Analyse & Kritik 02/2010 (© Lucius & Lucius, Stuttgart) S. 267–283

Margit Osterloh

Governance by Numbers. Does It Really Work in Research?

Abstract: Performance evaluation in research is more and more based on numbers of publications, citations, and impact factors. In the wake of New Public Management output control has been introduced into research governance without taking into account the conditions necessary for this kind of control to work efficiently. It is argued that to evaluate research by output control is counterproductive. It induces to substitute the 'taste for science' by a 'taste for publication'. Instead, input control by careful selection and socialization serves as an alternative.

1. Introduction

Today, governments spend more time and money on evaluation and performance measurement than ever before, mainly based on output based indicators. In the wake of New Public Management approaches such as management by objectives and pay-for-performance are transferred from private companies to public service organizations, e.g. to hospitals, schools and transport services to enforce market mechanisms and raise accountability, productivity, and efficiency within public service institutions. Also in universities a "governance by numbers" (Heintz 2008) has set in, aimed at an indirect regulation and control by quasi-market institutions and the establishment of an "enterprise university" (Clark 1998; Marginson/Considine 2000; Bok 2003; Willmott 2003; Khurana 2007; Donoghue 2008). Output indicators like citations, the number of publications in journals with a high impact factor or the amount of grants today serve as the 'currency' for scholarly performance. They are applied for four intended purposes.

First, they serve as basis of decisions about access to promotion or tenure. Second, they are used for the allocation of resources to universities. In some countries they serve (e.g. in Australia) or are intended to serve (e.g. in the UK) as a basis for determining research funds selectively. In many countries, such as Germany and Spain, salaries are linked to publications in high ranked journals. Increasingly, universities provide cash bonuses for publications in key journals, for example, in Australia, China, and Korea (Fuyuno/Cyranoski 2006; Franzoni/Scellato/Stephan 2010). Third, it is assumed that heating up the competition between scholars and paying bonuses according to the number of

publications in top-journals motivates scholars to do more research of higher quality. Fourth, some believe that output indicators give the public a transparent picture of scholarly activity and make universities more accountable for their use of public money. Academic rankings are intended to unlock the "secrets of the world of research" (Weingart 2005, 119) for journalists as well as for deans, administrators, and politicians who have no special knowledge of the field.

New Public Management intends to introduce modern management methods into public service and research institutions. However, it disregards research in management theory, in particular in management control theory and performance measurement theory. This kind of research asks, which kind of control and performance measurement system can successfully applied to which tasks. In this paper I apply management control theory and performance measurement theory to academic research. I will inquire, whether in this field the necessary conditions obtain under which output control works efficiently.

In the *next section* I give a brief introduction into management control theory dealing with the question which kinds of control exist and under which condition they are applicable. In the *following sections* I discuss the quality of the different kinds of control—process control, output control and input control—in academic research. I conclude, that though all kinds of control have advantages and disadvantages, output control or 'governance by numbers' in academic research should be applied with utmost care. Instead, input control could serve as an alternative.

2. Management Control Theory

Management Control Theory distinguishes three kinds of control, output control, process control, and input control and shows on which kinds of tasks they are applicable (Eisenhardt 1985; Kirsch 1996; Ouchi 1979; Turner/Makhija 2006). Although all organizations use a combination of these control mechanisms, two aspects are decisive for employing a certain mix of control mechanism. These aspects are: (1) the knowledge related to the measurability of outputs; and (2) the knowledge about the cause-effect relations (Thompson 1967), the transformation process (Ouchi 1979), or the appropriate rules to be applied. The relationships between control forms and knowledge available to the evaluator are summarized in Figure 1.

