

Evaluation

<http://evi.sagepub.com>

Does Evaluation Build or Destroy Trust?: Insights from Case Studies on Evaluation in Higher Education Reform

Christine Schwarz and Gerlinde Struhkamp

Evaluation 2007; 13; 323

DOI: 10.1177/1356389007078625

The online version of this article can be found at:
<http://evi.sagepub.com/cgi/content/abstract/13/3/323>

Published by:

 SAGE Publications

<http://www.sagepublications.com>

On behalf of:



The Tavistock Institute

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

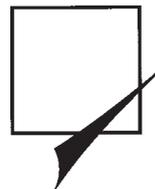
Email Alerts: <http://evi.sagepub.com/cgi/alerts>

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

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

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

Citations (this article cites 8 articles hosted on the SAGE Journals Online and HighWire Press platforms):
<http://evi.sagepub.com/cgi/content/refs/13/3/323>



Does Evaluation Build or Destroy Trust?

Insights from Case Studies on Evaluation in Higher Education Reform

CHRISTINE SCHWARZ
Independent Consultant, Germany

GERLINDE STRUHKAMP
Independent Evaluator, Germany

Trust isn't sustained easily. It can be destroyed in less than a second.
(Anonymous)

This article reflects on the relationship between evaluation and trust. Evaluation has become prominent in recent years as a way to control administrative action. The underlying assumption here is that administrative action, e.g. in the form of publicly funded programmes, requires control in order to gain or sustain its legitimacy. Two case studies of programmes of higher education reform in Germany, initiating so-called 'virtual universities', are used to investigate how stakeholders experienced evaluation. Informants argued that there was too much evaluation, that confusion and competition arose about the roles of evaluation, and that little instrumental use occurred. This situation caused frustration and begs the question whether evaluation, intended to increase trust through systematic and transparent inquiry and rational decision-making, contradicted its own claims. It is argued that evaluation can be endangered by 'inverse process use' or 'process damage'.

KEYWORDS: bureaucratic expansion; higher education reform; process use; qualitative research; trust

Introduction: Are we Living in an Audit Society?

As evaluation becomes more widespread, increasingly there are concerns raised among social scientists about its consequences. For example, evaluation as a 'growth industry' (Leeuw, 2002) was already identified as an issue in the 1970s, when the

first 'Cambridge Conference' was held. This was motivated by a 'rapid increase of evaluation activities', which it was thought would need an 'equivalent surge of new thinking' about evaluation (MacDonald and Parlett, 1973).

The relationship between evaluation and trust has received little attention in the evaluation literature, which is why it is introduced here for further discussion.

Generally, trust is a mechanism to reduce complexity. Since human beings are limited in their capacity to gather and process information, trust is used to come to decisions despite incomplete information. In a sociological sense, such as the explanation in system theory (Luhmann, 1968), trust reduces social complexity in the way that it relies on and 'enlarges' existing information and substitutes missing information with an internally assumed security. Besides this 'inner' mechanism of relying on trust, an individual also builds on other institutions developed in parallel, like law or an organization. 'Trust is not the only fundament of the world; but a complex yet structured society cannot be constituted without such trust' (Luhmann, 1968: 106; translation by authors).

Trust is essential in enabling social interaction, and hence societal functioning in general. Trusting a statement, person, assumption, institution or a set of rules allows a person to minimize risks; part of the risk is substituted by trust. However, at the same time trusting somebody or something *in itself* carries the risk of the trust being disappointed or broken. It is a constant, circular act of balance between offering trust, demanding and questioning it, and building and destroying it between individual members of society, as well as between individuals and institutions. Trust and mistrust are two sides of the same coin. A society always consists of (mis)trusting relationships within the realm of the ambiguity described.

In modern societies trust has its place between, among others, citizens and the political class and its administrative arm, the bureaucracy. Like other societal institutions, too, administrative action seeks to establish or sustain the trust that is invested in its actions. Max Weber (1980) focused on the (inter)action of bureaucracy in modern societies and how it has to seek trust in the form of legitimation for its activity on behalf of the voters. However, in his analysis about the side effects of democracy and bureaucracy he pointed to what he perceived as a paradox. On the one hand, government and its administrative arm requires control of its bureaucratic action in order to establish or sustain trust. On the other hand, bureaucratic control can grow to the extent that Weber warned against a dynamic of control: even measures of control might have to be controlled. So potentially control follows control in an endless cycle. In such cases the very trust that bureaucratic control attempts to create may, in turn, fail its aim in the long run. The historian of science Theo Porter (1995) labelled such dynamics 'trust in numbers'.

More recently social-anthropologist Marilyn Strathern (2000) and sociologist Michael Power (1994, 1997, 2003) have critically reflected on increasing evaluation and auditing activities and how this may affect societal relationships of trust. They characterized auditing and evaluation as non-personal 'regimes' of bureaucratic control and 'audit cultures' (Strathern, 2000). In such 'audit cultures', accountability practices have become so pervasive that observation has become more important than acting. Are we living in an 'audit society' (Power, 1997)?

The general idea about the potential expansion and dominance of measures of control from the beginning of the 20th century still applies today and with it the question of whether these measures of control actually do achieve their aim in creating trust or whether they enhance mistrust. Of course, over the course of a century the jargon, as well as society, has changed. Weber's terminology of bureaucratic control nowadays may be thought of as 'accountability' or 'government performance', as seen, for example, in the US Government Performance and Results Act (GPRA) of 1993. In the USA, as well as in European states, under the broad idea of transformed governance called New Public Management, evaluation often plays a crucial role as an instrument to exert and maintain control of bureaucratic action.

