EVALUATIONS: HIDDEN COSTS, QUESTIONABLE BENEFITS, AND SUPERIOR ALTERNATIVES

MARGIT OSTERLOH
Universitity of Zurich
Institute for Organization and Administrative Sciences
Email: osterloh@iou.unizh.ch

and

BRUNO S. FREY
Universitity of Zurich
Institute for Empirical Research in Economics
Email: bsfrey@iew.unizh.ch

EVALUATIONS: HIDDEN COSTS, QUESTIONABLE BENEFITS, AND SUPERIOR ALTERNATIVES

Abstract:

Research evaluation is praised as the symbol of modern quality management. We claim firstly, that output evaluations in academic research have higher costs than normally assumed, because the evaluated persons and institutions systematically change their behavior and develop counter strategies. Moreover, intrinsic work motivation is crowded out and undesired lock-in effects take place. Secondly, the benefits of output and sometimes process evaluations are questionable. These evaluations provide inadequate information for decision-making. Thirdly, there exist superior alternatives. They consist in process evaluation and careful selection and socialization – and then leave individuals and research institutions to direct themselves.

(97 words)

Key words: Evaluation, rankings, hidden costs, multi tasking, intrinsic motivation, control theory, selection.

JEL Classification: C44, D02, D61, D72, H52, I23, M12

PERFORMANCE EVALUATIONS IN ACADEMIA: EXPECTATIONS AND DISAPPOINTMENTS

Performance evaluations have become a standard procedure in quality and performance management in many organizations. This procedure today is applied in profit-oriented firms, non-profit and governmental institutions. A particularly important area has been *evaluation of academic research* on which this paper concentrates. Research evaluation is often praised as the symbol of modern quality management which today has diffused through the academic system. There are several academic journals (e.g. “Evaluation”, “Evaluation Quarterly”, “Evaluation Review”, “Research Evaluation”, “Scientometrics”) specifically devoted to the approach. “Evaluations of evaluations” or meta-evaluations are discussed (e.g. Cook & Gruder 1978).

“Evaluation” here is defined as a *formal*, *retrospective* and *external* appraisal of performance of persons or organizations. Evaluations are undertaken by other persons and not by the persons whose performance is to be assessed. Our definition corresponds both to the academic and practical use of the term. Thus, for example Brook (2002: 173) states „By evaluation, I shall mean the situation where visiting experts come from outside your organization or system and say what they think about it“. The usefulness of evaluations is rarely questioned though some observers call it “a fad” (Birnbaum 2000), or an “audit explosion” (Power 1994), leading to an “audit society” (Power 1997), and producing a booming evaluation bureaucracy and evaluation industry (Muller-Camen & Salzgeber 2005). But even the skeptics rarely discuss alternative instruments for raising the quality of research and allocating scarce resources (e.g. Daniel 1993; Gioia & Corley 2002; Holcombe 2004; Starbuck 2004; Weingart 2005).

This paper claims firstly, evaluations of academic research as currently practiced have *higher costs* than normally assumed in the evaluation literature. Secondly, the *benefits of this kind of evaluations are questionable*. Thirdly, there exist *superior alternatives*. We do not argue against evaluations in general but against the dominant form of output evaluation in academia. Instead, we discuss under which conditions process and input evaluations are appropriate.

We deal with the *hidden,* rather than the immediate *costs* of evaluation. The immediate costs of evaluation exercises in terms of academic manpower have been much complained about (e.g. Kieser 1998; de Bruijn 2001). This includes not only the effort and time expended by the evaluators. It also includes the costs borne by the evaluatees who have to provide voluminous evidence of their performance to meet the sophisticated systems of surveillance (Parker & Jary 1995:319) introduced for example in the United Kingdom.[1] Neither does our paper focus on the difficulties of measuring performance in general (see e.g. Neely & Neely 2002; Starbuck 2004, Latham et al. 2005;) nor in academic peer reviews and bibliometric indicators. Many empirical analyses of the peer reviews reveal that peer evaluations diverge, contradict each other and are not consistent over time (e.g. Cichetti 1991; Gans und Shepherd 1994; Campanario 1996, 1998; Bedeian 2004a, Starbuck 2005).[2] Bibliometric measurements of publications, citations, and co-citations as an answer to the growing skepticisms with peer reviews are also subject to major shortcomings, and can therefore not substitute for peer reviews with its own shortcomings (e.g. Üsdiken & Pasadeos 1995; Gmür 2003; Weingart 2005). These are important considerations but they have been extensively treated elsewhere. In this paper we focus on the *hidden*, and therefore often disregarded costs of performance evaluation.

[1] For a detailed description of the evaluation process of management research in the United Kingdom see Bessant et al. (2003)

[2] For a defense of peer reviews see e.g. Daniel (2005).

The *hidden costs of evaluation* consist of two main types, discussed in the first part of the paper:

(1) Evaluations *distort incentives* by (a) undue reliance on a restricted set of indicators ("you get what you measure"), known as the "multiple-task problem", (b) counter-strategies to the evaluation exercise; (c) crowding-out of intrinsic motivation of persons evaluated as well as evaluators;

(2) Evaluations *induce undesired lock-in effects* on the part of evaluated persons or institutions as well as on the part of evaluators.

The *benefits of evaluations* are usually taken for granted and remain unquestioned. The expected effects of evaluation exercises are efficient quality control, efficient allocation of resources, promotion of performance, and improved transparency of activity to principals, customers and suppliers. In particular, research evaluation is supposed to provide an effective signal to government agencies, sponsors, potential students, and their future employers about the reputation of research institutions. We argue in the second part of this paper that, firstly, evaluations provide inadequate information relevant for decision-making. Secondly, evaluations (in particular university rankings) lose importance due to new forms of scientific cooperation on the internet.

The third part of the paper is devoted to *alternatives to formal*, *retrospective* and *external* evaluations of output. They consist in

(1) An interactive and supporting coaching of persons and institutions (concurrent process control).

(2) A careful *selection and socialization* of persons charged with performing the activities in question (ex ante input control).

This paper does not provide a verdict on whether performance evaluations should be undertaken or not. This would, of course, require a comparison of *all* the costs with *all* the benefits. To the extent to which our analysis is correct, the hidden costs identified here should be added to the costs so far considered, and the benefits accordingly reduced. As a consequence, performance evaluation becomes less desirable. The intensity and frequency of evaluations should be markedly reduced and substituted by the alternatives sketched out.

HIDDEN COSTS OF EVALUATION

Evaluations Distort Incentives

In general it is taken for granted that evaluations have beneficial effects on performance. In standard economics as well as in everyday life it is assumed that performance measurement and performance pay raise performance. The search for monetary incentives has come to permeate firms, governments and research institutions. In contrast, the negative incentive effects of evaluations, in particular output evaluations, are rarely considered. Performance evaluations lead to costs, which arise because the evaluatees systematically change their behavior in an unintended way.

What is not measured is disregarded

Persons and institutions subject to an output evaluation have an incentive to perform well in those areas of their work considered in the evaluation process. In contrast, they have no incentive to perform well in those criteria which are disregarded (Kerr 1975). The phenomenon has been extensively studied in modern economics under the heading of "multiple tasking" (Holmström & Milgrom 1991; Gibbons 1998; Prendergast 1999; Fehr & Schmidt 2004). Academic research, characterized by highly incomplete contracts (Masten 2006), has to deal with this problem in several aspects.

