individuals with an equal standing and constitutes an expertise-oriented approach by peers who are usually not trained in evaluation. Their qualifications stem from the field of practice in which they work. In more traditional forms of peer review, a peer's expertise and the authority gained may be of primary importance, whereas in modern forms of peer review, organizational elements, structured systems for data gathering and feedback have been developed and often guide the work of groups of peers.

Peer evaluation lies between self- and external evaluation. It often interplays with other forms of evaluation and is subject to a distinct set of evaluation process rules. As with any other type of evaluation, the evaluation object can be situated on the micro-, meso- or macro-levels and can have ex-ante, ongoing and ex-post variants. Many forms of peer evaluation begin with the examination of a self-evaluation document, which becomes the basis for defining peer evaluation questions and areas for further investigation by peers. The peers usually only have time for a quick scan of such documents and are unable to develop their own studies, which forces them to partly rely on existing documents while also questioning them. In addition to documentary review, peers often use qualitative evaluation techniques, such as interviews, focus groups and observations. So, compared with traditional evaluation approaches, peer evaluation is in most cases a 'lighter' form of evaluation.<sup>2</sup>

Doubts regarding the usefulness of peer review have often been raised in terms of acceptance, bias, reliability and validity (e.g. Kornhuber, 1988: 377). Scriven (1991: 255) calls it 'extremely shaky'. However, the following carry out peer evaluations: large international organizations like the OECD (e.g. Pagani, 2002; Lehtonen, 2005) and the European Commission (e.g. Peer Review in Social Inclusion, 2005, 2009) as well as groups of schools or teachers. This article discusses peer evaluation and its underlying tensions and is based on a meta-evaluation of peer evaluation in initial vocational education and training (IVET), which was conducted by the author.

First, some prominent characteristics and forms of peer evaluation are discussed. This text cannot provide a comprehensive overview of peer evaluation, but can highlight some of its important forms, especially in relation to the field of education. Then, the design and results of a meta-evaluation of one modern form of peer evaluation – implemented in initial vocational education and training throughout Europe – will follow. The peer evaluation mechanisms – including their contexts, underlying rules and institutional settings – will be discussed on the basis of this meta-evaluation.

## Characteristics of Peer Evaluation

In evaluation theory, the peer evaluation approach is either neglected or just reflected upon very briefly.<sup>3</sup> In the *Encyclopedia of Evaluation* (Mathison, 2005: 300)

... peer review refers generally to the evaluation of professional performance or products by other professionals and, more specifically, to a set of procedures for evaluating grant proposals and manuscripts submitted for publication. For peer-reviewed journals, content-matter specialists are asked to judge a manuscript, often using specified criteria and are blinded to the author's identity.

This definition is mainly linked to the peer review of journals and grants, although the concept is used much more widely and might rely on different underlying procedures and assumptions.

Weiss (1998: 189–90) classifies peer evaluation as expert judgement among the ‘informal designs’. In particular, she points out that peers are probably better at judging the procedures and practices of a programme than judging the outcomes. Evaluating processes provides the peers with the opportunity to draw on their knowledge of other programmes and experiences. Peers might have difficulties in judging outcomes, especially when good data are not available. They may judge a situation in comparison with their own situation or with other situations known to them from practical experience, rather than comparing it with an ideal type of a situation. So peers may use subjective standards that differ from one peer to another and that are sometimes hard to articulate, ultimately leaving their judgements open to dispute.

Only Vedung (1999: 31–2) acknowledges peer review among the nine most prominent evaluation models. According to Vedung, peer review is a ‘professional model’, like self-evaluation, and it is defined by the person who is undertaking the evaluation, whose criteria stem from professional standards and goals. He points out that peer review is an institutionalized evaluation model, which determines how to organize the evaluation but not the questions to ask. The choice of the criteria and standards to be used remains at the discretion of colleagues. Vedung emphasizes that this method – although it is highly subjective – may be the most suitable for complex and complicated disciplines and for evaluating the quality of their output.

Thus, Weiss acknowledges peer review as a form of expert judgement mainly for evaluating processes, whereas Vedung stresses more its suitability for evaluating the quality of output.

Both approaches are integrated in the following discussion of internal and external evaluation stances; attempts at structuring and standardizing; links between self and external evaluation aspects; and the opportunities for learning and reciprocity through peer evaluation.

### *Insider and Outsider ‘Stances’*

First, peer evaluation can be located within the praxis or profession and be organized *inside* the same organization, i.e. evaluating the organization from which the peers are drawn, or parts of it. Peer evaluation of teachers is a classic example of peer evaluation within an organisation, where peers evaluate each other on a reciprocal basis. This kind of peer evaluation can be organized quite spontaneously on a voluntary basis or become routinized as part of peer evaluation systems. When the peers are confined within a particular organizational world-view, in which structures and processes are ‘normal’ for insiders, and even more so if they also lack qualifications in evaluation design and empirical methods, the peer might be considered a ‘prophet in his own country’. The obligations and identities of the peers are also problematic. In terms of obligations, the evaluated and the peers might favour reports made on a reciprocal basis, creating the potential for bias. Regarding identity, the person being evaluated might be afraid of ‘being spied on’ by colleagues/peers.

But peer evaluation is also often characterized by the evaluation of programmes from *outside* their institution, for example by professionals like university professors. The evaluator is often an outsider who may carry out a variety of functions. But this peer will be less able to ensure that the evaluation results will have any effect on future behaviour, in contrast to a peer who comes from inside the organization and who can try to foster the utilization of the peer evaluation results. However, the boundary between peers from within the praxis or profession and those from outside the praxis or profession is *blurred* in some cases. For example, a peer evaluation team may consist of peers from broader disciplinary backgrounds as well as evaluators who are integrated into the

teams and serve as methods experts. Professional evaluators then help the peer who is an expert in the field. Although these experts cannot be regarded as peers as such, they are a part of the peer evaluation process and in such cases the majority of the members of the peer team are usually considered 'real' peers. The peer evaluation perspective is extended mostly due to the peers' lack of methodological training and a general desire to integrate broader perspectives.