Power's and Strathern's reflections motivated the empirical study presented in this article. Their critical analyses created the starting point for an investigation of how evaluation can be understood as an instrument affecting trust. Two programmes of German higher education reform where multiple evaluations had been built in were the subject of the investigation. Based on empirical observation of these two programmes, this article attempts to introduce the notion of trust to the evaluation literature. The investigators also hope to contribute to an understanding about how mechanisms of trust (and mistrust) are linked to evaluation. The analysis of the two cases *exemplifies* a train of thought, and readers are invited to connect the reflections presented here to their own experiences.

The Cases

Commonly, administrative interventions come in the form of 'programmes' aiming to address social problems, or to initiate or facilitate the development of innovative ideas. Accordingly, in the late 1990s, the German Federal Ministry of Education and Research and almost all of the 16 German state (*bundesland*) ministries launched a number of programmes to promote the international competitiveness of German universities. One of the core strategies (largely paralleling the 'dot-com'/'new economy' boom) was promoting the use of new information and communication technologies in higher education – the vision of Online Universities or Virtual Universities. Various evaluations were outlined in the programmes' descriptions that accompany the programmes' development.

Several activities had been considered broadly as evaluation, for example, 'accompanying research' carried out by an internal evaluation team, a 'midterm evaluation' by a scientific board, the 'self-evaluation'¹ of the programme members (valuees²) or 'consultancy' by a competence centre. All of these evaluative activities are summarized under the term 'evaluation' in this analysis and hence this is also the general term used in the discussion that follows. Evaluation is broadly considered an institutionalized strategy for building trust in bureaucratic societies.

Case Descriptions

Both virtual university programmes considered in the case studies took place over a five-year period. One was located in northern and one in southern Germany.

Each programme was administered by the respective state's educational ministry with their funds and funds from the federal ministry. The two programmes were chosen out of approximately ten of the largest virtual university programmes that were set up during the late 1990s. The two that were selected were the most advanced and comparable to each other.

The programme description of one of the case studies included an internal evaluation team, intended to carry out a formative evaluation for the didactical improvement of the virtual courses offered and for 'quality assurance' purposes. When the programme eventually started, nine staff members were recruited from four different social science institutes that had applied for the programme based on the terms of reference. This team brought together knowledge of more than four disciplines (among them pedagogics, psychology, communication science and economics), their representatives all having different conceptions of how to do an evaluation. The evaluation focused on the specific virtual courses and courses with e-learning elements. The team gave feedback to the evaluatees regularly, but a substantial portion of their time was also committed to writing up evaluation reports for the competence centre and the ministry and presenting the results at conferences to find a forum for their results. Throughout the second half of the programme especially, the conflicts that had existed from the beginning grew stronger and resulted in high staff turnover. Only four of the initial team stayed focused on the formative task, but they concentrated their efforts in working with just one programme partner (one faculty member and his assistant), and publicized their evaluation results in research-oriented forums.

A second, initially central, unit was the competence centre. Its responsibility was to synthesize reports of experiences of virtual learning environments and communicate this to external audiences. Three pedagogues formed the team. They had to coordinate the evaluations and self-evaluations, were consultants to internal (e.g. groups that wanted to plan a project and write a proposal) and external (e.g. external projects that requested consultancy) audiences and were also responsible for PR work. They positioned themselves close to the evaluatees and the academic environment. However, they lacked recognition from the scientific board, which recommended closing down the competence centre at midterm, regrettably in the eyes of many other stakeholders. Towards the end of the programme, though, a new centre was founded, called 'Centre for Knowledge Management'.

A scientific board of four members was appointed during the programme's start-up phase. Their obligation was to hand in annual 'status reports' to the commissioners (the federal and state ministries that were responsible). After three years the board had to decide about the programme's continuation based on three site visits, which were not specifically structured, and framed only by the board's chairperson. For these visits, they requested documents about the programme upfront, the e-learning environment was explained to them, and they could ask questions and test the e-learning arrangements 'live'. These evaluations were meant to inform the commissioners about the 'sustainability' of the programme investments and activities.

To summarize the involvement and investment of evaluative activities in these programmes: the first virtual university programme was evaluated in some way

by the following parties (cf. Figure 1): (1) the programme staff themselves ('self-evaluation'), (2) an external third-party project (research study of an assistant professor, 'evaluation research'), (3) the 'competence centre' (later: Centre for Knowledge Management), (4) experts from the scientific board, as well as (5) the 'Centre for Higher Education System Development'.

Similarly, the second programme was evaluated by (1) an evaluation department with four different subdepartments, and (2) a visiting researcher, whose original aim was to write her dissertation but who ended up being entrusted with the evaluation. These parties, initially recognized as internal evaluations, were supplemented with external evaluations, namely (3) another 'competence centre', (4) a scientific board of four external experts and (5) again the above mentioned 'Centre for Higher Education System Development' – with different staff than the ones that were concerned with the first programme.