Pressure towards "normal science" in the sense of Kuhn (1962). Measurement exerts pressure to produce predictable but unexciting research outcomes that can be published quickly (Prichard & Wilmott 1997; Muller-Camen & Salzgeber 2005). More importantly, path-breaking contributions are exactly those at variance with accepted criteria. Indeed innovative research creates novel criteria which before were unknown or disregarded. The referee process, based on the opinions of average peers often leads to the rejection of creative and unorthodox contributions and rewards the mainstream (Frey 2003). Many rejections in highly ranked journals are documented regarding papers that later were awarded high prizes, even the Nobel Prize (Gans & Shepherd 1994; Campanario 1996). It would be mistaken to assume that these are atypical and unfortunate special cases (Weingart 2005). Many luminaries in science such as Newton, Darwin, Kant, Mendel, Schopenhauer, Semmelweis, Wittgenstein and Goedel would have had little chance in a peer review system (Fischer 1998; Pasternak 2000; Gillies 2005a). Their path breaking radical innovations could only be appreciated after decades but not within short evaluation cycles such as the four to five year British Research Assessment Exercise (RAE).

Eliminating type II errors instead of type I errors. Referring to the theory of statistical tests Gillies (2005a, 2006) shows severe shortcomings of evaluations like the British Research Assessment Exercise (RAE). These evaluations are not devoted to eliminate errors of type I, but only aim to eliminate type II errors.

Type I errors occur if the test leads to the rejection of a hypothesis which is in fact true. A type I error in evaluation leads to funding being withdrawn from a researcher or a research program which would have obtained important results had it been continued. The solution of the research problem will remain undiscovered for a long time. A notable example is the case of Semmelweis (Gillies 2005b) who did research on the causes of the childbed fever in 1848, which was at the time the main cause of death in childbirth. Semmelweis (1861) followed a

procedure similar to Popper`s conjectures and refutations (Popper 2002). The vast majority of the medical profession rejected his approach and ignored the practical recommendations based upon it. Semmelweis had to leave the Vienna Maternity Hospital in 1850, where he did his research. Hospitals went through a severe crisis and thousands of patients lost their lives. Only in the mid 1880s the new antiseptic methods based on Semmelweis`s work had become general.

Type II errors occur if the test leads to the confirmation of a hypothesis which is in fact false. These errors are less serious than type I errors. A type II error leads to funding being continued which in the end obtains no good results. A type II error leads to money being spent for nothing. Unfortunately most evaluations concentrate exclusively on eliminating type II errors in order to make research more cost effective.

To avoid the serious consequences of type I errors, Gillies (2005a) goes so far as to suggest that funding bodies should make sure that at least some funding is given to *every* research school rather than concentrating on the hopeless task trying to foresee which approach will prove to be successful in the long run.

Homogenization of research endeavors. On the institutional level departments that have their eyes on rankings and output evaluations tend to become more and more similar, undermining the necessary variety in scientific discourse (see Holcombe 2004 for case of economics departments). They prefer to hire in the standard fields that get more citations or grants and avoid individuals whose work is outside the mainstream. Gioia and Corley (2002) argue that in business schools, rankings raise incentives to invest in short term management fads instead of investments in resources for basic research, which in management research follows heterogeneous approaches.

Underestimation of work committed to the dialogue between academics and practitioners. Academic work of relevance to real world phenomena is often disregarded.[3] Research departments rarely give credit to faculties who write books and magazine articles designed to intermediate between the research community and the general public because they don't contribute to the citation record. As a consequence, the gap between rigor and relevance of research is deepened and the dialogue between science and practice is undermined. [4]

Incentives to conduct expensive research. The tendency to measure research performance by the size of grants received creates an incentive to undertake more expensive, rather than relevant research (Holcombe 2004). Grants, of course, do not measure research output but rather resource input and have little to do with the scientific capabilities of an academic (Toutkoushian et al 2003).

Counter strategies to evaluation exercises

Evaluatees risking to lose from the evaluation exercise will rationally muster resources to fight against the findings. They are often able to neutralize the evaluation results *ex post* and *ex ante*.

Ex post neutralization of evaluation results. Evaluatees have strong incentives to manipulate unpropitious results ex post. Firstly, they can come forth with new criteria according to which they perform well or which are important to the decision-makers. Examples in research evaluation are criteria like teaching and administrative duties. Secondly, they claim that the low judgment of their performance is due to inadequate resources, discrimination of minorities, health problems, or simply bad luck. Thirdly, losers refer to alternative ratings or rankings in which they fare better. It is well known that there are always such alternatives

[3] The Academy of Management seeks to redress this neglect by bestowing a "Distinguished Scholar Practitioner Award".
[4] See Kieser and Nicolai (2005) for a discussion of rigor versus relevance in management research, which in their view is inevitable.

available. Inconsistent university rankings are a striking example.[5] All these arguments against the validity of a particular evaluation may be well taken or not; what matters is that they consume a large amount of attention, time, and effort, thus seriously distracting from research.

Ex ante manipulation of performance criteria. Less visible, though even more important, are strategic ex ante reactions of evaluatees intending to make forthcoming evaluations more favorable to themselves and their organization. There are many possibilities for such strategies that individually are rational but represent collectively unproductive and distorting rent-seeking. In political economy this effect is well known as „Goodhart's Law"(1975) or „Lucas Critique" (1976) and it has been empirically confirmed (e.g. Chrystal & Mizen 2003, Brück & Stephan 2006). At the micro level it is known that schools evaluations can be manipulated by "teaching to the test", excluding bad pupils from tests (for empirical evidence in USA see Figlio & Getzler 2003), or putting lower quality students in special classes which are not included in the measurement sample (Corley & Gioia 2000). Scholars are encouraged to rank competing schools low to avoid the risk of lowering the rank of their school (Argenti 2000). Much has been written about managers who "cook the books" to raise their "pay for performance" (Bebchuk & Fried 2004; Frey & Osterloh 2005; Osterloh & Frey 2000, 2006; Foss et al. 2006, Harris & Bromiley, forthcoming). In scientific publishing such strategies are undertaken by authors, editors and referees wearing different hats as evaluatees (Lawrence 2003).

Authors raise their number of publications by dividing their research results to a "least publishable unit" (Weingart 2005:125), slicing them up as thin as salami and submitting them to different journals (Lawrence 2003: 259). Authors may also offer to include another scholar among the authors in exchange for being put as co-authors on his or her paper. Time and

[5] For rankings of business schools and economic departments see e.g. Stake (1998), Dichev (1999), Thursby (2000).

energy is wasted by trying to influence editors by courting them e.g. by unnecessarily citing them. More serious are manipulations of data and results. Cheating seems to be a wide-spread misbehavior. In a questionnaire among 8,000 scientists in the United States about one third admitted behavior such as “cooking” research data, using the idea of others without appropriate reference, or failing to present data that contradict one`s own previous research (Martinson et al. 2005). The change of publication practices in response to funding based on evaluations has been demonstrated in a study for Australia (Butler 2003). The mid-1990s saw a linking of the number of peer-reviewed publications to the funding of universities and individual scholars. The number of publications increased dramatically but the quality measured by citations decreased. As a result, Australia fell behind all OECD countries with respect to quality of research (Weingart 2005:126). Of course, the process does not stop there. When the evaluators realize that their evaluation is undermined by maximizing quantity and disregarding quality (as reflected in citations) they will introduce the number of citations as quality measure. When authors adjust to this measure by maximizing the number of citations (e.g. by using citation cartels), the next step is weighing citations by impact factors, whose validity in turn is subject to strong criticism (see e.g. Adam 2002). As a consequence, an endless process of more and more sophisticated but useless evaluations emerges. As a whole, „Success in the evaluation process can become a more significant target than success in research itself“ (Brook 2002: 176).