### *Structuring and Standardizing Peer Evaluations*

External programme evaluations often do not use specific rules and approaches. Self-evaluation can also be a very open process; institutions or professionals often choose the most appropriate approach by first referring to the relevant evaluation questions. In the field of education, there is a long-standing tradition of self-evaluation in the teaching profession (McNamara and O'Hara, 2004). Sometimes there are rules for choosing certain ways of self-evaluation and quality management on the organizational level, for example, the European Foundation for Quality Management (EFQM) model for school self-evaluations. Peer evaluations can also be standardized to a certain extent. Instead of a non-systematic individual peer evaluation, a peer review system can be constructed with structuring rules, clearly defined tasks for the peers and evaluative criteria. Predominantly, peer reviews are organized through a network of peers and institutions, which want or ought to be peer-evaluated; a coordinating body then manages these peer reviews. This is because the timeframe for peer evaluation is much shorter than for any other form of evaluation. Rules are often laid down in guidance documents which define who the peers are and who selects them. The documents also establish methods and areas for evaluation, and sometimes they also include criteria and details about the organization of the peer review. The peers then have 'choices within the rules' but not a 'choice of rules'.<sup>4</sup> This kind of structured peer review can lead to more transparency in the evaluation process.

### *Possible Links between Self-evaluations and External Evaluations*

Individual evaluation is no longer the sole form of evaluation. Streams of information (Rist and Stame, 2006) are often integrated into some form of evaluation system, in which peer evaluation builds on the self-evaluation perspective or the external evaluation perspective. Since evaluation plays an ever more important role at various levels, single policy fields have become permeated by evaluation systems (Leeuw and Furubo, 2008). For example standardized testing and measurement of scholastic achievement has increasingly been introduced into educational fields, in recent years. Sometimes peer evaluation has links to macro external evaluations and benchmarking systems, and it can thereby constitute an additional level of evaluation within an overall evaluation system. For more than a decade self-evaluation has been implemented in the education sector, at both the level of the school and the classroom. Self-evaluation and quality management are newer to the VET sector (Di Battista et al., 2009) and have often been introduced in conjunction with the European Quality Assurance Reference Framework for VET (EQARF) (EC, 2008). A peer evaluation may start by examining a self-evaluation document as already noted, but may be reinforced by incorporating other self-evaluations, especially when the validity of self-evaluation is contested (Kyriakides and Campbell, 2004: 29). In these circumstances peer evaluation can be seen as incorporating meta-evaluative components, even if it is not a meta-evaluation as such.

Various evaluation approaches can also be combined with one another. An external evaluation may be driven by a need for accountability and it can deliver comparative data, which then can also be used as a management tool for benchmarking (Speer, 2001: 64). A peer evaluation may combine

this with an improvement perspective on the micro-level and has a more immediate purpose. Dialogical elements are more focused on the internal stakeholders and their wish to improve quality. Combining different evaluation approaches is especially useful when they each contribute to different evaluative goals.

Evaluation always creates its own incentives and peer evaluation is no different. Peer evaluation can affect incentives especially for the improvement of self-evaluation activities and also for expanding its scope (Nevo, 2001: 98). This is sometimes also the logic of inspections (Ehren et al., 2005: 63). The peer evaluation can also benefit from self-evaluation insofar as the scope of the peer evaluation may be expanded or the interpretation of findings may be improved and the utilization of the peer evaluation findings may be increased (Nevo, 2001: 99–100). However, it might also be the case that peer evaluation is crowding out other forms of programme evaluation. For example, in the case of higher education in Canada, Cabatoff (2001: 82) reports that there is a trend away from programme evaluation and a shift towards accreditation based on peer review.<sup>5</sup>

### *Learning within Peer Evaluations*

The audience for peer evaluation might vary from specific users – for example, the evaluated – to the general public. In the latter case, peer evaluation could be a basis for a public debate, whereas in the first case the results might be kept confidential. Within peer evaluation, two main concepts coexist concerning the direction of the learning process. First, the classic approach to peer evaluation is where peers evaluate the evaluandum. The evaluated receives feedback either regarding a decision on resources (accountability) or for capability development. In a second type of peer evaluation, peers receive information on the previously evaluated evaluandum. Afterwards, the peers can learn from the evaluandum. This type of peer evaluation is also called ‘peer-learning’ or ‘learning from good practice’. The latter is, for example, practised on the macro-level in various single policy fields within the European Open Method of Coordination (OMC), including VET. Here peer learning should ‘replace the traditional indirect approach of policy decision in which policy advisers, through the preparation of reports and recommendation, inform policy decision makers’ (ENQA-VET, 2009: 2).

### *Reciprocity within Peer Evaluations*

On the one hand, peer evaluation has the reputation of being a weak instrument, because peers are embedded in the value system of the evaluated and because of the peers’ self-interest, which is sometimes reinforced through a mechanism of reciprocity – an individual is a peer on one day and then is evaluated the next day. This may create conflicts of interest. So the evaluation of peers might be especially at risk of bias. However, if the main goal of peer evaluation is learning, win–win situations can be created for both the peer and the evaluated. Leeuw (2002) stresses the need for balancing independence and reciprocity as a prerequisite for building social capital and trust between the parties involved in educational evaluations. He distinguishes two kinds of reciprocity. The first dimension is the ‘give-and-take-dimension’, i.e. giving information and learning. The second dimension is the ‘you-too-me-too-dimension’ characterizing reciprocity in terms of the transparency and evaluation that the one institution asks from the other. This variant of reciprocity is not always present, as for example in school inspection in Europe. This is especially interesting as there is a tendency for activities related to school inspections to be further developed into activities that are closer to the concept of peer evaluation.

**Table 1.** Quality Areas within the IVET Peer Review

---

Quality Area 1: Learning and Teaching
Quality Area 2: Curricula
Quality Area 3: Assessment and Tests
Quality Area 4: Social Environment and Accessibility
Quality Area 5: Infrastructure and Financial Resources
Quality Area 6: Management and Administration
Quality Area 7: Institutional Ethos and Strategic Planning
Quality Area 8: Staff Allocation, Recruitment and Development
Quality Area 9: Working Conditions of Staff
Quality Area 10: External Relations and Internationalization
Quality Area 11: Social Participation and Interactions
Quality Area 12: Learning Results and Outcomes
Quality Area 13: Gender Mainstreaming
Quality Area 14: Quality Management

---

Source: Peer Review in Initial VET (2005)

## Peer Review in Initial Vocational Education and Training in Europe

The following analysis and discussion is based on a meta-evaluation of a pilot project on peer review conducted by the author (Speer, 2007).