Output control (cell 1) is useful if well-defined unambiguous output-indicators are available and can be attributed to persons or groups so that the results can be contracted. It is not necessary to specify the processes that produce the desired output. Output control relies on desired results that can be linked to the use of incentives without knowing how the result can be achieved. Therefore, output control is applied when processes or cause-effect relationships are difficult to determine, are not understood by non-experts or are varying according to external changes that are difficult to predict. This is the reason why many journalists, politicians, administrators and other external observers are keen on

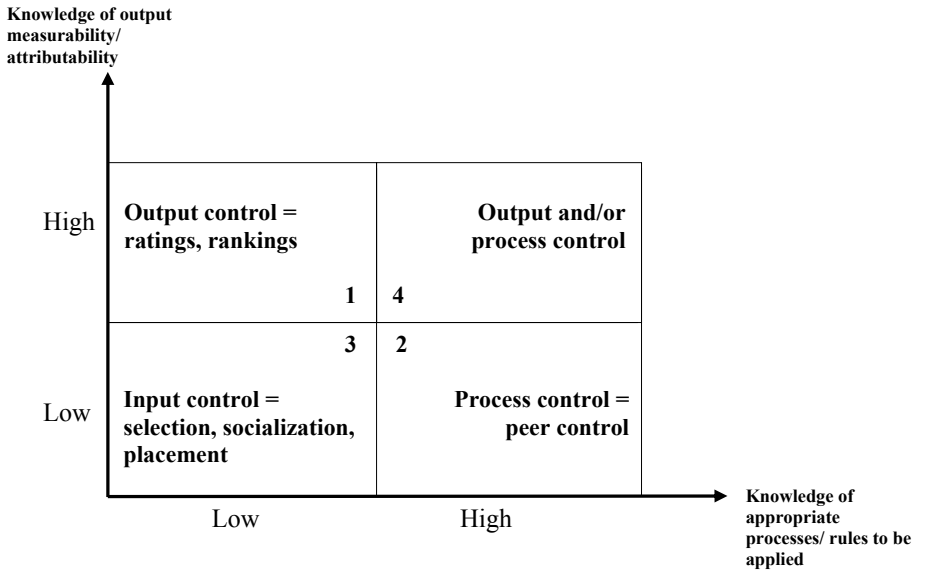

Figure 1: Control modes and task characteristics adopted from Ouchi (1979)

output control. Output indicators seem to provide easily understandable quality data. They promise to change lack of knowledge into knowledge (Heintz 2008).

However, there are some preconditions for efficient output control (e.g. Merchant/Van der Stede 2003; Snell 1992; Eisenhardt 1985) that—as I will show below—fall short in the case of research. *First*, the output indicators measure unambiguously what should be measured, namely research performance. *Second* the output indicators are clear-cut and stable. *Third*, measurement of research output motivates researches in a way so that unintended side-effects of measurement do not compensate the intended performance increases. *Fourth*, the research institution knows and controls its production function, i.e. they can allocate resources in an optimal way to produce the desired output in the future.

Process control (cell 2) might be an alternative.[1] It specifies the appropriate behaviour and processes instead of the desired outputs which often cannot be determined in advance. The preconditions are that evaluators (a) have the appropriate knowledge of cause-effect relationships or appropriate methodologies to be applied, and (b) use their knowledge without biases and malevolence. In contrast to output control, process control can only be conducted by people who are familiar with the field. It encompasses highly formalized standard operating procedures as well as scholarly methodologies that are state of the art in the respective field. Process control in the academic field takes the form of peer control. Peers are (or should be) familiar with processes and methodologies in their research. Peer control—e.g. in the form of refereeing submitted articles—

[1] Process control sometimes is called behaviour control or action control.

also supports (or should support) scholars by giving advice how to improve research.

Process control has it disadvantages too. It is dependent on knowledgeable and fair evaluators. It sometimes restricts researchers' ability to deal with the high level of uncertainty inherent in the research process. In such situations process control might meet with criticism due to its inflexibility, rigidity and sometimes malevolence and cronyism among peers.