Some *editors* freely admit that they induce authors to cite as many publications in their journal as possible in order to raise their impact factor (Smith 1997, Garfield 1997). They also like to see themselves cited in the papers submitted.

Referees are prone to judge more favorably papers that approvingly cite their own work and tend to reject papers threatening their previous work (Lawrence 2005: 260).[6] Willingly or unwillingly referees induce authors to change their papers in order to secure acceptance. Bedeian (2003) finds evidence that not less than 25 percent of authors revised their manuscripts according to the suggestions of the referee though they knew that the change was incorrect. Frey (2003) calls this behavior "academic prostitution". It has often been lamented that referees steal ideas from the papers they evaluate. Similar strategic reactions occur in other peer review processes like evaluations of research projects or academic institutions, which are a major source of funding in all countries today.

If one considers the whole process including the reactions of authors, editors, and referees, the output evaluation exercise has gained a life of its own even among those who should be involved in the very content of the scholarly discussion. As a consequence, the institutionalized evaluation process might *not* improve the outcome but rather move competition to an area where the critical academic discourse is weakened and substituted by signaling through publishing in peer reviewed journals. The "quality stamp" given by peer-reviewed journals induces scholars to cite research results without ever checking them. According to the study by Simkin & Roychowdhury (2003) on average only twenty percent of cited papers were ever read by the citing authors. As a result of these distortions a "rat-race" (Akerlof 1976) is taking place where competition does not lead to positive outcomes but rather leads to an inefficient use of scholarly resources. It intensifies and rewards those scholars engaging in collectively unproductive rent-seeking activities and distorts the working of, and undermines trust in the scholarly system. These costs normally are not taken into account when evaluation exercises are considered. Therefore, the usefulness of evaluations is highly questionable. Academia is moving towards the direction of a

[6] Such problems of sabotage in tournaments have been extensively discussed in personnel economics see e.g. Lazear (1995).

bureaucratic planning system in which the formal evaluation system has become an independent entity exerting a new kind of censorship (Biagioli 2002). Instead of competition, the institutions of a planned economy are introduced.

This process is comparable to the introduction of benefit-costs analyses into political decision-making. In contrast to what has been claimed by its supporters, and expected by parliaments and other political decision-making bodies, the main effect was not to improve the decision-making process. Rather, the discussion moved towards a more abstract and formalized level. A negative consequence is that the ordinary citizens find it more difficult to participate in the deliberations because they are not able to use the language of costs-benefit analysis. The well organized groups, including public bureaucracy, benefit at the cost of ordinary citizens (see Self 1975). This is true not only in political decision-making, but also in the academic field: Users of research results are excluded from the evaluation process though it would be crucial to consider their views (Kieser 1989).

Crowding-Out of Intrinsic Work Motivation

The application of performance evaluation, in particular output evaluation, under known conditions has negative consequences on the intrinsic work motivation of evaluatees as well as evaluators (see Frey 1997; Osterloh & Frey 2000; Lindenberg 2001; Frey & Osterloh (eds.) 2002; Fehr & Gächter 2002; Bénabou & Tirole 2003; Weibel 2006; Falk & Kosfeld 2006). The crowding-out effect has been analyzed by psychologists and economists in hundreds of laboratory experiments. The most recent and extensive meta-studies are Deci, Koestner and Ryan 1999 & Cameron, Banko & Pierce 2001 as well as the econometric studies of real life events (Frey & Oberholzer 1997; Gneezy & Rustichini 2000; for a survey of the empirical evidence, see Frey & Jegen 2001). For simple tasks, this is of little consequence, but it is generally acknowledged that for qualified, innovative and artistic work,

intrinsic incentives are of decisive importance (Osterloh 2007). This aspect has been almost totally neglected with respect to research evaluations.

An evaluation exercise has two undermining effects on the intrinsic motivation of researchers. The first is that researchers cannot pursue their intrinsic motivated research but rather must legitimize their research activity according to criteria appreciated by the evaluators. They are forced to find and state goals in line with the evaluators. The second undermining effect is brought about by the close connection between evaluation and monetary rewards. Intrinsically motivated scholars getting funds may shift their interest from the research itself to the money. This is known as the "overjustification effect" which substitutes internal by external incentives (Kruglanski et al. 1978) and thus may undermine creative research.

Performance evaluation does not necessarily have to undermine intrinsic work motivation for two reasons. Firstly, if the evaluation is perceived to be *supportive*, work morale tends to be enhanced. The same occurs if the evaluatees enjoy being the center of attention of the evaluators (the so-called "Hawthorne Effect"). The two conditions are likely to be obtained when performance evaluation is newly introduced. But once it has become a continual exercise, it increasingly tends to be perceived as controlling, crowding out intrinsic work motivation. Secondly, performance evaluation may increase work effort if the crowding out effect is overcompensated by the relative price effect (Gneezy and Rustichini 2000). In the extreme case, if intrinsic motivation is totally crowded out by evaluation exercises, research can only be induced by further monetary incentives. But it is doubtful whether innovative research can therewith be achieved (Amabile 1996, 1998). Evaluations are moreover taken as signaling a loss of trust in the evaluatees, tending to erode professional ethical norms (Kieser 1998).

Evaluation exercises may also negatively affect the intrinsic motivation of evaluators. In many cases (e.g. in the approach pursued by of the Royal Netherlands Academy of Arts and Sciences 2005) the evaluators are put into a straightjacket of formalized procedures leaving little or no room for aspects considered important from the personal point of view. They have to perform within a bureaucratic formal output evaluation system. In such an environment evaluators have no scope to act in a role of participants in an intellectual discourse, which often was the reason to engage in a particular evaluation process.

Lock-in Effects

Once an evaluation system is in place in a particular organization even those members who are aware of the hidden costs and questionable benefits are rational not to oppose it. If they would, they are easily accused of being afraid of the consequences. Therefore, it is a better strategy to go along, understand the rules and play the game. As a consequence it appears that most, if not all members of the organization evaluation exercises accept the evaluation imposed, though in fact they are only resigned because they see no other option.

Evaluators are also locked into the system. Scholars are invited to participate in evaluation exercises often because they have a personal or professional relationship to the persons or institutions to be evaluated. Once they have agreed, they have little or no incentive to oppose the bureaucratic procedures involved. They therewith become part of a quasi-socialist planning system, supporting its further existence.

Because both evaluatees and evaluators are locked into the system, formal evaluations are expected to persist for a considerable time.

QUESTIONABLE BENEFITS OF EVALUATIONS

Performance evaluations are often of little or no use for three quite different reasons.

Redundant or Contestable Information Provided

In many cases information provided by evaluations is used only to bolster political decisions already made beforehand (Mittelstrass 2000). These decisions are mostly based on obvious criteria. In the case of a formal evaluation of particular individuals, substantial decisions (in particular promoting a person to an important new position or, in contrast, firing him or her) will only be taken if the evidence seems to be clear – and in the majority of the cases this is quite well known *before* the formal evaluation has been undertaken. The decision makers are often able to see well whether a person performs above or below average, according to the established criteria.[7] The same holds for the evaluation of projects and institutions. In that case also, the formal evaluation at best serves to support what already seems to be clear.