### *The Peer Review Pilots*

Within the project 'Peer Review in Initial Vocational Education and Training', a European peer review procedure has been developed into a transnational cooperation between twelve European countries as a European instrument for quality assurance and improvement at the institutional level of single IVET providers. Here IVET is understood to be formal vocational education that is usually completed before entering work. The procedure was tested in fifteen transnational peer reviews in eight European countries ('peer review pilots'). This was organized as a mutual network: the IVET providers have been involved in the project as so-called 'operative partners' from the beginning. The peers were prepared through participation in web-based training. Over two to three days, a group of four peers was invited to assess the quality of different fields of the IVET institution. The quality areas for peer evaluation could be chosen from fourteen predefined areas (see Table 1).

The eight different European countries in which the IVET providers are located have very different evaluation cultures and represent various forms of organization, both public and private as well as both small specialized and large multifaceted providers. The underlying peer review approach and methodology is described in the 'European Peer Review Manual' (Peer Review in Initial Vocational Education and Training, 2005; Gutknecht-Gmeiner, 2007). Like most peer evaluations, it consists of the four phases outlined in Box 1.

The Peer Review Manual has been developed by a subgroup of the peer review project team, so the criteria stem from 'within', which is often the case for accreditation systems. As it was a pilot phase, wider implementation in Europe has not occurred yet. The reviews are not currently planned to be repeated at specified intervals.

**Box 1.** Four phases of peer evaluation

1. The preparatory phase:
  - the peer review is organized,
  - the peers must be recruited and trained,
  - a self-report is sent to the peers,
  - the peers prepare the leading questions,
  - arrangements are made for the peer visit.
2. The peer visit:
  - the peers carry out the evaluation on-site and
  - provide feedback to the provider at the end of the visit.
3. The reporting:
  - the peers write a draft report, which may be commented on by the provider and then
  - the final report is issued.
4. The improvement:
  - results and recommendations from the peer review should be transferred into concrete actions for improvement,
  - these are planned and implemented.

*The Blurred Boundaries of the Peers*

The Peer Review Manual states that the peer review team should be composed of four peers and at least half of the peers should be ‘real’ peers, defined as persons of equal standing. Additionally, it is recommended that one member of the peer team should be an ‘evaluation expert’ with expertise in evaluation, moderation and communication. Also, this team member might work within the specific field of practice under review but he or she may also extend the peer evaluation perspective, if the (peer) evaluation expert is external to the praxis. It is also possible for one or more members of the peer team to be representatives of stakeholders. One of the four peer team members should be competent in gender mainstreaming and the peers should have diverse backgrounds and values. It was rarely clear cut whether the underlying values stem more from outside or inside the evaluandum. Thus, the peer team may end up being composed of ‘real’ peers as well as by ‘evaluation methods and techniques experts’, which ultimately extends the traditional peer perspective, as it makes systematic data gathering more likely (see Table 2).

The IVET peer evaluation system combines some of the advantages of both internal and external evaluation: the knowledge of and familiarity with the evaluandum and the stakeholder groups of insiders as well as the impartiality, objectivity, credibility, the outside perspectives and the evaluation expertise of external peers. Additionally, the possible inclusion of a ‘stakeholder peer’ introduces a *participatory* element into peer evaluation. Thus stakeholder peers can be drawn from ‘indirect beneficiaries and injured parties’ or ‘civil society and citizens’ and are equally involved in the peer evaluation process. When looking at the level of participation, one can conclude that they have a high ‘extent of involvement’ in this kind of peer evaluation (see Daigneault and Jacob, 2009).

The extension of the peer team to include ‘outsiders’ can increase the credibility of the peer review. Stakeholders have different interests in the peer review process to the ‘real’ peers. The inclusion of

**Table 2.** Members of a Peer Team

Members of a peer team	Occupational background	Required competencies	Evaluative positions
'real' peer	<ul style="list-style-type: none"> <li>- teachers and other professionals from IVET schools</li> </ul>	<ul style="list-style-type: none"> <li>- expertise in the review topics under scrutiny</li> <li>- teaching experience (at least 5 years)</li> <li>- experience with quality management and quality development</li> </ul>	<ul style="list-style-type: none"> <li>- references from outside the school under review</li> <li>- knowledge of the working conditions and framework, in which the provider operates (the same IVET system/country)</li> </ul>
stakeholder peer	<ul style="list-style-type: none"> <li>- representative of a stakeholder group, like a company rep of the region or a labour union rep, parents, etc.</li> </ul>	<ul style="list-style-type: none"> <li>- knowledge of the quality areas under review</li> <li>- experience from other quality management systems</li> </ul>	<ul style="list-style-type: none"> <li>- more an outsider view to the teaching praxis and profession</li> <li>- is most likely familiar with other stakeholder groups and their interests</li> <li>- knows the environment in which the provider operates</li> </ul>
transnational peer	<ul style="list-style-type: none"> <li>- one of those mentioned, usually with strong VET background</li> </ul>	<ul style="list-style-type: none"> <li>- inside knowledge of the quality areas under review</li> <li>- teaching experience</li> <li>- experience with the implementation of quality management (systems)</li> </ul>	<ul style="list-style-type: none"> <li>- more of an outsider view because of large differences between national VET systems, school cultures and evaluation cultures</li> <li>- insider view from own teaching experience</li> </ul>
evaluation expert	<ul style="list-style-type: none"> <li>- evaluator; quality manager; ideally at the same time with experiences from evaluation in the field of education, such as higher education or VET</li> </ul>	<ul style="list-style-type: none"> <li>- knowledge of the quality areas under review</li> <li>- expertise in evaluation methods</li> <li>- expertise in moderation and communication techniques</li> </ul>	<ul style="list-style-type: none"> <li>- more of an outsider perspective;</li> <li>- a great distance from provider under review;</li> <li>- references from v. different projects possible</li> </ul>
gender mainstreaming	<ul style="list-style-type: none"> <li>- one of those mentioned</li> </ul>	<ul style="list-style-type: none"> <li>- additionally: expertise in gender mainstreaming</li> </ul>	<ul style="list-style-type: none"> <li>- as above</li> </ul>

Source: Peer Review in Initial VET (2005) and own elaborations

'stakeholder peers' and 'evaluation experts' in addition to the 'real' peers might also signal to teachers as well as to other stakeholder groups a strong desire for change management and improvement.