Methodology and Questions Asked

Two programmes aiming to introduce information and communication technologies to higher education in Germany were analysed using ethnographic research techniques. These two programmes dealt with the 'Virtual University'. In reconstructing the evaluation processes within these two programmes, 35 narrative interviews with

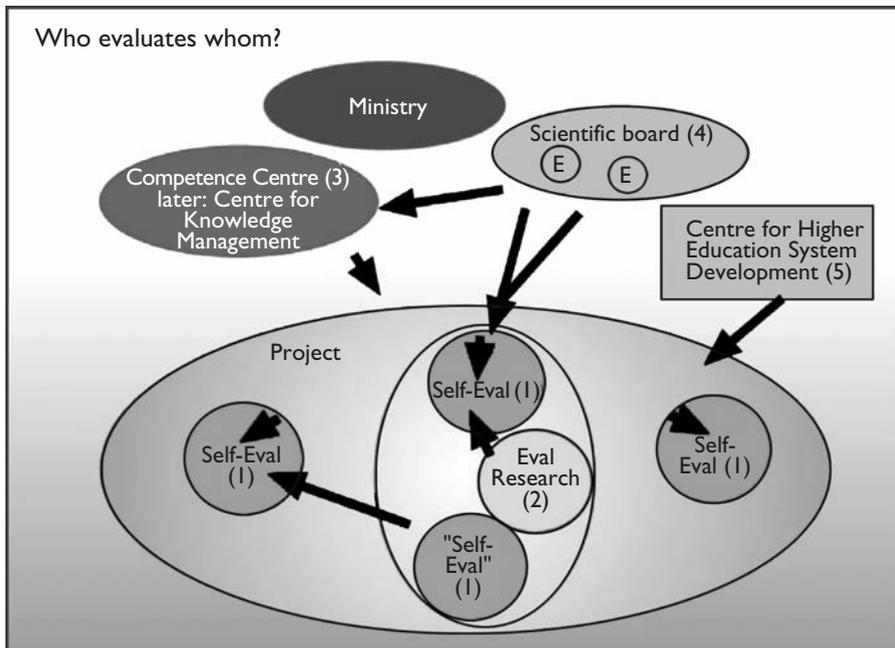


Figure 1. Parties Engaged in Evaluative Activities within the First Virtual University Programme

various stakeholders involved in the programmes were conducted in 2001 and 2002. These were with:

- members of the German federal and state ministries;
- employees from the so-called ‘competence centres’ and other units like the ‘centre for higher education system development’ and a public evaluation agency guiding the programmes’ implementation, as well as members from steering committees like scientific board members – all these taking some kind of ‘evaluator’ role;
- evaluatees – in this case, assistant and full professors of the universities who took part in these programmes and whose task was to ‘provide content’ for the e-learning modules to be developed; and
- other programme staff members, such as research assistants and guest researchers.

In addition to the interviews, documents were analysed, visualization techniques were used when working with the interview partners, participant observation was employed and a research diary was kept.

The central research question that guided the interviews, observations and analyses was how people think about and reflect upon the growing evaluation activities. The focus of the interviews was on the following specific questions: ‘How does evaluation take place in your programme?’, ‘What is the function of evaluation in the programme?’ and ‘How is evaluation used by the stakeholders?’

The examples in this article are taken from a larger study;³ in the following discussion only selected results are shared. Even though comparisons between the two cases are also worthwhile, in this article the insights from both case studies are summarized to present consistent patterns that occurred in both programmes. Therefore, the two programmes described earlier will be presented as one unit of analysis, so that the two cases will be dealt with as one programme in the explication that follows.

The discussion first considers the quantitative expansion of evaluation; second, people’s perceptions of evaluation’s role(s) in this development; and third, the tension between expected and actual use of evaluation results and forms of use regarding the evaluation process. Finally, mechanisms of trust through evaluation are considered.

Stakeholders’ Experience of Evaluation: Simply Too Much

The multiple and many-layered evaluation activities that occurred in these programmes in some cases exceeded the programme activities. For example, one person was employed to evaluate only one seminar, and four people were responsible for developing the design of the evaluation to evaluate a seminar, run by one colleague. More human resources were invested to monitor outcomes than were used to produce the outputs and outcomes. For example, each e-learning programme was evaluated by at least six different institutions or teams specifically formed for evaluative purposes. One interviewee stated:

From now on everything will be evaluated! Even though some people in Germany maybe didn’t yet get it! (Staff member of public evaluation agency)

It is worth noting that the employees involved in the evaluations in each of the two programmes were not informed clearly about the existence of the other evaluations taking place within the same programme – no one was aware of this totality of all the evaluations. Some became aware of each other more or less by chance. Ultimately, it was the case study that revealed that this many evaluations were taking place of both programmes. Interviewees realized that they had become aware of the multiple layers of the evaluation only through the researcher and it caused irritation among them that they had had no idea, let alone been informed, about the multiple layers of evaluation.

It makes me cry to think about what's all been evaluated here for so much money!
(Programme staff member)

In view of the perceived overproduction of evaluations, we tend to agree with Power's diagnosis of an 'audit society' (1994, 1997) in which citizens observe more than they act. But as the following reveals, the evaluations were not mere observation as an end in itself, but rather there were overt and covert negotiation processes that led to the 'licence to evaluate'.

The 'Licence to Evaluate': Role Confusion and Role Conflicts

It was observed that the two case studies appeared to incorporate a dynamic process of evaluation: what it is, ought to be, or turns out to be. Three questions help to highlight how people thought and acted concerning the role evaluation played in those programmes:

- Who is allowed to call one's actions 'evaluation'?
- What are the purposes of that evaluation?
- How do evaluators become more or less powerful among each other and with other parties (evaluatees, client/sponsor)?