However, in some cases, new information is provided by the result of output evaluations – but such additional information rarely leads to changes in decisions. When the evaluation produces surprising information about performance, it is easy to put it into doubt or to manipulate it at will (e.g. by changing the weight of the various components in the evaluation). For example, there have been many evaluations of possible sites for nuclear refuse deposits. There have virtually been hundreds of studies costing millions and even billions of dollars and euros but in almost no case a clear decision could be reached on the basis of that method. (e.g. the discussion on the Yucca Mountain in the US, see Portney 1991, Rabe 1994, Easterling and Kunreuther 1995).

Information Irrelevant for Decision-making

An evaluation should capture the *marginal effect on performance of additional resources.* How would performance change if an institution or a researcher could dispose over more or less means? This question is difficult to answer because it depends on a large number of

[7] Nevertheless they often fail in assessing the performance of individuals, see Latham et al 2005.

conditions. An evaluation considering *marginal performance effects* is much more costly to undertake than the evaluations considering absolute performance levels. Hence, the balance between the benefits and costs of evaluations is worsened. Accordingly, the consequence of a positive or a negative evaluation of a person or institution remains unclear. Should the resources available to the institutions and researchers evaluated as "bad" be reduced? Or should they be given *additional* means so that they can improve the quality of their research? Should the resources available to those evaluated as "good" or "excellent" be kept constant (or even be reduced) because they prove to be successful? These questions suggest that the result of an evaluation based on performance *levels* leaves an essential aspect unanswered.

Rankings Lose Importance

The internet has fundamentally modified the production process taking place in research institutions as well as in knowledge production in general. A thoroughly conducted study by Kim, Morse and Zingales (2006) reveals that the position of top universities leading to the Matthew Effect loses importance.[8] While in the 1970s residence in an elite university had a sizeable impact on individual research productivity (measured by impact equivalent pages per year), in the 1990s this effect disappeared. Also the percentage of co-authored papers with scholars of non-elite universities doubled. It follows that the externality of having better local research colleagues declines because it is easier to collaborate by internet with researchers located at any other university.

ALTERNATIVES TO FORMAL OUTPUT EVALUATIONS

The much disregarded hidden costs and underperformance of currently practiced, output evaluations would matter little if there were no alternatives. Indeed, it has often been claimed that no reasonable alternatives exist (e.g. Royal Netherlands Academy of Arts and Sciences

[8] Medoff (2006) finds that the Matthew Effect applies to the Universities of Harvard and Chicago, only.

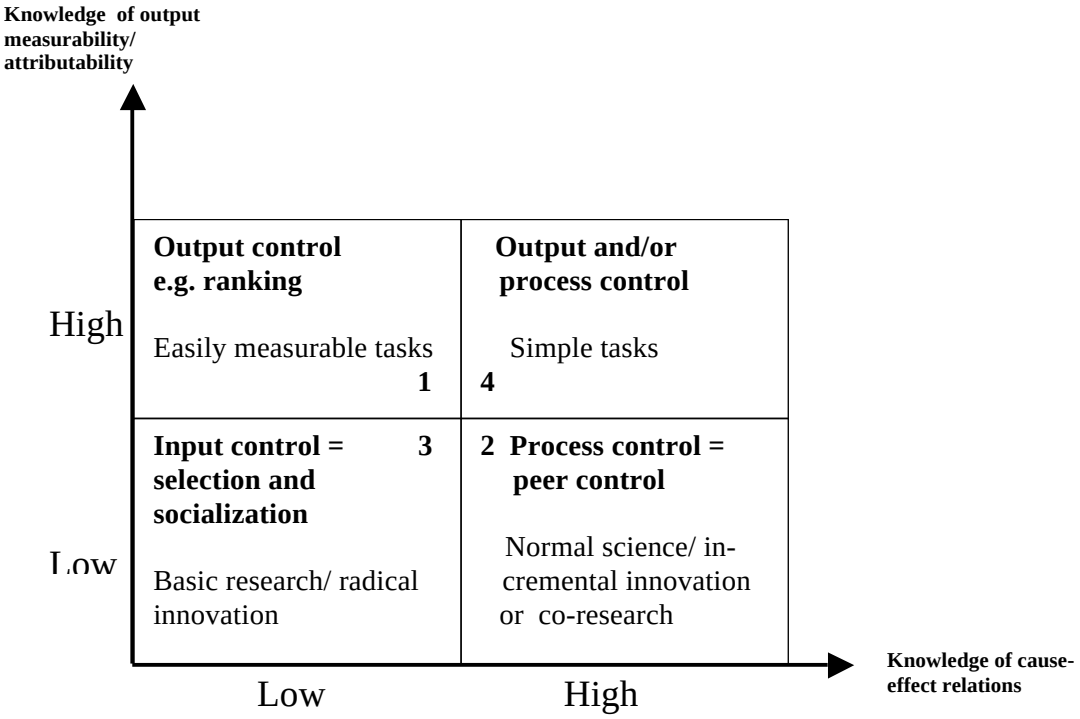
2005). But there *are* superior alternatives. In the following, three alternatives to retrospective output evaluations are discussed: Process control, input control, and a combination of process and input control.

We draw on insights from managerial control theory (e.g. Thompson 1967; Ouchi 1977, 1979; Eisenhardt 1985) which differentiates between output control, process control and input control. A major result of that theory is that the kind of evaluation or control system must fit the characteristics of the knowledge available to the evaluator (Turner and Makhija 2006). These characteristics are defined on the one hand by "knowledge of measurability and attributability of outputs". On the other hand, they are defined by "knowledge of cause/effect relations" (Thompson (1967) or "knowledge of the transformation process" (Ouchi 1979). The relationships between control forms and knowledge available to the evaluator are summarized in Figure 1 and discussed below. It is important to keep in mind that all organizations, including research organizations, employ a combination of control types, though with different emphasis.

We concentrate on process control and input control (cell 2 and 3). We do not discuss cell 4, because it is quite clear that activities that can be controlled by outputs *and*/*or* well defined processes characterize simple tasks apart from research. To make clear why output control of research (cell 1) should be avoided, we start with comments on cell 1 from the point of view of managerial control theory.

FIGURE 1:

Relationship Between Control Forms and Knowledge Available to the Evaluator



Output control (cell 1) is based on well-defined indicators, which is the essence of output evaluations and in particular rankings. Examples for such indicators in research are e.g. the number of published articles in refereed journals or citations. Since output controls do not specify the processes or cause-effect relations that produce that outputs, such a control type is most appropriate when process or cause-effect relations are difficult to specify as long as the outputs are measurable. Output controls are attractive to non-experts, like politicians, journalists, and bureaucrats. This is the reason why newspapers love output controls in the form of rankings, because rankings seem so provide easy to understand quality signals. Moreover, this kind of control gives the controlled person or institution a certain measure of discretion how to reach certain goals or sub-goals, which has been discussed intensely in the "management by objectives" approach (e.g. Lawler and Rhode 1976) and in the literature

about modularization (e.g. Baldwin and Clark 2000; Fleming and Sorenson 2001; Langlois 2002).