Ideally a peer evaluation process should be a win-win situation for peers and IVET providers, in which both learn from each other. The concept of reciprocity ('give and take' dimension) is crucial here to create a trust-building relationship between the evaluated and evaluators. So it is very different to other peer evaluation systems, which are expected to exert their influence by using 'peer pressure' as a mechanism for soft persuasion or even coercion. The peers in this peer evaluation

system have no power to undertake any decisions, for example, on accreditation, funding or even any terminations.

The underlying meta-evaluation identified differences between the pilot peer reviews; and tensions that arose within the various forms of the peer review implementations that were chosen; these will be described here. The more detailed results can be found in Speer (2007).

### *Evaluation Design of the Meta-evaluation*

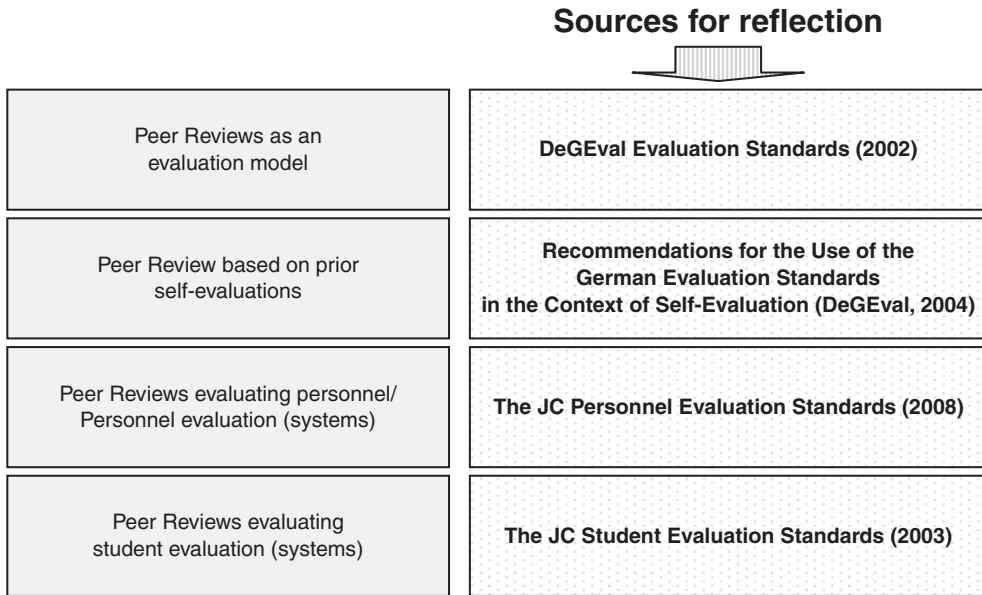
The evaluation of the peer review pilots is seen here as a *meta-evaluation* because the peer review is itself an evaluation system. Scriven (2009: p. iv) calls meta-evaluation ‘the consultant’s version of peer review’. Following his view, the meta-evaluation of peer review would be a peer review of peer review. If peer evaluation is interpreted as meta-evaluation itself – because it may partly evaluate previous self-evaluations – then the meta-evaluation of peer evaluation can also be regarded as a meta-meta-evaluation.

This meta-evaluation is formatively focusing on the peer review design as laid down in the Peer Review Manual and its implementation during the peer review pilots. The meta-evaluation analysis was undertaken by the author using a number of evaluation standards. The meta-evaluation of the peer review pilots involved triangulating sources. After each pilot, the peers were asked to summarize their experiences of the peer review procedure as well as of the peer training programme. Second, the peers were asked to provide feedback, with a greater focus on the further development of the peer review procedure. Feedback was gathered by telephone interviews for the first five pilots and in writing for the subsequent pilots. The peer review facilitators, who organized the peer review within the IVET institutions, were also asked for their suggestions regarding possible improvements to the process; and the directors of the IVET providers also provided feedback on the peer review process and the benefits of the pilots. The mostly qualitative answers from the questionnaires and the telephone interviews were analysed by grouping the text around dominant issues. Additionally, the ‘self-reports’ (provided by the IVET providers prior to the peer review) and the peer review reports (written by the peers as feedback for the IVET providers under review) were analysed so that areas could be identified where the providers and the peers needed additional information and where the procedures might require adjustment.

### *Reflections on the Peer Review Manual Guided by Evaluation Standards*

For a critical analysis of the peer review method in IVET, as specified within the Peer Review Manual, four different evaluation standard sets were used for meta-evaluation.<sup>6</sup> This IVET peer review method is an evaluation model itself; it is based on prior self-reports and sometimes personnel or personnel evaluation systems were included in the peer review. They could also include the evaluation of student evaluations or student evaluation systems.

The Peer Review Manual has been reflected upon by using four different standard sets (see Figure 1). These different sets of standards are organized around four important attributes of sound evaluation practice: propriety, utility, feasibility and accuracy. The evaluation standards that stem from the four different sets of standards have been used to guide the reflections of the author for highlighting where possible pitfalls or tensions may lie when implementing the strategies mentioned in the Peer Review Manual. All these standards can be regarded as ‘*ideal*’ standards: it is not possible to fulfil them all equally at the same time, as is true for any peer review handbook. Between some of the single standards trade-offs exist and must occur. However, these standard sets can be used to reflect on and improve current practices.