Input or personnel control (cell 3) builds on selection, socialization and placement of individuals (Merchant/Van der Stede 2003, 75). It has to be applied when the preconditions of output and process control are not given. It intends to make sure that individuals have internalized norms and professional standards so that they follow them even when there is no output or process control feasible. Input control takes place inside professional groups. Senior colleagues function as supervisors and role models for younger co-workers in the process of selection, socialization and placement. After a candidate has passed the input control he or she becomes the member of a profession. Output and process control loses importance. Autonomy is curtailed only by professional norms that are confirmed by institutionalized rituals. According to Ouchi (1979), input control should be applied with complex and ambiguous tasks.[2] This is the case not only with researchers, but also with professions characterized by a low degree of observable outputs and processes, such as life-tenured judges (e.g., Benz/Frey 2007; Posner 2010) and executive search companies (Zehnder 2001). However, input control has a major disadvantage: It is in danger of being submitted to groupthink (Janis 1972) and cronyism.

Output and/or process control (cell 4) is not relevant 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 summarize, Management Control Theory clarifies when to apply output, process, or input control and when to avoid it.[3] In reality, there is always a mix of these types of control. Depending on the type of the task and the knowledge of the evaluator about the task an optimal mix of control types should be aimed for. In the next sections I will discuss to which extent the criteria for the different kinds of control are met in research.

3. The Quality of Process Control in Research

Process control in the form of peer control is the basis of all other kinds of control in academic research—be it output or input control. Research is characterized by a cumulative process of knowledge building, by a great uncertainty of research outcomes, and by an even greater uncertainty whether the research results will be ever marketable. In addition, research often produces serendipity effects; that is, it provides answers to unasked questions (Stephan 1996; Simon-

[2] Ouchi (1979) calls this kind of control "clan control".

[3] Many authors (Eisenhardt 1985; Kirsch 1996; Ouchi 1979; Turner/Makhija 2006) provide empirical support for Management Control Theory.

ton 2004). Therefore in academia evaluation by the market has to be replaced by peer evaluation to determine whether a piece of research represents an advance. Only peers are able to evaluate on the one hand the degree of novelty of a research result and on the other hand who earns the reputation to be the discoverer according to the "priority rule" (Merton 1973; Dasgupta/David 1994). The decision about priority often is a matter of controversial scholarly discussion (Merton 1961).

The preconditions of a high quality process control are that peers are well informed about the state of the art in their discipline, and have an unbiased judgement. However, as the recent heated discussion about the quality of peer reviews show, there are serious problems. (e.g. Frey 2003; Bedeian 2004; Starbuck 2005; 2006; Tsang/Frey 2007; Gillies 2005; 2008; Abramo/Angelo/Caprasecca 2009; Bornmann/Daniel 2009).[4]

First, the extent to which reviewing reports conform to each other is low. The correlation between the judgments of two peers falls between 0.09 and 0.5 (Starbuck 2005). In clinical neuroscience, it was found that the correlations among reviewers' recommendations "was little greater than would be expected by chance alone" (Rothwell/Martyn 2000, 1964). It is important that the correlation is higher for papers rejected than for papers accepted (Cichetti 1991). This means that peer reviewers are better able to identify academic low performers than excellent research (Moed 2007).

Second the correlation of a particular reviewer's evaluation with the actual quality as measured by later citations of the manuscript reviewed is also quite low between 0.25 and 0.30 (Starbuck 2006, 83–84).

Third, many rejections of papers in highly ranked journals are documented that later were awarded high prizes, including the Nobel Prize (Gans/Shepherd 1994; Campanario 1996; Lawrence 2003).

Fourth, reviewers find methodological shortcomings in 71 percent of papers contradicting the mainstream, compared to only 25 percent of papers supporting the mainstream (Mahoney 1977).

Fifth, in the process of grant giving, reviewers often are biased by the reputation of the potential recipient of the grant. Much attention was given to a gender bias with respect to Swedish female researchers (Wenneras/Wold 1997).