However, there are some preconditions of output control which are lacking in the case of research. *Firstly*, the knowledge relating to the output must be stable and not subject to change (Snell 1992). In research, desired outputs are ambiguous and the criteria of what constitutes good research are changing, in particular when radical innovations (Christensen and Bower 1996) or paradigm shifts (Kuhn 1962) are concerned. In these cases, the usefulness of established knowledge and of well-defined indicators is put into question. *Secondly*, outputs must be observable and attributable (Eisenhardt 1985). Today's research is characterized by intensive pooled interdependencies between the various scholars involved in a research enterprise and therefore often lacks attributability to particular scholars and even institutions. As a consequence, research quality cannot well be captured by output-oriented evaluations like e.g. rankings.[9] If they are nevertheless applied, multi-task problems emerge which distort incentives (as discussed above) and lead to risk-adverse and imitative behavior (Ouchi 1977).[10]

Fortunately, according to managerial control theory there exist adequate alternatives to output controls, namely process and input control or a combination of both.

Process control: Interactive and supportive coaching

As output control, and therefore formal output evaluations, do not work well in the field of research, process control may present a useful alternative (cell 2). The preconditions are that evaluators (a) have the appropriate knowledge of cause/effect relations or of the transformation process of inputs into outputs and (b) have a shared understanding of the rules

[9] At best, one could argue for a variety of rankings with diverse criteria to make clear that no single ranking can meet the requirements of scholarly work, see Gioia and Corley (2002: 118).
[10] But see Cardinal (2001). She finds that output control enhances radical innovation in the pharmaceutical industry. But the dependent variable – FDA approvals - might not well serve as an indicator of radical innovation, and might be subdue to the multiple task effect itself.

obtained. This is often the case in academic peer reviews. Peer reviews are useful to make sure that well established standards of research methodology are met. However, as argued above, peer reviews have major shortcomings when unorthodox contributions are to be evaluated. As a consequence, process control works well only in the case of "normal science" (Kuhn 1962) or incremental innovations.[11] Normal science exploits potentials of established knowledge and introduces relatively little changes to it. Applying process control to radical innovations or to paradigmatic shifts has major shortcomings, because there is little agreement about which criteria are applicable and which are not. In these cases, process control or peer evaluation can only be used if it takes the form of an interactive research process between evaluator and evaluatee in which new criteria are jointly developed. However, in this case the roles of evaluator and evaluatee are intermingled. The evaluator is no longer an external, independent outsider but is inevitably involved as a contributor to the research itself. The danger arises that he or she is no longer a person who evaluates from a neutral point of view.

Input control: Careful Selection and Training as Alternative

Neither *output control* nor *process control* work sufficiently when measurability and attributability of outputs is not given *and* the external evaluators´ knowledge of the transformation process is limited. This is the case in most knowledge work and in particular in basic research endeavors. In these cases, *input control* in the form of careful selection and training is central (cell 3). In managerial control theory this type of control is also called "clan control" (Ouchi 1979). Selection and socialization has to make sure that candidates become members of a community in which aligned norms and values are internalized and are part of their intrinsic motivation. Input control represents prospective controls, regulates

[11] It should be noted that sometimes there are conflicting logics between science and innovation in industry, see Gittelman and Kogut (2003)

antecedent conditions of performance and manages potentials (Snell 1992). This is of particular importance for knowledge work, as social relationships matter for knowledge creation, retention, and transfer (Kogut and Zander 1996; Duncan, Ford, Bousculp & Ginter 2002, Argote et al. 2003; Nickerson and Zenger 2004, Bedeian 2004b)[12]. If input control is successful, mutual tolerance for ambiguity is possible, which is important in production processes where output measurement is questionable *and* procedural rules are unclear. This makes input control the main form of control for all kinds of knowledge work, in particular for basic research.

The strategy is to use resources to find the persons best suited for a job and to consider how he or she is likely to perform *in the future* - and then to have trust that he or she will indeed perform well. Thus, after a careful selection and training one has to abstain from external evaluations in terms of output and to some extent also in terms of process control. Such a control approach to scientific research was emphasized by the famous President of Harvard University James Bryan Conant (Renn 2002):

> „There is only one proved method of assisting the advancement of pure science – that is picking men of genius, backing them heavily, and leaving them to direct themselves.“ (Letter to the New York Times, 13. August 1945).

This view is still part of the „Principles Governing Research at Harvard", stating:[13]

[12] But see Nooteboom (2000) for a discussion of the problem of optimal cognitive distance and cognitive proximity in knowledge production

[13] See http://www.fas.harvard.edu/research/greybook/principles.html.

> „ The primary means for controlling the quality of scholarly activities of this Faculty is through the rigorous academic standards applied in selection of its members“

Input control as main control form does not apply to the whole academic career. During the socialization and selection process much process control and some output control must take place. They have to ensure that the candidate knows the rules of scholarly work, meets the established standards of research methods, is performing in an efficient way, and has sufficient intrinsic motivation to work on its own. These criteria serve as basis of trust in the ability, willingness and creativity of a researcher and are the preconditions to meet ambiguous tasks. But it makes a great difference being submitted to process and output control during a whole life than being submitted to it during a certain, clearly delineated phase knowing that the appointment will imply a wide extent of autonomy and self-determination.

The avoidance of output and process control after the appointment as a full professor is likely to lead *some* researchers not to perform after having been appointed. But this is the necessary price to pay for the most able and innovative scholars to flourish.

Selection and socialization processes combined with essentially putting faith in the persons chosen have been used with much success in many areas in society. It suffices to indicate three cases pertaining to some of the most important functions accorded to in society:

(1) In most countries committed to the rule of law high judges after having been appointed are not subjected to any sort of ex post formal evaluation process. This applies to chief judges in the United States, or the federal judges in Germany and in Switzerland. In many countries, the highest judges are elected for life, and thus are not even subject to a re-election constraint.

(2) The presidents and other top members of Central Banks are selected, and their performance is not evaluated by any sort of ex post evaluation. Often, they are elected for rather long periods.

(3) University professors were appointed for life, and not submitted to formal ex post evaluations, in countries such as Germany, Austria and the United Kingdom when they were the leading centers of scientific research (Gillies 2005a).

In all three cases, the selection and socialization process is expected to be so careful that once it is taken one can trust that the persons selected have sufficient intrinsic motivation to perform well. For high judges and top members of central banks this principle has remained intact. In contrast, for universities this principle has recently been undermined. Notwithstanding, it is crucial for research and radical innovations because of three reasons: *Firstly*, individuals as well as institutions are not induced to only focus on easily measurable dimensions of their task and to disregard tasks difficult to identify and measure. The distorting effects of multi-tasking and rat races are evaded. *Secondly*, the process of selection does not induce crowding out of intrinsic work motivation because the affected persons do not perceive it as controlling. As long as the selection process is considered to be procedurally fair (see Lind and Tyler 1988; Frey and Stutzer 2001 for empirical evidence on procedural utility) their intrinsic motivation tends to be raised rather than lowered (Osterloh and Weibel 2006). This effect occurs for the persons chosen, but less so for the persons not selected. Institutions are therefore well advised to apply the "up or out" rule. This selection rule makes sense in cases where intrinsic motivation is crucial and when the not selected persons are expected to have a lower work morale.[14]

[14] This explanation for the "up or out" rule differs from the explanation in orthodox economics (e.g. Milgrom and Roberts 1992: 379-382). In this literature "up or out" is understood as an instrument to overcome opportunism of decision makers and potential partners when performance is not contractable.

Combination of input and process control

Input and process control may usefully be combined when institutions, not persons, are to be evaluated. The selection process and the autonomy of researchers being selected and appointed constitute the essential condition for scholarly work. The process control of institutions therefore must ensure, *firstly*, that the selection process meets the requirements discussed above, namely a rigorous evaluation of the potential candidates, and, *secondly,* that they are granted as much autonomy as possible once appointed.[15] Indeed, some highly productive knowledge producing companies today at least partly follow this rule. Examples are 3M, Siemens and Google.[16] These companies allow their researchers to spend fifteen to forty percent of their work time in pursuing self-chosen goals (Brand 1998).[17] This should, of course, apply even more strongly to academic research.