**Figure 1.** Relevant Sets of Evaluation Standards

Four evaluation standards from the DeGEval Evaluation Standards (2002) have been identified as especially important for the peer review procedure as it is laid down in the manual. These are:

- ‘Information Scope and Selection’ (U4),
- ‘Transparency of the Values’ (U5),
- ‘Complete and Fair Investigation’ (P3) and
- ‘Unbiased Conduct and Reporting’ (P4).

The latter is especially important because there is a tendency to regard peers as being too close to the evaluandum. But also the number of peers and the composition of the peer team will have considerable influence over ‘P3’ and ‘P4’. The interaction with the IVET institution and the quality of the evaluation design will be crucial for ‘U4’ and ‘U5’.

Based on the *Recommendations for the Use of the German Evaluation Standards in the Context of Self-Evaluation*<sup>7</sup> (DeGEval, 2004) the ‘context’ of the IVET provider should be described in detail (‘Context Analysis’, A2). Furthermore, it should be mentioned whether the self-report is based on a voluntary or a compulsory self-evaluation and the underlying evaluation questions should be described in detail (‘Described Purposes and Procedures’, A3) as well as possible different opinions between various stakeholder groups (‘Transparency of values’, U5). The sources of information used in the self-report should be described in enough detail that the reliability and adequacy of the information can be assessed (‘Disclosure of Information Sources’, A4). Critical information must not be given to the peers (‘Protection of Individual Rights’, P2).

On the basis of the *Personnel Evaluation Standards* (JC, 2008), there are two aspects that could be highlighted more within the peer review manual. First, the qualification of the peers (‘Evaluator Qualifications’, U3) should be emphasized. This may be in terms of principles for sound personnel

evaluation; this is sometimes relevant in the case of classroom observations. Alternatively, peer knowledge of national, or rather regional, IVET systems – and in particular their compulsory evaluation systems – could be included, in relation to preparation for a peer visit. This would be important background information, although the peer evaluation would never include a personnel evaluation as such. Second, for observations of the ‘validity orientation’ (A1) the following issues that may be crucial for validity were identified: decisions about scheduled or unscheduled observations; determining whether the persons under review are informed about the evaluation questions or not; and establishing who is selecting the classes or teachers under review.

On the basis of reflections on *Student Evaluation Standards* (JC, 2003), the following points regarding the peer review procedure are relevant. Within the ‘Quality Area 3 Assessment and Test’ it could be stressed that:

- the evaluation of students should provide information that identifies strength and weaknesses (‘Balanced Evaluation’, P6) and
- the student evaluation procedures should be practical, efficient and non-disruptive (‘Practical Orientation’, F1).

Many other areas identified during these reflections were also considered when undertaking the following survey.

## Results and Tensions within the Peer Review Pilots

The meta-evaluation was based on various sources, as already explained. Here, the results from the meta-evaluation are presented and discussed within the context of the dominant issues. The response rates were as indicated in Table 3.

**Table 3.** Sources of Meta-Evaluation

Sources <sup>x</sup>	cases	target sample <sup>y</sup>	response	response rates
(a) Meta-evaluations	15	15	8	53%
(b) Experiences peers	63	63	34	54%
(c) Telephone interviews peers	23	20	14	70%
(d) Questionnaires peers	34	29	18 <sup>z</sup>	62%
(e) Questionnaires facilitators	15	13	6	46%
(f) Questionnaires for OP reports	15	15	3	20%
(g) Questionnaires directors	15	15	5	30%
(h) Self-reports	15	15	15	100%
(i) Peer review reports	15	15	9	60%

<sup>x</sup> (a), (b) and (f) distributed by the project management; (c), (d), (e), (g) carried out by the evaluator; (h) prepared by the providers and (i) has been written by the peers; (h) and (i) based on the procedures of the Peer Review Manual. Therefore the instruments developed by the evaluator and the other questionnaires and documents have been analysed to give a complete picture of the experiences from the pilot projects.

<sup>y</sup> The difference of 10 cases between the total number of cases and the targeted number of interviews results from the fact that some peers participated in more than one pilot. Their experiences from all their pilots were covered in the same interview.

<sup>z</sup> Some peers filled out one questionnaire but were reflecting on more than one pilot they participated in. This fact has lowered the response rate to 62%; the experience of more peers and pilots are reflected in these 18 questionnaires. Source: Speer (2007).

When the peers were asked about the lessons learned from the peer reviews, many mentioned that they reflected on their own competencies and ways in which they could improve their own practices. Most of the peers would like to be involved again and gave positive feedback. The realization that other providers are not doing much better was useful, and examples from the peer review have already been used for staff development training. The self-interest and recompense of the peers for acting as peers, according to their own comments, took the form of pleasant travel, some money, learning and raising their professional image and networking. The enhancement of their own reputation can be seen as a trade-off with pleasing the institution being evaluated. Usually the peers are not full-time peers and they would not risk damaging their careers if they were not recommended to continue as peers, which is possible within this system.

The staff of the IVET providers reacted overall very positively. They appreciated the fact that they could have discussions with their peers instead of an 'audit authority' and the atmosphere was reported to be very open and full of ideas. When the IVET providers were asked about their benefits, they answered that their own goals for initiating the peer review were achieved. A positive effect was, for example, that teachers and staff shared the goals set by the head of the school. In some cases, the implementation of changes has already started with respect to planning and setting up preliminary schedules for areas that need improvement. The best practices from other schools could be shared and contact with schools abroad could be established. The positive effects reported by one provider included a constructive, critical and independent external analysis of their business practices.

*Simple versus professional peer reviews.* For some IVET providers, the peer review procedure was the starting point for a general quality improvement process in their institution; for other IVET providers the peer review was a new tool that was added to their already elaborate evaluation culture. The beginners faced a heavy workload while preparing the self-reports whereas experienced IVET providers use self-evaluations routinely in formative and summative ways. Therefore it is easy for the latter to introduce the peer review as a new method because they already have a sound basis of data and knowledge on how to organize similar reviews. The experienced providers were better prepared and were better at 'steering their peers'.