Sixth, sometimes editors are prone to severe errors. A famous example is the 'Social Text'-Affair: The physicist Alain D. Sokal published an article in a (non-refereed) special issue of the journal *Social Text*, which was written as a parody. The editors did not realize that the bogus article was a hoax (Sokal 1996).

Seventh, it is sometimes not in the interest of rational and selfish reviewers to accept a piece of research or to give advice how to improve it. Reviewers might reject papers that threaten their previous work (Lawrence 2003) or draw attention to competing ideas. In a simulation it was shown that unless the fractions of such rational reviewers (as well as the fraction of unreliable, uniformed

[4] See also the special issue of *Science and Public Policy* 2007 and the Special Theme Section on "The use and misuse of bibliometric indices in evaluating scholarly performance" of *Ethics in Science and Environmental Politics* 2008.

reviewers) are kept well below 30 percent, peer review will not perform better than throwing a coin (Thurner/Hanel 2010).

As a consequence, the quality and credibility of process control in research is far from being unquestioned. In particular in situations of radical innovations or paradigm shifts (Kuhn 1962) process control might hinder instead of support path breaking research.

Although the shortcomings of peer reviews are substantial, they are counterbalanced by the heterogeneity of scientific views produced. If rejected by the reviewers of one journal, an article often is accepted by the reviewers of an equivalent journal; an unsuccessful application to one university may be overcome by applying to another university. Such heterogeneity is an essential feature of scholarly endeavors. It is of utmost importance as long as groupthink and cronyism does not lead to an undue dominance of a single view. The overall effectiveness of the decentralized evaluation process by peer review often is overlooked by critics of the peer review process. However, for the public as well as for politicians and university administrators, such a controversial scholarly communication process is not easy to comprehend. They prefer an easy to understand single metric in the form of rankings based on the number of publications, citations, and impact factors.[5] In addition, it is expected that some of the problems of peer reviews can be avoided.

4. The Quality of Output Control in Research

Output control in research measures citations, the amount of publications in refereed journals or the amount of grants. These numbers are the basis of rankings or ratings. However, output control in research is based on process control in the form of peer evaluations. Process control by peers is turned into output control by disembedding it from its multifaceted context that peers consider, by crystallizing it into indicators and by aggregating them into numbers and rankings that are understandable by non-experts (Heintz 2008). In order to estimate to which extent these numbers are able to measure research quality and to give non-experts a transparent picture of scholarly activity I will proceed in three steps. First, I consider the advantages of rankings compared to peer reviews on which they are based. Second, I will ask to which extent the aggregation of peer judgments in the form of rankings is correct. Third, I will discuss behavioural effects of output control as they are discussed in the performance measurement literature and I will apply these insights to the measurement of research performance. Fourth, I will ask to which extent output measures could be used to determine a strategy to produce a desired output in the future. The aim is to examine whether output control in research meets the four conditions mentioned for this kind of control, namely first to be a precise measure of performance, sec-

[5] For example, the British Government decided to replace its Research Assessment Exercise based mainly on qualitative evaluations with a system based mainly on bibliometrics. Interestingly, the Australian Government, which used mostly bibliometrics in the past, in the future plans to strengthen qualitative peer review methods (Donovan 2007).

ond, to be clear cut and stable, third, to motivate people in an efficient way and fourth to help allocate resources efficiently.

Which are the advantages of output control compared to process control in research?

Although in research rankings are based on process control in the form of peer evaluation they have several advantages that help explain why they have become so popular in the last years (e.g. Abramo et al. 2009).

First, rankings are based on more than the three or four evaluations typical for qualitative approaches. Through statistical aggregation, individual reviewers' biases may be balanced. Second, the influence of the old boys' network may be avoided. An instrument is provided to dismantle unfounded claims to fame (Khurana 2007, 337). Third rankings facilitate the comparison between a large numbers of scholars or institutions. They admit updates and rapid intertemporal comparisons. Fourth they are intended to give research administrators, politicians, journalists, and students an easy to use device to evaluate the standing of the research.