For example, among the variety of evaluations of these programmes, members of the scientific board that had been appointed by the ministry became 'midterm evaluators' during the programme. However, they gained this function only after the programme had, in the view of the ministry, developed somewhat problematically. In reaction to this development, those who were originally authorized as evaluators in accordance with the programme's description began to oppose the scientific board. Others, like actors from the 'competence centre', who considered themselves as 'consultants' and explicitly not as 'evaluators', came to be perceived as 'evaluators' by others. It was the official task of the competence centre to provide consultancy and coaching for programmes, and to assist especially in training the programmes' staff and helping with public relations work. Nevertheless, for most of the programme staff this was seen as a kind of evaluation, although the competence centre staff refused repeatedly to be called or treated as 'evaluators' or controllers of any kind, since their intention was to support and not to judge the programme.

It was very important to us to clearly tell the programme staff from the very beginning: 'We are not the evaluation, but your consultancy!' It was our aim at the foundation of the competence centre to assist and support the programmes (it's difficult enough for them already) in a cooperative way. And then suddenly the ministry wants us to turn

into controllers in the midst of it! That's not going to work, you can't do it that way.
(Staff member of competence centre)

It appeared that who was perceived – i.e. respected or feared – as evaluator, and what rights and duties went along with the (self-)appointment as 'evaluator' was socially constructed. Stakeholders found it difficult to grasp who trusts whom and who is trustworthy to whom. Rather than roles of evaluation and evaluators having been declared and made clear beforehand, it was the actors' perceptions that made somebody an evaluator or someone's action an evaluation. What was considered 'evaluation' depended on the rules and contingencies of (group) dynamics and had evolved and changed over time. Who actually was allowed to call herself/himself an 'evaluator' and who claimed to do an 'evaluation' appeared more or less arbitrary to the stakeholders.

In this context, there was also confusion concerning the evaluand's objectives:

When I met one of our ministry officials responsible for the programme for the first time at a programme meeting, I asked her: 'Good to meet you: Can you tell me what is the main target of the e-learning programme?' she was shocked by my question and I was shocked by her answer. She replied to me: 'Well, please, you are the evaluator. I thought I'm here today at this meeting to find that out from you!' (Staff member of competence centre)

This situation occurred at a time of transformation with a heightened level of uncertainty about what to do and how to do it well. In a similar vein, evaluation also seemed to play a diffuse role of slowing down decision processes. In the words of a programme staff member:

If nobody knows what to do, they start evaluating!

Additionally, in both programmes the – mostly incompatible – evaluation roles underwent ongoing changes: variously evaluation was undertaken as a formalized way of self-reflection, a style of research or enlightenment, consultancy or moderation, a process of finding consensus about programme objectives, a chance to settle (old) quarrels or simply as an end in itself. This made perfect terrain for role conflicts.

For example, a team of young social scientists were in general not accepted as 'real' evaluators – although the team was convinced by their proper data collection methods that they were qualified to be evaluators. In contrast, members of the scientific board, who appeared only seldom and when they did it appeared to be a special and costly event, were perceived by interviewees to have formed their judgement by generalizing their personal impressions of the programme based on sporadic conversations with other external experts instead of using thoroughly the data available. Everybody from the internal evaluation team was annoyed that there was 'no interaction' (research assistant of internal evaluation team) between the scientific board and the internal evaluation. Yet, informally, presumably through the competence centre, members of the scientific board did access data produced by the internal evaluators:

Additional to their site visits and hearings, the scientific board obviously got hold of our programme reports that were actually meant for the programmes themselves. It was too obvious that they copied from that, but results were distorted and interpretations

Schwarz and Struhkamp: Does Evaluation Build or Destroy Trust?

only partly backed up. Apart from that, the board judged completely a-factually [without facts]. And towards midterm they even produced different evaluation criteria! But none of the board members actually read all the reports. They were without a plan but wanted to be part of the decision! (Staff member of internal evaluation team)

After the quantity of evaluation information and its unclear role(s) were explored, the interviewees' perception of the relevance and actual use of evaluation results was considered so that their impressions regarding the evaluation process could be presented. The role confusion extended to different perceptions and expectations towards the use(s) of evaluations.

Not Use but 'Uses'

Most interviewees reacted with perplexity to the interview question about the use of the evaluation activities. The majority of the interviewees voiced the opinion that evaluation would require more time for its insights and recommendations to be transformed into productive actions. For some interview partners, however, it remained a mystery how the scientific board – despite the available detailed data from internal evaluators – came to their cursory and misleadingly generalized judgement about the programme, and how their recommendations were then actioned 'one-to-one':

Well, for our decision, midterm of the 5-year period, to continue the programme's funding, we put the recommendations of the scientific board into practice one-to-one. What else could we have done? When we asked whether the internal evaluation had been considered, he [a scientific board member] answered in the negative and said those had been intended more internally for didactical improvements. (Ministry official)

Some interviewees argued that as time went by and knowing the ministry was behind them, the scientific board members behaved more and more like over-elaborated judges, bosses or even patrons that were allowed to censure the programme staff in public like their wards:

Could you please be so kind and explain to me why I should invest another one million Euro into you and your programme?

using these words a scientific board member interrogated a programme staff member in front of approximately 50 of his colleagues. In the view of the evaluatees and internal evaluators they behaved like a powerful, detested tribunal:

The scientific board came, saw, and conquered. (Research assistant of internal evaluation group)

The administrators from the ministry, meanwhile, enthroned the scientific board and took their advice:

The scientific board fulfilled its evaluation function to 100%. (Ministry official)