*Transnational versus national peers.* For some peer reviews, the outside view/perspective of the transnational peer was a *conditio sine qua non* for gaining new experiences, insights and ideas. The transnational peer could ask questions about topics that the national peers would perhaps take for granted. The transnational peer is – perhaps implicitly, perhaps explicitly – relying on a different set of standards and values. Furthermore, the transnational peer is 'allowed' to ask questions that nationals would not ask in such a direct manner, because the nationals may not consider it to be politically correct. For others, the lack of proficiency in the national language and/or of knowledge concerning the local IVET system turned out to be a crucial weakness of the transnational peer. In some cases, the transnational peer seemed to be the 'weakest' peer review team member, and in others, the most inspiring one. Most of the peers described experiencing good cooperation with the provider and only a few peers felt unduly influenced. In one case a peer felt under pressure to 'conform'. That was mainly concerning sampling and the freedom to operate outside the provider's remit.

*Many versus few selected quality areas.* Fourteen quality areas are defined in the Peer Review Manual and they provide a framework for the evaluation questions, which are then formulated by the IVET provider. Some pilot peer reviews were designed as broad peer reviews with many quality areas, while others focused on specific questions within a few selected quality areas. Only very

experienced peers are able to complete sufficient reviews of more than three areas within two days. Most of the peers recommend only reviewing two or three quality areas in this amount of time.

*All versus selected criteria.* The peers are in the position to choose the criteria for their judgements; these will be in line with the encompassing framework of quality areas and the evaluation questions suggested by the evaluated. In some cases the peer review was more standardized, using all the criteria of the quality areas selected for a general review by the providers. The questions to be asked by the peers were then also derived with the help of the indicators mentioned in the Peer Review Manual. Other peer reviews focused responsively with respect to specific questions asked by the IVET provider. The use of all criteria and indicators for certain quality areas would allow for better comparability and possible future certification. On the other hand, some IVET providers are especially interested in the review of specific review questions and not in the review of all criteria and indicators.

*Few versus many review techniques.* Because of the tight timeframe of the visits, the choice of techniques was limited. Each of the pilot peer reviews consisted of many interviews – mostly in the form of focus groups, which included teachers, students, company representatives, alumni, management and parents. However, other techniques were also used, for example, observations: tour of the school, classroom observations, company visits, photo-evaluation and analysis of documents. In nine out of fifteen peer review visits, classroom observations were included in the peer review design. Most of the interviewed peers had the impression that the teachers were informed about the classroom observations beforehand. It was assumed by the interviewed peers that not all of the teachers were informed about the quality areas chosen for the review. Regarding the quantity of classroom observations, the peers recommended not having more than four classroom observations in the given timeframe (because otherwise the amount of data would become too great to analyse). It is important to get the balance right between time spent on the review, and time spent on documentation and reflection: when as much time as possible was used for the review itself, little time was left for documentation and reflection among the peer team and vice versa.

*Independent versus efficient sampling.* The IVET providers who requested the peer review took the initiative and set the agenda. The providers also defined the central issues and quality areas on which to focus. The peers had limited time and little knowledge of the local stakeholder groups, so the IVET providers gave them information on the stakeholder groups, and provided them with information by which representatives could be chosen and who was available at the time of the peer visit. This definitely resulted in scope to influence stakeholder involvement. For reasons of practicability, the IVET providers organized the interviews and determined the persons to be interviewed. This could lead to an unbalanced review if the provider does not look for a ‘representative’ mix of interviewees. In cases where gaining favourable PR for the institution is the driving force behind participating in the peer review process, there is considerable room for window dressing.

*Feedback for a small versus large group.* Different kinds of feedback sessions led to different effects: some peers gave their feedback to the management only; some to a wider stakeholder group; and other peers recorded a feedback video. The feedback varied from an analysis to proposals for organizational development. The peers have no influence on the diffusion of the written report; they can only suggest a format for immediate feedback after the peer visit.

## Overview and Conclusions

Many forms of peer evaluation approaches exist: they vary from traditional peer review, as often used in scientific fields, to more participatory approaches and systematically organized evaluation systems. Also, within the same system various nuances of peer review may occur, especially when the focus is on formative evaluation. It is rare for a peer evaluation to be carried out by a single peer. It is more common to involve groups of peers because they can combine different expertise and a collective evaluation would be more accurate than an individual evaluation. To a certain extent peers have to rely on self-evaluations completed beforehand and to build their evaluation upon previous results but they also have to question them. As the time allotted for a peer review is usually very limited, the evaluated will suggest or organize the meetings with the different stakeholder groups. The stakeholder selection might therefore be biased. As soon as systems of peer evaluation with structuring guidelines are set up, the leeway of the peers is limited in advance.

Concerning the evaluative 'stance', the IVET peer review contains many external perspectives, such as transnational peers, stakeholder peers and evaluation expert peers, who all can provide outside views. Within this pilot project, the peers were organized in a loose network. The peer reviews were formative and linked to previous self-evaluations by the organizations under review. It was the general intention to make the learning process reciprocal: the peers learned from the evaluation object and the organizations under review learned from the peer feedback.

The results of the meta-evaluation showed that each IVET provider, in collaboration with the relevant peer team, enacts the procedure differently. The European Peer Review Manual sets the framework but variations remain subjective, because the peer review procedure is not completely standardized and leaves room for interpretation as to how to implement it. Different institutions will emphasize different areas for review and some institutions will have a stronger commitment to the review process than others. If this kind of peer evaluation were to become compulsory for IVET schools, the incentives would certainly change and aspects of accountability within the peer review system would have to be strengthened, such as independence, the use of self-evaluations and the use of quality criteria. The rich experiences from the pilot studies illustrate the options available for peer review and show how the procedures are already 'alive'. The incentives not only depend on the form of the peer review system but also on the institutional circumstances in which the peer evaluation is implemented. So far the incentives for seeing peer evaluation as an opportunity for learning have dominated and it has been considered especially important to have teams comprised of mixed peers and have an international peer play an important role. Because of the indirect reciprocity between schools and peers, it is very likely that the evaluation will be on friendly terms. The peers will most likely help the directors improve teaching quality. In many cases, the peers learned new and useful information, as did the institutions under review.