These advantages hold only, if the process of aggregation of peer review judgments is correct and the picture of research performance given to the public is adequate.

Does the aggregation of peer judgments provide a correct picture of research performance?

In recent times, it became clear that rankings might counterbalance some problems of qualitative peer reviews but that they have disadvantages of their own (Butler 2007; Donovan 2007; Adler/Ewing/Taylor 2008; Adler/Harzing 2009).

There exist *technical problems* consisting in errors in the citing-cited matching process. This leads to a loss of citations to a specific publication. Moreover errors are made in attributing publications and citations to the source, for example, institutes, departments, or universities (Moed 2002). It has been shown that small changes in measurement techniques and classifications can have large effects on the position in rankings (Ursprung/Zimmer 2006; Frey/Rost 2010).

Methodological problems of constructing meaningful indices to measure scientific output consist firstly in selection problems (e.g. only journals are selected) (Frey 2003; 2009; Adler et al. 2008; Adler/Harzing 2009). Secondly, citations often only promote visibility and a "Matthew effect" in science (Merton 1968) rather than to disseminate knowledge. On the basis of analyses of misprints it has been found that most papers that are cited are not read (Simkin/Roychowdhury 2005). Thirdly, using the impact factor of a journal as a proxy for the quality of a single article leads to substantial misclassifications (Laband/Tollison 2003; Oswald 2007; Adler et al. 2008). Fourth, there are difficulties comparing citations and impact factors between disciplines and even between sub-disciplines (Bornman/Mutz/Neuhaus/Daniel 2008).

Technical and methodological problems can be mitigated. What cannot be mitigated is the problem of *one-dimensionality* of all kinds rankings (Fase 2007). They press the multifacetness and ambiguity of scholarly endeavours into a simple one-dimensional order that leads the public astray. Accountability to the public should encompass the effort to make clear that there are fundamental differences between rankings e.g. in a football league and scholarly work. A one-dimensional approach contradicts the idea of research as "institutionalized scepticism" (Merton 1973) that builds on controversial scholarly disputes. Moreover, one-dimensional rankings motivate to invest in the first place in improving ones position in the rankings rather than in improving the performance that is intended.

In summary, the first precondition of output control, namely that output indicators measure unambiguously what should be measured, is far from being fulfilled.

Which are the behavioral reactions to output control?

The discussion of the behavioral reactions of output control deals with two questions that are interdependent. First, are output indicators clear-cut and stable? Second, does output control motivate researchers in the right way so that unintended side effects do not compensate the intended performance increases?

In management performance measurement theory there is some debate about the so called 'performance paradox' (Meyer/Gupta 1994; Meyer 2005). It is caused by the tendency of performance measures to lose variance with use, in particular with complex tasks. As a consequence, performance indicators lose the ability to discriminate good from bad performance. The relationship between reported and actual performance declines. The performance paradox is the stronger the fewer the number of performance indicators and the more complex the task is. It follows that performance indicators that really measure performance are not clear-cut and stable, but consist of ever changing measures. Why is this the case?

The 'performance paradox' is caused by two effects which in reality cannot be differentiated. First, performance may be improved since the feedback by performance indicators causes positive learning as well as a selection and self-selection of bad performers. Nevertheless such an intended effect of performance measurement causes the tendency that performance indicators lose variance with use (e.g. grades in schools). Second, it could as well be the case, that performance measurement causes an unintended perverse learning. This is the case when people focus on indicators but not on the performance that it sought (e.g. teaching to the test). It is often hard to determine, whether a positive or a perverse learning has taken place. In any case, a permanent adaption of performance indicators is necessary that undermines the precondition of output indicators to be clear-cut and stable. There are numerous examples for such a 'performance paradox' not only in command economies but also in market economies. Examples are the shift of dominant performance criteria for incentive compensation in companies. During the 1960s market share dominated. During the 1970s return

on investment (ROI) became important, during the 1990s the criteria changed to short-time shareholder value (Meyer/Gupta 1994). After the recent financial market crisis sustainable shareholder value has become prominent.