CONCLUSIONS

This paper's goal has been to identify and analyze hidden costs, as well as the questionable benefits of evaluations in academia, in particular of output and process evaluations, that have been neglected or treated lightly. It is not intended to present an overall judgment of the desirability of such methods. This would only be possible if all the benefits and costs were considered. However, in so far as the benefits and the costs normally considered remain constant, the identification of additional costs result in formal evaluation exercises being less desirable.

Two main kinds of hidden costs were discussed:

[15] This is the kind of evaluation the German Science Council (Deutscher Wissenschaftsrat) usually conducts. Recently the Wissenschaftsrat has decided to engage in a pilot study concerning a rating of research institutions similar to the British Research Assessment Exercise (see Wissenschaftsrat 2004), which is the kind of output control we criticize in this paper.
[16] Another impressive example for how autonomy in knowledge productions furthers productivity is open source software production, see Osterloh and Rota (forthcoming)
[17] The Google way is documented in e.g. http://www.infoworld.com/article/04/02/20/08OPconnection_1.html

(1) Distorted incentives due to (a) identifying and measuring only some, but not all aspects of performance (the multi-task problem), (b) Counter strategies to the evaluation exercises, which induce blocking reactions by the persons negatively affected, and (c) crowding out intrinsic work motivation crucial for creative scholarly work;

(2) Lock-in effects of evaluatees and evaluators leading to an undue persistence and expansion of bureaucratic interventions.

While formal ex post evaluations have undoubted benefits, they are normally overestimated. The benefits of these evaluations are questionable for three reasons:

(1) They often do not produce new information not already known by peers;

(2) The information produced is not helpful for a reasoned decision-making;

(3) Rankings of universities lose importance due to the new possibilities of scientific cooperation via the internet.

Some general, but important, conclusions follow from the analysis:

- Retrospective formal evaluations are not a method to be successfully used nearly everywhere as it is the case today. They can *not* be considered a “modern” system of quality management, in particular for individuals and institutions engaged in research. Rather, retrospective formal evaluations should be used sparingly.

- Careful (ex ante) selection and socialization presents a superior alternative to ex post evaluation. The characteristic of a selection system is that once a decision has been made the principals put faith in the persons selected. Important positions in society (such as top judges and presidents of Central Banks) are elected either for life or for a very long time

period without formal evaluations for good reasons. It is questionable why these reasons should not apply to research.

- The behavior of creative persons fulfilling ambiguous tasks of low programmability needs to be based on intrinsic motivation and the ability to direct themselves. The selection process must above all be directed to this goal.

- Research institutions are to be evaluated by a combination of process and input control. The evaluation primarily has to assess whether the research organizations (a) select the most promising scholars, (b) back them with institutional rules securing them autonomy, and (c) then leave them to direct themselves.

REFERENCES

Adam, D. 2002. The Counting Houses. ***Nature***, 415: 726-729.

Akerlof, G. A. 1976. The Economics of Caste and Rat Race and other Woeful Tales. ***Quarterly Journal of Economics***, 90: 599-617.

Amabile, T. 1996. ***Creativity in Context: Update to the social Psychology of Creativity***. Boulder: Westview Press.

Amabile, T. 1998. How To Kill Creativity. ***Harvard Business Review***, 76(5)(Sep-Oct): 76-87.

Argenti, P. 2000. Branding B-Schools: Reputation Management for MBA-Programs. ***Corporate Reputation Review***, 3(2): 171-178.

Argote, L., Mc Evily, B., & Reagans, R. 2003. Managing Knowledge in Organizations: An Integrative Framework and Review of Emerging Themes. ***Management Science***, 49(4): 571-582.

Baldwin, C. Y., & Clark, K. B. 2000. ***Design Rules. The Power of Modularity***. Boston: MIT Press.

Bebchuk, L. A., & Fried, J. 2004. ***Pay Without Performance. The Unfulfilled Promise of Executive Compensation***. Cambridge, Mass.: Harvard University Press.

Bedeian, A. G. 2003. The Manuscript Review Process: The Proper Roles of Authors, Referees and Editors. ***Journal of Management Inquiry***, 12: 331-338.

Bedeian, A. G. 2004a. Peer Review and the Social Construction of Knowledge in the Management Discipline. ***Academy of Management Learning and Education***, 3 (2): 198-216.

Bedeian, A. G. 2004b. The Gift of Professional Maturity. ***Academy of Management Learning and Education***, 3 (1): 92-98.

Bénabou, R., & Tirole, J. 2003. Intrinsic and Extrinsic Motivation. ***Review of Economic Studies***, 70(3): 489-520.

Bessant, J., Birley, S., Cooper, C., Dawson, S., Gennard, J., Gardiner, M., Gray, A., Jones, P., Mayer, C., McGee, J., Pidd, M., Rowley, G., Saunders, J., Stark. A. 2003. The state of the field in UK management research: Reflections of the Research Assessment Exercise (RAE) panel. ***British Journal of Management***, 14(1): 51-68.

Biagioli, M. 2002. From Book Censorship to Academic Peer Review. ***Emergences: Journal for the Study of Media & Composite Cultures***, 12(1): 11-45.

Birnbaum, M. 2000. MLA millennial questionnaire on university issues – Reply. ***PMLA Publications of the modern Language Association of America***, 115(7): 2044-2045.

Brand, A. 1998. Knowledge Management and Innovation at 3M. ***Journal of Knowledge Management***, 2(1): 17-22.

Brook, R. 2002. ***The Role of Evaluation as a Tool for Innovation in Research*.** Proc. 2000 Max Planck Forum 5, Innovative Structures in Basic Decision Research. Munich: Ringberg Symposium, 173-179.

Brück, T., & Stephan, A. 2006. Do Eurozone Countries Cheat with their Budget Deficit Forecasts? ***Kyklos***, 59: 3-16.

Butler, L. 2003. Explaining Australia's increased share of ISI publications - the effects of a funding formula based on publication counts. ***Research Policy***, 32(1): 143-155.

Cameron, J., Banko, K. M., & Pierce, W. D. 2001. Pervasive negative effects of rewards on intrinsic motivation: The myth continues. ***The Behavior Analyst***, 24: 1-44.

Campanario, J. M. 1996. Using Citation Classics to study the incidence of serendipity in scientific discovery. ***Scientometrics***, 37(1): 3-24.

Campanario, J. M. 1998. Peer review for journals as it stands today - Part 1. ***Science Communication***, 19(3): 181-211 - Part 2. ***Science Communication***, 19(4): 277-306.

Cardinal, L. B. 2001. Technological Innovation in the Pharmaceutical Industry: The Use of Organizational Control in Managing Research and Development. ***Organization Science***, 12(1): 19-36.

Christensen, C. M., & Bower, J. L. 1996. Customer Power, Strategic Investment, and the Failure of Leading Firms. ***Strategic Management Journal***, 17(3): 197-218.

Chrystal K., Mizen, A., & Mizen, P. D. 2003. Goodhart's Law: its origins, meaning and implications for monetary policy. In P.D. Mizen (Ed). ***Central Banking, Monetary Theory and Practice: Essays in Honour of Charles Goodhart***: 221-243. Volume 1. Cheltenham, U.K. and Northampton, MA, USA: Edward Elgar.