When evaluators had tried to present themselves as responsive, dialogical consultants and advisers in the service of the programmes' staff, their pledge of being advisers had led staff to hope for less hierarchical management structures and for shared decision-making. But in comparison with the investment

Evaluation 13(3)

in evaluation of the two programmes, its influence over practical matters was questioned:

In the sense of 'constructive help' the scientific board did, in my view, not contribute to the programme at all. Instead of providing advice they [the scientific board members] behaved like a scientific testing group that only dictates guidelines, reformulates objectives. But when you want to evaluate, you first have to be willing to understand things! . . . Well, I don't have any need for such an evaluation. When the ministry thinks it wants to set up something like that – please do. But I don't need anything like that. (Head of internal evaluation team)

Apart from smaller discoveries and helpful tips, many of the stakeholders – mostly programme staff – saw the 'use' of the many evaluation activities mainly as legitimizing the political administration:

Evaluation is simply duty and must be set in place, especially to legitimate larger public expenditures. (Guest researcher in evaluated programme)

Evaluation is just for the marketing of the ministry, so that they know where to pump in the money. (Programme staff member)

In hindsight and from a different perspective, the lack of utilization of the evaluation was attributed to the evaluation proceeding too slowly:

Every scientific analysis is always too slow for the political decision and rarely does it contain automatically a 'plot to act'. So with evaluation results: they almost always run behind the political processes. (Scientific board member)

In the ministry the fast and scientifically questionable procedure of judgements made by the scientific board members was defended as efficiency:

Evaluation is just always too late for political decisions. . . . The reproach to have acted too fast, for example, in selecting board members or when starting new programmes, I accept happily. That's the trade-off between republican efficiency and democratic legitimation. (Ministry official)

Political needs seemed to be in the forefront when a ministry official – feeling discouraged by mostly negatively critical evaluation results – pointed out:

To be honest, I don't care how well you could prove the failure of a programme with an exact evaluation. That happens all the time! What's important is to show successes – regardless of how well that can be proven by an evaluation. (Ministry official)

Considering the need for timeliness of evaluative information within political processes, administration officials seemed to actually benefit from a slow, somewhat late running, less powerful and less directly tied-to-use evaluation, because for them it opened up more possibilities and options for actually using the evaluation results – that is, using them at their own convenience and at their own choosing. A version of 'republican efficiency', as an administration official called it. But what is evaluation then? A silver bullet in political day-trade to administrative decision-makers, while ignoring what the evaluation process does to the other stakeholders?

Evaluation Affecting Trust

This recapitulation of the two case studies reflected on the ways and mechanisms where trust possibly got lost 'along the way' during the evaluation processes described. Even though one might think that these were just two unfortunate examples of poor evaluation practice, the insights from them might – as suggested here – apply more broadly to the field and stimulate further thinking among evaluators about evaluations and their consequences.

As evaluation becomes an increasingly used instrument – as suggested by these two cases and it could also be argued for Germany more broadly (e.g. Beywl, 1999) – this is part of a larger audit culture, as Power (1994, 1997) stated, for our current situation in (western) countries, especially the UK. The problem is not checking per se, but 'checking gone wild' (Power, 1997: 14). Taking a pessimistic stance and considering the relative chaos of evaluative procedures in the two case examples, this may be not too far from accurate. Do not take this the wrong way – this position is not meant to question checking, evaluation, auditing and similar activities *as such*. All of these activities are legitimate and necessary procedures of bureaucratic control in modern societies; in fact, there cannot be a society without these and other mechanisms for building trust. However, we do want to emphasize 'reflexivity' (Power, 1997: 145) in how evaluation is linked to mechanisms of building and destroying trust on the verge of a rising profession – so that it is not 'simply duty' (as one of the interviewees had suggested).

We have seen that quantity does not equal quality – but rather the contrary. Programme administrators requested evaluation as integral parts of their programmes in the hope of understanding the effects of their administrative actions. In this case, a true construction process started which, when viewed as a whole, appeared to be quite unproductive. Guidance and coordination or more negotiation towards a clear(er) purpose would have been helpful. The result of this vacuum was that evaluation, supposedly an instrument to help with and guide decisions, was (mis)used to defer responsibilities. Evaluation achieved pretty much the opposite of what it generally claims to achieve – transparency and systematic inquiry and rational decision-making. That politicians and administrators at least at times seem to hope for some miraculous effects from evaluation could be seen in the two case examples presented here – multiple 'evaluations' were written into the programme descriptions without further specifications. It seemed there was some hope that evaluation 'will do', will take care of whatever is necessary: to hold programmes accountable, to uncover and cure weaknesses, to develop innovations, to moderate and mediate, to legitimate and to provide evidence. Due to the lack of specification which function(s) evaluation was to have played in these reform programmes, a vacuum emerged, inevitably leading to role conflicts between individual evaluators and among evaluation teams.

Different roles for evaluators are not a problem per se; in fact, some authors perceive the variety of roles an evaluator can take on as enriching for evaluation practice (Brown, 1995; Morabito, 2002). The linchpin for simultaneous roles, though, is a set of clearly defined purposes for each evaluation taking place. Otherwise, confusion and (micro)political struggles result, as in the case studies presented here.

Evaluation should be conducted in ways that are more careful with the stakeholders' trust. If uncontrolled, unplanned or unreflected evaluation is used as a potentially flexible instrument of bureaucratic control and governance to grant and withdraw trust, it involves a certain danger of destroying trust while being (or claiming to be) a possible trust-building tool.