For some providers the peer review is a learning exercise; for others it is an opportunity to gain favourable PR; and still for others it is an exercise of organizational development linked to previous self-evaluation activities; or perhaps even the other way round: participating in the peer evaluation leads to getting started with self-evaluation activities. Self-evaluation is not a popular activity; it is sometimes perceived as something threatening and as a form of social control (Vanhoof et al., 2009). Peer evaluation might help to open communication structures within schools and create more participatory initiatives, thereby enabling peer evaluations to lay the ground for further self-evaluations. Different interests of the participants can be incorporated into this kind of peer evaluation system.

During the last few years more formative and participatory peer-evaluation systems have become common, although older forms of peer review continue to exist in parallel. The summative peer

evaluations have more of a rationalistic tradition and the newer formative peer evaluations more of an argumentative tradition, in terms of giving feedback on the process-level and making a contribution to an informed debate. The latter makes it easier for the subjects of peer evaluation to accept recommendations from peers, sensitive to local circumstances, than from external evaluators. Newer models of peer evaluation usually have more diverse peer teams with greater stakeholder involvement.

In mainly standardized peer evaluation systems, the peers need only limited knowledge of evaluation theory and design. They mainly bring their expertise from the field and apply their social competence. In terms of less-structured peer evaluations, like in the IVET example, there is a greater need for some peers to be competent in evaluation. The evaluation design should be developed with flexibility, and in accordance with the evaluation questions and needs. So the spectrum of the roles for peers seems to be as big as it is for evaluators in general; the roles include everything from 'facilitator' to 'consultant' (Morabito, 2002), and from 'coach' to 'co-learner'. Despite the relevance of peer evaluation, surprisingly little empirical research is available that examines the variety of models. The challenge is to further meta-evaluate peer evaluation systems and to highlight what kind of approach is the most appropriate and what kind of conditions are required. Further investigations should also be undertaken to better understand the complementary existence of peer evaluation and other evaluation approaches and potential crowding-out effects.

### **Acknowledgments**

The author wishes to thank the participants of the underlying meta-evaluation, which was financed by LdV AT/04/C/F/TH-82000. Comments by two anonymous reviewers contributed to improving the text.

### **Notes**

- 1 The peer review of publications has always been a target of criticism. Many empirical analyses of peer reviews reveal that they diverge from one another, contradict each other and are not consistent over time (e.g. Cichetti, 1991; Gans and Shepherd, 1994; Campanario, 1998a, 1998b; Starbuck, 2005). Frey (2002: 4) puts it in these terms: the scientist needs to . . . prostitute oneself in order to publish and be academically successful. It seeks to overcome the veto power of (anonymous) referees, whose interests are not aligned to those of the journal, because they have no property right on the journal they advise, while the editor who does have a property right in the present setting is unwilling to overrule the referees.
- 2 The framework of peer evaluations often just allows for brief on-site visits of e.g. organizations or countries under review. Of course, qualitative evaluation techniques can also be very time-consuming. However, the point is that certain evaluation techniques and models will typically not be used by peers, e.g. experimental evaluation.
- 3 A look at standard textbooks shows that peer evaluation is not included or reflected upon separately from other evaluation approaches, see e.g. Stufflebeam and Shinkfield (2007), McDavid and Hawthorn (2006), Davidson (2004), Owen and Rogers (1999), Rossi et al. (1999).
- 4 This differentiation goes back to Buchanan (1987) and his concept of constitutional economics, which focuses on 'choice among constraints' respectively 'among rules', whereas neoclassical economics is on the 'choice within constraints'.
- 5 This does not mean that peer evaluation is regarded as superior to other forms of evaluation.
- 6 For the use of the programme evaluation standards in the framework of the VET sector in Europe, see Beywl and Speer 2004; the other three standards sets are already closely related to the education field so that all four standard sets seem to be appropriate for use and thus provide the basis for the underlying reflections in this context.

- 7 The German Evaluation Standards are very close to the JC Standards (1994). In Germany, there is a long tradition of self-evaluation, especially in the fields of education and social services. The Evaluation Society (DeGEval, 2004) published these recommendations for the conditions of self-evaluation.

## References

- Beywl W, Speer S (2004) Developing standards to evaluate vocational education and training programmes. In: Descy P, Tessaring M (eds) *The Foundations of Evaluation and Impact Research*. Luxembourg: CEDEFOP, Reference Series 58, 51–129.
- Buchanan JM (1987) Constitutional economics. In *The New Palgrave: – A Dictionary of Economics*. Cambridge: Palgrave Macmillan, 585–8.
- Cabatoff K (2001) The long march from evaluation to accreditation: Quebec’s new ‘government management framework’. *Canadian Journal of Program Evaluation* 16 (Special Issue): 73–88.
- Campanario JM (1998a) Peer review for journals as it stands today: – part 1’, *Science Communication* 19(3): 181–211.
- Campanario JM (1998b) Peer review for journals as it stands today – part 2. *Science Communication* 19(4): 277–306.
- Cichetti DV (1991) The reliability of peer review for manuscript and grant submissions: a crossdisciplinary investigation. *Behavioral and Brain Sciences* 14(1): 119–35, discussion 135–86.
- Daigneault P-M, Jacob S (2009) Toward accurate measurement of participation. *American Journal of Evaluation* 30(3): 330–48.
- Davidson EJ (2004) *Evaluation Methodology Basics*. London: SAGE.
- Deutsche Gesellschaft für Evaluation (DeGEval) (2002), *Standards für Evaluation*, Cologne.”
- Deutsche Gesellschaft für Evaluation (DeGEval) (2004) *Empfehlungen zur Anwendung der Standards für Evaluation im Handlungsfeld der Selbstevaluation*. Alfter: DeGEval.
- Di Battista G, Palomba L, and Vergani A (2009) Evaluating the quality of Italian local vocational training systems. *Evaluation* 15(2): 185–203.
- Ehren MCM, Leeuw FL, and Scheerens J (2005) On the impact of the Dutch Educational Supervision Act. *American Journal of Evaluation* 26(1): 60–75.
- ENQA-VET (2009) *Policy Brief: ENQA-VET Peer Learning Activities*. URL (consulted Aug. 2009): <http://www.enqavet.eu/documents/PolicyBriefPLA000.pdf>.
- European Commission (2008) *Recommendation of the European Parliament and of the Council on the Establishment of a European Quality Assurance Reference Framework for Vocational Education and Training*, SEC (2008) 440, SEC (2008), 441. Brussels: EC.
- Frey B (2002) *Publishing as Prostitution? Choosing between one’s own Ideas and Academic Failure*. Zurich: Institute for Empirical Research in Economics, University of Zurich, Working Paper, 117.
- Gans JS, Shepherd GB (1994) How are the mighty fallen: rejected classic articles by leading economists. *Journal of Economic Perspectives* 8(1): 165–79.
- Gutknecht-Gmeiner M (2007) Peer Review als Verfahren zur externen Evaluierung von Schulen. *Zeitschrift für Evaluation* 6(1): 31–60.
- Joint Committee on Standards for Educational Evaluation (1994) *The Program Evaluation Standards: How to Assess Evaluations of Educational Programs*. Thousand Oaks, CA: SAGE.
- Joint Committee on Standards for Educational Evaluation (2003) *Student Evaluation Standards: How to Improve Evaluations of Students*. Thousand Oaks, CA: Corwin Press.
- Joint Committee on Standards for Educational Evaluation (2008) *The Joint Committee Personnel Evaluation Standards*, Thousand Oaks, CA: SAGE.