Cichetti, D.V. 1991. The reliability of peer review for manuscript and grant submissions: A cross-disciplinary investigation. ***Behavioral and Brain Sciences***, 14: 119-135. Discussion 135-186.

Cook, Th. D., & Gruder, Ch. L.. 1978. Metaevaluation Research. ***Evaluation Quarterly***, 2: 5 – 51.

Corley, K. G., & Gioia, D. A. 2000. The Rankings Game. Managing Business School Reputation. ***Corporate Reputation Review***, 3(4): 319-333.

Daniel, H. - D. 1993. ***Guardians of Science: Fairness and Reliability of Peer Reviews***. Winheim: Wiley-VCH.

Daniel, H. - D. 2005. Publications as a measure of scientific advancements and of scientists´ productivity. ***Learned Publishing***, 18: 143-148.

De Bruijn, H. 2002. ***Managing Performance in the Public Sector***. London and New York: Routledge.

Deci, E. L., Koestner, R., & Ryan, R. M. 1999. A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation. ***Psychological Bulletin***, 125(6): 627-668.

Dichev, I. 1999. How good are business school rankings? ***Journal of Business***, 2: 201-213.

Duncan, W.J., Ford, E., Rousculp, M.D. 2002. Community of Scholars: An exploratory study of management scholars. ***Scientometrics***, 55: 395-409.

Eisenhardt, K. M. 1985. Control: Organizational and economic approaches. ***Management Science***, 31(2): 134-149.

Easterling, D., & Kunreuther, H. 1995. ***The dilemma of siting a high-level nuclear waste repository***. Boston: Kluwer Academic Publishers.

Falk, A., & Kosfeld, M.2006. The hidden cost of control. ***American Economic Review***, 96(5): 1611-1630.

Fehr, E., & Gächter, S. 2002. ***Do Incentive Contracts Crowd Out Voluntary Cooperation?*** Working Paper No. 34, Institute for Empirical Research in Economics.

Fehr, E., & Schmidt, K. M. 2004. Fairness and Incentives in Multi-Task-Principal-Agent Model. ***The Scandinavian Journal of Economics***, 106(3): 453-474.

Figlio, D., & Getzler, L. 2003. **Accountability, Ability and Disability: Gaming the System**. NBER Working Paper No 9307.

Fischer, K. 1998. Evaluation der Evaluation. ***Wissenschaftsmanagemen****t*, 5: 16-21.

Fleming, L., & Sorenson, O. 2001. The Dangers of Modularity. ***Harvard Business Review***, 79(8): 20-23.

Foss, K., Foss, N. J., & Vasquez, H. 2006. "Tying the Manager`s Hands: Constraining Opportunistic Mangerial Intervention. ***Cambridge Journal of Economics***, 30(5): 797-818.

Frey, B. S. 1997. ***Not Just for the Money: An Economic Theory of Personal Motivation***. Cheltenham, U.K.: Edward Elgar

Frey, B. S. 2003. Publishing as prostitution? – Choosing between one's own ideas and academic success. ***Public Choice***, 116: 205-223.

Frey, B. S., & Jegen, R. 2001. Motivation Crowding Theory. ***Journal of Economic Surveys***, 15(5): 589 - 611.

Frey, B. S., & Osterloh, M. (Eds.) 2002. ***Successful Management by Motivation – Balancing Intrinsic and Extrinsic Incentives***. Berlin , Heidelberg, New York: Springer Verlag.

Frey, B. S., & Osterloh. M. 2005. Yes, Managers Should Be Paid Like Bureaucrats. ***Journal of Management Inquiry***, 14: 96-111.

Frey, B. S., Oberholzer-Gee, F., & Bohnet I. 1997. Fairness and competence in democratic decisions. ***Public Choice***, 91(1): 89-105.

Frey, B. S., & Stutzer, A. 2001. Procedures Matter in Political Economy. ***Journal of Public Choice and Constitutional Economics***, 2: 3-18.

Gans, J.S., & Shepherd, G. B. 1994. How Are the Mighty Fallen: Rejected Classic Articles by Leading Economists. ***Journal of Economic Perspectives***, 8: 165-179.

Garfield, E. 1997. Editors are justified in asking authors to cite equivalent references from same journal. ***British Medical Journal***, 314(7096): 1765.

Gibbons, R. 1998. Incentives in Organizations. ***Journal of Economic Perspectives***, 12: 115-132.

Gillies, D. 2005a. ***Lessons from the History and Philosophy of Science regarding the Research Assessment Exercise***. *Paper read at the Royal Institute of Philosophy on 18 November 2005. (www.ucl.ac.uk/sts/gillies).*

Gillies, D. 2005b. Hempelian and Kuhnian Approaches in the Philosophy of Medicine: The Semmelweis Case. ***Studies in History and Philosophy of Biological and Biomedical Sciences***, 36: 159-181.

Gillies, D. 2006. Why Research Assessment Exercises Are a Bad Thing. ***Post-autistic economics Review***, 37: 2-9.

Gioia, D., & Corley, K. G. 2002. Being Good Versus Looking Good: Business School Rankings and the Circean Transformation From Substance to Image. ***Academy of Management Learning and Education***, 1(1): 107-120.

Gittelman, M., & Kogut, B. 2003. Does Good Science Lead to Valuable Knowledge? Biotechnology Firms and the Evolutionary Logic of Citation Patterns. ***Management Science***, 49(4): 366-382.

Gmür, M. 2003. Co-Citation Analysis and the Search for Invisible Colleges: A Methodological Evaluation. ***Scientometrics***, 57(1): 1-31.

Gneezy, U., & Rustichini, A. 2000. Pay Enough or Don't Pay at All. ***Quarterly Journal of Economics***, 115(3): 791-810.

Harris, J. & Bromiley, P. forthcoming. Incentives to Cheat: The Influence of Executive Compensation and Firm Performance on Financial Management. ***Organization Science***.

Holcombe, R. G. 2004. The National Research Council Ranking of Research Universities: Its Impact on Research in Economics. ***Econ Journal Watch***, 1: 498-514.

Holmstrom, B., P. Milgrom. 1991. Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. ***Journal of Law, Economics, and Organization*** 7(2) 24-52.

Kerr, St. 1975. On the Folly of Rewarding A while Hoping for B. ***Academy of Management Journal***, 18: 769-783.

Kim, H. E., Morse, A., & Zingales, L. 2006. ***Are Elite Universities Losing Their Competitive Edge?*** NBER Working Papers No. 12245.

Kieser, A. 1998. Going Dutch - Was lehren niederländische Erfahrungen mit der Evaluation universitärer Forschung? ***DBW***, 58: 208-224.

Kieser, A., & Nicolai, A. T. 2005. Success Factor Research. Overcoming the Trade-Off Between Rigor and Relevance? ***Journal of Management Inquiry***, 14(3): 275-279.

Kogut, B., & Zander, U. 1996. What Firms Do?: Coordination, Identity, and Learning. ***Organization Science***, 7: 502–518.

Kruglanski, A.W., Schwartz, J. M., Maides, S., & Hamel, I. Z. 1978. Covariation, discounting and augmentation: Toward a clarification of attributional principles**.** ***Journal of Personality***, 46: 176-180.

Kuhn, Th. S. 1962. ***The Structure of Scientific Revolution***. Chicago USA: University of Chicago Press.

Langlois, R. 2002: Modularity in Technology and Organization. ***Journal of Economic Behavior and Organization***, 49(1): 19-37.