The case studies demonstrated that the vacuum led to (micro)politically driven confusion about evaluation, its purposes, potential uses (and ignorance of uses) and who is responsible for each. However, leaving actions to (micro)political struggles makes evaluation no better than other means of deciding about or 'arranging' an issue. Evaluation will then develop the sour aftertaste of being like club law – a weak currency for legitimation in our political culture that is led, at least officially, by the postulate of rational decision-making.

The question of evaluation utilization is the third aspect we highlight in this article. Patton (1997) advanced the debate on utilization by distinguishing between the notion of use of evaluation results and use of the evaluation process itself, called 'process use'.

Marvin Alkin (1990: 293) provided a taxonomy of forms of use and opposite variations of it (non-use, misuse, abuse, etc.). In this case, the 'one-to-one' use of those cursory, scarcely data-based evaluations by the scientific board through administration officials might be classified as one instance of 'misuse' (i.e. use by an informed user of a poorly undertaken evaluation). However, perceptions of use differed among stakeholders within the same evaluation process and were subject to changes over time. How could evaluations deal appropriately with such plurality of expectations towards use? Whereas the massive presence of evaluations taking place gave rise to programme staff members hoping for (more) instrumental use of evaluation results, they became frustrated with what they perceived as little to no immediate utilization. A number of programme staff members thought that their evaluation work served merely to legitimate the ministry's actions, which is a legitimate form of evaluation use. From the perspective of most of the programme staff though, it was this 'legitimation-only use' that undermined their trust in the instrument of evaluation, even though it did not appear from the case studies that the commissioners of the evaluation had planned upfront for a 'symbolic purpose' (according to Alkin's taxonomy also a form of misuse).

But there are more dimensions to use than immediate and direct (instrumental) use. Patton (1997) suggested the conceptual use of evaluations, and other authors have called for broader conceptions like 'evaluation influence' (Henry and Mark, 2003; Kirkhart, 2000) or 'enlightenment' (Weiss, 1987).

Even though administration officials often are considered 'intended users', when interviewed, the representatives from the ministries involved actually did not show much interest in 'intended uses' of a specific evaluation – other than 'proven successes' (ministry official). Rather, they may have intended not to have intentions regarding use, or at least not to express them. For them, it seemed more valuable to remain flexible ('republican efficiency') in using whatever evaluation results – or none of them – appeared to be most appropriate for the situation at hand.

Peter Dahler-Larsen and Jonathan Breul (2004) pointed to a related aspect based on a different argument. In their paper presented during the EES conference

they spoke of an 'arena' to which the political world belongs and in which instrumental use of evaluation becomes less likely, if not impossible, for an evaluator to induce. The decreasing influence of evaluators with regard to instrumental use, in their view, is associated with lower 'degrees of predictability' that accompany evaluation arenas in which some stakeholders are not in personal contact (any more).

Karin Haubrich and Christian Lüders (2004) approached the utilization postulate from a normative perspective: according to their view it poses the question 'who has the legitimate right to decide what actually is useful to whom and when?'. This also became apparent in the case studies considered where role confusion and role conflicts led to (micro)political struggles and ultimately prioritized the form of use the ministry preferred – legitimation for bureaucratic action, in the eyes of many interviewees. Haubrich and Lüders (2001) cautioned against such bias towards certain stakeholders. According to them, focusing on certain users, mostly those declared as 'intended users', may actually increase conflicts.

Empirical studies have revealed the difficulty of controlling the use of research results (for Germany, cf. Beck and Bonß, 1989; Lüders, 1993; for the US, cf. Weiss, 1980; Weiss and Bucuvalas, 1980). These studies emphasized that scientific knowledge is used independently, detached from the research context and from the researcher's intentions and that the researcher cannot effectively steer when and how and by whom her knowledge is or will be used. It is difficult to assume that knowledge from evaluation studies would be tied any closer to the evaluator's intentions how, when and by whom it ought to be used (Haubrich and Lüders, 2004). Dahler-Larsen and Breul's (2004) notion of (lack of) predictability of use corresponds with this argument.

Another aspect of use is associated with the notion of process use being viewed as an *exclusively* positive notion; this is challenged by the results of the case studies. In the light of the notion of trust, evaluators should be aware that evaluation processes may have their down-sides, too. The very trust that evaluation attempts to build or sustain may be at risk in the case of evaluation processes that leave behind an evaluation 'hangover'.⁴

Patton considers as process use 'ways in which being engaged in the processes of evaluation can be useful quite apart from the findings that may emerge from those processes' (1997: 88). Patton quotes a programme director who described how the staff's attitude had changed, how they paid more attention to outcomes and were more reflexive and more focused. Patton states that 'any evaluation can, and often does, have these kinds of effects'.

In contrast, in these case studies the evaluation process led to what one might call 'inverse process use' – or 'process damage' if you will – because the evaluation setting in these two programmes left behind much frustration, mostly on the side of the evaluatees.⁵ We wonder if evaluators could serve their own profession better – or even whether they have a professional obligation to do so – by 'talking straight' about the chances for (instrumental) use on the one hand, as well as the risk of misuse, non-use and other potentially frustrating variations. The utilization postulate being so widespread in the evaluation field may divert the views of evaluators, clients and evaluatees from the risks that evaluation *also* bears.

Finally, returning to the title question – does evaluation build or destroy trust? – we suggest that it does both. In the two cases of virtual university programmes, evaluation did not seem to help building trust as an instrument of bureaucratic control. Rather, ambiguity replaced transparency, confusion replaced systematicness and suspicion replaced legitimation.