- Kornhuber HH (1988) Mehr Forschungseffizienz durch objektivere Beurteilung von Forschungsleistungen. In Daniel H-D, Fisch R (eds) *Evaluation von Forschung: Methoden – Ergebnisse – Stellungnahmen*. Konstanz: Universitätsverlag, 361–82.
- Kyriakides L, Campbell, RJ (2004) School self-evaluation and school improvement: a critique of values and procedures. *Studies in Educational Evaluation* 30(1): 23–36.
- Leeuw FL (2002) Reciprocity and educational evaluations by European inspectorates: assumptions and reality checks. *Quality in Higher Education* 8(2): 137–49.
- Leeuw FL, Furubo J-E (2008) Evaluation systems: what are they and why study them? *Evaluation* 14(2): 157–69.
- Lehtonen M (2005) OECD Environmental Performance Review Programme: accountability (f)or learning? *Evaluation* 11(2): 169–88.
- McDavid JC, Hawthorn LRL (2006) *Program Evaluation and Performance Measurement*. Thousand Oaks, CA: SAGE.
- McNamara G, O'Hara J (2004) Trusting the teacher: evaluating educational innovation. *Evaluation* 10(4): 463–74.
- Mathison S (2005) *Encyclopedia of Evaluation*. London: SAGE.
- Morabito S (2002) Evaluator roles and strategies for expanding evaluation process influence. *American Journal of Evaluation* 23(3): 321–30.
- Nevo D (2001) School evaluation: internal or external? *Studies in Educational Evaluation* 27(2): 95–106.
- Owen JM, Rogers PJ (1999) *Program Evaluation: – Forms and Approaches*. London: SAGE.
- Pagani R (2002) *Peer Review: A Tool for Co-operation and Change*. Paris: OECD SG/LEG (2002)1.
- Peer Review in Initial Vocational Education and Training (2005) *European Peer Review Manual for Initial VET*. Vienna.
- Peer Review in Social Inclusion (2005) *Final Report: Implementation of the Peer Review Programme 2005: Main Results of the Peer Reviews 2004 and 2005*. URL (consulted June 2009): [www.peer-review-social-inclusion.eu](http://www.peer-review-social-inclusion.eu).
- Peer Review in Social Inclusion (2009) *Operational Guide*. URL (consulted June 2009): [www.peer-review-social-inclusion.eu](http://www.peer-review-social-inclusion.eu).
- Rist RC, Stame N (eds) (2006) *From Studies to Streams*. New Brunswick, NJ: Transaction Publishers.
- Rossi PH, Freeman HE, and Lipsey MW (1999) *Evaluation: A Systematic Approach*. Thousand Oaks, CA: SAGE.
- Scriven M (1991) *Evaluation Thesaurus*. Newbury Park, CA: SAGE.
- Scriven M (2009) Meta-evaluation revisited. *Journal of MultiDisciplinary Evaluation* 6(11): pp. iii–viii.
- Speer S (2001) Evaluation und Benchmarking: – Ein Methodenvergleich am Beispiel der betrieblichen Personalarbeit. In: Geise W (ed.) *Ökonomische Bildung zur Bewältigung von Lebenssituationen*. Bergisch Gladbach: Verlag Thomas Hobein, 51–67.
- Speer S (2007) *Evaluation Report of the Leonardo da Vinci-Project Peer Review in Initial Vocational Education and Training*. Vienna. URL (consulted June 2009): [http://www.peer-review-education.net/\\_TCgi\\_Images/peerreview/20071218155137\\_Evaluation%20Report%20Peer%20Review.pdf](http://www.peer-review-education.net/_TCgi_Images/peerreview/20071218155137_Evaluation%20Report%20Peer%20Review.pdf).
- Starbuck WH (2003) Turning lemons into lemonade: where is the value in peer reviews? *Journal of Management Inquiry* 12(4): 344–51.
- Starbuck WH (2005) How much better are the most-prestigious journals? The statistics of academic publication. *Organization Science* 16(2): 180–200.
- Stufflebeam DL, Shinkfield AJ (2007) *Evaluation Theory, Models, and Applications*. San Francisco, CA: Jossey-Bass.
- Vanhoof J, van Petegem P, and de Maeyer S (2009) Attitudes towards school self-evaluation. *Studies in Educational Evaluation* 35(1): 21–8.

Vedung E (1999) *Evaluation im öffentlichen Sektor*. Vienna: Böhlau.

Weiss CH (1998) *Evaluation*. Upper Saddle River, NJ: Prentice Hall.

Sandra Speer is an independent evaluator, based in Wiesbaden, Germany; the focus of her work is on national and international evaluation research for public agencies as well as evaluations in various fields; Please send correspondence to: Untere Matthias-Claudius-Str. 11, 65185 Wiesbaden, Germany, [email: [sandra\\_speer@t-online.de](mailto:sandra_speer@t-online.de)]