Latham, G. P., Almost, J., Mann, S., & Moore, C. 2005. New Developments in Performance Management. ***Organzational Dynamics***, 34: 77-87.

Lawrence, P. A. 2003. The politics of publication - Authors, reviewers and editors must act to protect the quality of research. ***Nature***, 422(6929): 259-261.

Lawler, E. E., & Rhode, G. 1976. ***Information and control in organizations***. Pacific Palisades, CA: Goodyear.

Lazear, E. P. 1995. ***Personnel Economics***. Cambridge, Massachusetts: MIT Press.

Lind, A. E., & Tyler, T. R. 1988. ***The Social Psychology of Procedural Justice***. New York: Plenum Press.

Lindenberg, S. 2001. Intrinsic Motivation in a New Light. ***Kyklos***, 54: 317-343.

Martinson, B. C., Anderson, M. S., & de Vries, R. 2005. Scientists behaving badly. ***Nature***, 435(7043): 737-738.

Masten, S. E. 2006. Authority and Commitment: Why Universities, like Legislatures, are not Ortganized as Firms. ***Journal of Economics & Management Strategy***, 15(3): 649-684.

Medoff, M. H. 2006. Evidence of a Harvard and Chicago Matthew Effect. ***Journal of Economic Methodology***, 13 (4): 485-506.

Milgrom, P. R., & Roberts, J. 1992. ***Economics, Organization and Management***. Englewood Cliffs, NJ: Prentice-Hall.

Mittelstrass, J. 2000. Exzellenz und Mittelmass. ***Gegenworte, Zeitschrift für den Disput über Wissen***, 5: 22-25.

Muller-Camen, M., & Salzgeber, St. 2005. Changes in Academic Work and the Chair Regime: The Case of German Business Administration Academics. ***Organization Studies***, 26(2): 271-290.

Neely, A. D., & Neely A. 2002 (Eds.). ***Business Performance Measurement: Theory and Practice.*** Cambridge: Cambridge University Press

Nickerson, J. A., & Zenger, T. R. 2004. A Knowledge-Based Theory of the Firm – The Problem-Solving Perspective. ***Organization Science***, 15(6): 617-632.

Nooteboom, B. 2000. Learning by Interaction: Absorptive Capacity, Cognitive Distance and Governance. ***Journal of Management and Governance***, 4: 69-92.

Osterloh, M. 2007. 'Human Resources Management and Knowledge Creation', In I. Nonaka, & I. Kazuo (Eds.) ***Handbook of Knowledge Creation***. Oxford: Oxford University Press: 158 – 175.

Osterloh, M., & Frey, B. S. 2000. Motivation, Knowledge Transfer, and Organizational Forms. ***Organization Science***, 11: 538-550.

Osterloh, M., & Frey, B. S. 2006. Shareholders Should Welcome Knowledge Workers as Directors. ***Journal of Management and Governance***, 10(3): 325-345.

Osterloh, M., & Rota, S. (forthcoming) Open Source Software Production – Just Another Case of Collective Invention? ***Research Policy***.

Osterloh, M., & Weibel, A. 2006. ***Investition Vertrauen***. Wiesbaden: Gabler Verlag.

Ouchi, W. G. 1977. The relationship between organizational structure and a organizational control. ***Administrative Science Quarterly***, 22: 95-113.

Ouchi, W. G. 1979. A conceptual framework for the design of organizational control mechanisms. ***Management Science***, 25(9): 833-848.

Parker, M., & Jary, D. 1995. The McUniversity: Organization, management and academic subjectivity. ***Organization***, 2/2: 319-338.

Pasternack, P. 2000. Besoldete Qualität? Qualitätsbewertung und leistungsgerechte Besoldung. ***Wissenschaftsmanagement***, 4: 8-13.

Popper, K. 2002. ***Conjectures and Refutations: The Growth of Scientific Knowledge***, Reprint, London/New York: Routledge.

Portney, K. E. 1991**. *Siting hazardous waste treatment facilities: The Nimby Syndrome***. New York: Auburn House.

Power, M. 1994. ***The Audit Explosion***. London: Demos.

Power, M. 1997. ***The Audit Society. Ritual of Verification***. Oxford: Oxford University Press.

Prendergast, C. 1999. The Provision of Incentives in Firms. ***Journal of Economic Literature***, 37(March): 7–63.

Prichard, C., & Willmott, H.. 1997. Just how managed is the McUniversity? ***Organization Studies***, 18(2): 287-316.

Rabe, B. G. 1994. Alternatives to NIMBY Gridlock – voluntary approaches to radioactive-waste facility siting in canada and the United States. ***Canadian Public Administration – Administation publique du Canada***, 37(4): 644-666.

Renn, J. 2002. ***Challenges from the Past. Innovative Structures for Science and the Contribution of the History of Science***. Proc. 2000 Max Planck Forum 5, Innovative Structures in Basic Decision Research. Ringberg Symposium, Munich, 25-36.

Royal Netherlands Academy of Arts and Sciences. 2005. ***Judging Research on its Merits***. Amsterdam.

Self, P. 1975. ***The Politics and Philosophy of Cost-Benefit Analysis***. London: Macmillan.

Smith, R. 1997. Journal accused of manipulating impact factor. ***British Medical Journal***, 314: 463.

Snell, S. A. 1992. Control Theory in strategic human-resource management – The mediating effect of administrative information. ***Academy of Management Journal***, 35(2): 292-327.

Simkin, M. V., & Roychowdhury, V. P. 2005. Copied citations create renowned papers? ***Annals of Improbable Research***, 11(1): 24-27.

Stake, J. E. 1998. ***Rankings are Dangerous!*** http//monoborg.law.indiana.edu/LawRank/rankgame_danger.html.

Starbuck, W. H. 2004. Methodological Challenges Posed by Measures of Performance. ***Journal of Management and Governance***, 8: 337-343.

Starbuck, W. H. 2005. How Much Better Are the Most-Prestigious Journals? The Statistics of Academic Publication. ***Organization Science***, 16(2): 180-200.

Thompson, J. D. 1967. ***Organizations in Action - Social Science Bases of Administrative Theory.*** New York; St.Louis; San Francisco; Toronto; London; Sydney: McGraw-Hill Book Company.

Thursby, J. G. 2000. What Do We Say about Ourselves and What Does it Mean? Yet Another Look at Economics Department Research. ***Journal of Economic Literature***, 38: 383-404.

Toutkoushian, R. K., Porter, St. R., Danielson, Ch., & Hollis, P. R. 2003. Using Publication Counts to Measure an Institution´s Research Productivity. ***Research in Higher Education***, 44(2): 121-148.

Turner, K. L., & Makhija, M. A. 2006. The Role of Organizational Controls in Managing Knowledge. ***Academy of Management Review***, 31(1): 197- 217.

Üsiken, B., & Pasadeos, Y. 1995. Organizational Analysis in North America and Europe: A Comparison of Co-citaion Network. ***Organization Studies***, 16: 503-526.

Weibel, A. Forthcoming. Formal Control and Trustworthiness - Shall the Twain Never Meet? ***Group and Organization Management***.

Weingart, P. 2005. Impact of Bibliometrics upon the Science System: Inadvertent Consequences? ***Scientometrics***, 62(1): 117-131.

Wissenschaftsrat (German Science Council). 2004. ***Recommendations for Rankings in the System of Higher Education and Research. Part 1: Research***. Drs. 6285-0. Hamburg, 12 November 2004.