Alternatively, Ernest House (2004) provided insight into a different example in his keynote address at the EES conference: ‘The passions on both sides were highly inflamed. . . . For years the school district and the plaintiffs have displayed a complete lack of trust in each other. In conducting a monitoring evaluation, I tried to reduce the distrust . . .’ What began as a most fiercely debated and contested evaluation setting evolved over time into what may not be described as perfect, but a *more* trustworthy environment – trust evolved through the conduct of the evaluation.

So, evaluation does have the potential to build or sustain trust. It seems important, in our view, to bring to the evaluation community’s attention how evaluation as an instrument – new or revived – potentially affects trust. This is one step towards the reflexivity that Power (1997) called for.

The ‘audit explosion’ that Power (1994, 2003) spoke of seemed to come true in these German higher education reform programmes. The various evaluation processes being only poorly linked may have been partly due to this expansion, but it was also due to a lack of specification of evaluation purposes and roles. Within this terrain of role confusion and role conflicts, not one use but many ‘uses’ emerged, changed over time and clashed, due to different perceptions of use by the stakeholders who were operating at cross-purposes.

Both the cases presented, as well as House’s case, underline the importance of the process of an evaluation – one could think: ‘It’s the process, stupid!’ Patton (1997) brought this to our attention years ago, but focused only on the positive side (process *use*) and in an ‘arena’ (Dahler-Larsen and Breul, 2004) in which an evaluator is mostly addressing intended users. However, not every evaluation setting offers such possibilities; there are also situations bearing risks in which the evaluation process may harm more than help. This suggests that taxonomies or frameworks of the use of evaluation findings need to be further clarified, especially with regard to the evaluation process (cf. Kirkhart, 2000).

From a sociological perspective, trust is part of society’s fabric and evaluation is part of the threads of trust. Therefore, evaluators should pay attention to how evaluation processes (best) build trust, or conversely prevent the erosion of trust. This article has highlighted this connection and the involvement of evaluators in the societal network of trust, because trust is not sustained easily, but can be destroyed in less than a second.

Notes

We wish to thank Stuart Astill and Gareth Patterson, UK National Accounting Office (VFM Development Team), and Marilyn Grantham, University of Minnesota Extension Service/USA, for their proofreading and helpful comments as well as the three anonymous reviewers for their thoughtful feedback on the earlier version of this article.

1. In Germany this has become a technical term and established evaluation approach (in German: 'Selbstevaluation').
2. Here, evaluatees means programme staff whose programmes or projects were subject to an evaluation; evaluatee here does not mean an individual person being evaluated as in personnel evaluation. Scriven (1991: 157) defines the term more narrowly. A term for the interpretation used here – staff member of evaluand (evaluated programme) – is missing.
3. The first author of this article wrote her dissertation based on this study, financially supported by the Heinrich Böll Foundation, Berlin, Germany (Schwarz, 2003, 2005, 2006).
4. Kirkhart (2000) did substantially expand the framework of evaluation use in integrating the three dimensions source (process/results), intention (intended/unintended) and time (immediate/end-of-cycle/long-term). She states (p. 13) that 'Taken together, the three dimensions of influence offer a framework within which to examine both the positive and negative impacts of evaluation'. However, within her integrated theory an evaluative dimension (good/bad, etc.) is missing. Such an evaluative dimension would highlight that forms of use (both results- as well as process-based) may have a positive or negative impact – like process use or 'process damage' – on the respective stakeholders.
5. Negative experiences in and through an evaluation have also been described under the phenomenon of evaluation anxiety, and excessive evaluation anxiety (XEA) as its serious form (Donaldson et al., 2002). 'Process damage' as introduced in this article may be considered being related to it. However, evaluation anxiety is a psychological concept whereas in this article trust is seen in a sociological context. Whereas Patton does perceive process use as 'indicated by individual changes in thinking and behavior', for him process use also refers to 'program or organizational changes in procedures and cultures' (1997: 90).

References

- Alkin, M. C., ed. (1990) *Debates on Evaluation*. Newbury Park, CA: SAGE.
- Beck, U. and W. Bonß, eds (1989) *Weder Sozialtechnologie noch Aufklärung? Analysen zur Verwendung sozialwissenschaftlichen Wissens* [Neither Socio-Technology Nor Enlightenment? Analyses of Use of Social Science Knowledge]. Frankfurt/Main: Suhrkamp.
- Beywl, W. (1999) *Zur Weiterentwicklung der Evaluationsmethodologie. Grundlegung, Konzeption und Anwendung eines Modells der responsiven Evaluation* [Developments in Evaluation Methodology: Foundation, Conception and Application of an Approach of Responsive Evaluation]. Berne: Lang; repr. from 1988.
- Brown, P. (1995) 'The Role of the Evaluator in Comprehensive Community Initiatives', *New Approaches to Evaluating Community Initiatives*, vol. 1. Washington, DC: Aspen Institute. Available at: <http://www.aspeninstitute.org/Programmet1.asp?bib=1276&i=83> (site visited 9 Aug. 2004).
- Dahler-Larsen, P. and J. Breul (2004) 'Evaluation in the Public Arena', paper presented at the European Evaluation Society conference, Berlin, Germany, 1 Oct.
- Donaldson, S. I., L. E. Gooler and M. Scriven (2002) 'Strategies for Managing Evaluation Anxiety: Toward a Psychology of Program Evaluation', *American Journal of Evaluation* 23(3): 261–73.
- Haubrich, K. and C. Lüders (2001) "'Nutzungsorientierung" – ein anregendes Konfliktpotenzial' ['Orientation towards Utilization': A Stimulating Conflict Potential], paper presented at the German Evaluation Society conference, Speyer, Germany,

- 4 Oct. Available at: <http://www/degeval.de/de/dokumente/Speyer2001/Jaubrich-Text-Nutzerorientierung.pdf> (site visited 20 March 2005).
- Haubrich, K. and C. Lüders (2004) 'Evaluation – mehr als ein Modewort?' [Evaluation: More than a Fashion Word?], *Recht der Jugend und des Bildungswesens* 3/2004: 316–37.
- Henry, G. T. and M. M. Mark (2003) 'Beyond Use: Understanding Evaluation's Influence on Attitudes and Actions', *American Journal of Evaluation* 24(3): 293–314.
- House, E. (2004) 'Democracy and Evaluation', keynote address at the European Evaluation Society conference, Berlin, Germany, 2 Oct.
- Kirkhart, K. (2000) 'Reconceptualizing Evaluation Use: An Integrated Theory of Influence', in V. J. Caracelli and H. Preskill (eds) *The Expanding Scope of Evaluation Use*, pp. 5–24. New Directions for Evaluation, 88. San Francisco: Jossey-Bass.
- Leeuw, F. L. (2002) 'Evaluation in Europe 2000: Challenges to a Growth Industry', *Evaluation* 8(1): 5–12.
- Lüders, C. (1993) 'Spurensuche: Ein Literaturbericht zur Verwendungsforschung' [Looking for Traces: A Literature Review about Research on Utilization], in J. Oelkers and H.-E. Tenorth (eds) *Pädagogisches Wissen* [Pedagogical Knowledge], pp. 415–37. Weinheim: Beltz.
- Luhmann, N. (1968) *Vertrauen: Ein Mechanismus der Reduktion sozialer Komplexität* [Trust: A Mechanism for Reducing Social Complexity]. Stuttgart: Enke.
- MacDonald, B. and M. Parlett (1973) 'Rethinking Evaluation: Notes from the Cambridge Conference', *Cambridge Journal of Education* 2: 74–81.
- Morabito, S. M. (2002) 'Evaluator Roles and Strategies for Expanding Evaluation Process Influence', *American Journal of Evaluation* 23(3): 321–30.
- Patton, M. Q. (1997) *Utilization-Focused Evaluation: The New Century Text*, 3rd edn. Thousand Oaks, CA: SAGE.
- Porter, T. M. (1995) *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton: Princeton University Press.
- Power, M. (1994) *The Audit Explosion*. London: Demos.
- Power, M. (1997) *The Audit Society: Rituals of Verification*. Oxford: Oxford University Press.
- Power, M. (2003) 'Evaluating the Audit Explosion', *Law and Policy* 25(3): 185–202.
- Schwarz, C. (2003) 'Von der Messtechnik zur Mediation? Rollenerwartung und Selbstverständnis in der Medienevaluation' [From Measurement Technology to Mediation? Role Expectation and Self-Conception in Media Evaluation], *Zeitschrift für Evaluation* 02/2003: 201–21.
- Schwarz, C. (2005) *Bürokratie oder Anarchie? 11 Thesen zum Evaluationsboom* [Bureaucracy or Anarchy? Eleven Theses Regarding the Evaluation Boom]. Available at: <http://evanet.his.de/evanet/positionen/positionen/positionen-2005.php> (site visited 1 April 2005).
- Schwarz, C. (2006) *Evaluation als modernes Ritual: Zur Ambivalenz gesellschaftlicher Rationalisierung am Beispiel Virtueller Universitäten* [Evaluation as a Modern Ritual: About the Ambivalence of Societal Rationalization Taking the Example of Virtual Universities], dissertation, University of Hanover. Münster: Waxmann.
- Scriven, M. (1991) *Evaluation Thesaurus*, 4th edn. Newbury Park, CA: SAGE.
- Strathern, M. (2000) *Audit Cultures: Anthropological Studies in Accountability, Ethics and the Academy*. London: Routledge.
- Weber, M. (1980) *Wirtschaft und Gesellschaft* [Economy and Society]. Tübingen: Mohr.
- Weiss, C. H. (1980) 'Knowledge Creep and Decision Accretion', *Knowledge: Creation, Diffusion, Utilization* 1(3): 381–404.

Schwarz and Struhkamp: Does Evaluation Build or Destroy Trust?

Weiss, C. H. (1987) 'The Circuitry of Enlightenment', *Knowledge: Creation, Diffusion, Utilization* 8(2): 274–81.

Weiss, C. H. and M. J. Bucuvalas (1980) *Social Science Research and Decision-Making*. New York: Columbia University Press.

CHRISTINE SCHWARZ works as an independent consultant in the field of organizational development and certification for the public health and non-profit sector. Her area of interest is the social shaping of public management. Please address correspondence to: Von-Alten-Allee 3, 30449 Hanover, Germany. [email: christine.schwarz@htp-tel.de]

GERLINDE STRUHKAMP is an independent evaluator and researcher located in Boston, Massachusetts, USA. Her areas of interest are evaluation of human services, values in evaluation and evaluation theory. Please address correspondence to: 142 Kenrick St., Apt. 12, Brighton, MA 02135, USA. [email: Gerlinde.Struhkamp@gmx.de]