In research such shifts have occurred from counting publications to counting publications in refereed journals. Later publications in journals according to their impact factors or citations became important. New indicators like the H-index have been invented. There is no doubt, that we can expect further new indicators in the future as soon as the old ones lose discriminatory power by strategic behavior. This will be the case e.g. by increasing the amount of new journals, or by editors requesting authors to cite their journals in order to raise their impact factor. A lock-in effect sets in which forces scholars and institutions to adapt their behavior even if they are aware of the deficiencies of such a development.

The background of such perverse learning is strategic behavior. It is well studied and has been called "hitting the target and missing the point", "tunnel vision", "goal-displacement" (Merton 1940; Perrin 1998), "reactivity" (Espeland/Sauder 2007), or "multiple-tasking effect" (Ethiray/Levinthal 2009; Holmstrom/Milgrom 1991; Kerr 1975). There is much empirical evidence of this behavior (Ordonez/Schweitzer/Galinsky/Bazerman 2009;[6] Fehr/Schmidt 2004). One step further goes "cream skimming" or "cherry picking" (van Thiel/Leeuw 2002) by "gaming the system". Examples are chronically ill patients excluded in healthcare, teachers responding to evaluations by excluding bad pupils from tests (for empirical evidence in the U.S., see Figlio/Getzler 2002) or putting lower quality students in special classes that are not included in the measurement sample (Gioia/Corley 2002).

In academia, examples of such strategic behavior can be found, for example, as 'slicing strategy', whereby scholars divide their research into as many papers as possible to increase their publication list (Butler 2003). Another example of goal displacement is the lowering of standards for PhD candidates when the amount of completed PhDs is used as a measure in rankings.

One step further go counterstrategies that consist in altering research behavior itself. Examples are scholars that distort their results to please, or at least not to oppose, prospective referees. Bedeian (2003) finds evidence that no less than 25 percent of authors revise their manuscripts according to the suggestions of the referee although they know that the change is incorrect. Frey (2003) calls this behavior "academic prostitution". Authors cite possible reviewers because the latter are prone to judge papers more favorably that approvingly cite their work. To meet the expectations of their peers—many of whom consist of mainstream scholars—authors may be discouraged from conducting and submitting creative and unorthodox research (Horrobin 1996; Prichard/Willmott 1997; Armstrong 1997; Gillies 2008).

The negative effects of the 'performance paradox' are enforced if a second kind of unintended consequences takes place, the decrease of intrinsically moti-

[6] Locke and Latham (2009) in a rejoinder provide counterevidence to Ordonez et al. 2009. However, they disregard that output control might well work for simple but not for complex tasks within an organization.

vated curiosity which generally is acknowledged to be of decisive importance in academic research (Amabile 1996; 1998; Spangenberg et al. 1990; Stephan 1996). In psychology and psychological economics, there exists considerable empirical evidence that there is a crowding-out effect of intrinsic motivation by externally imposed goals linked to incentives that do not give a supportive feedback and are perceived to be controlling[7] (Hennessey/Amabile 1998; Frey 1992; 1997; Gagné/Deci 2005; Falk/Kosfeld 2006; Ordonez et al. 2009).

From that point of view, output control tends to crowd out intrinsically motivated curiosity. First, in contrast to process control, rankings do not give a supportive feedback as they do not tell scholars how to improve their research. Second, because rankings are mostly imposed from outside, the content of research is in danger of losing importance. The "taste for science" (Merton 1973; Dasgupta/David 1994) is substituted by a 'taste for publication'. As a consequence, the dysfunctional reactions of scholars (e.g., goal displacement and counterstrategies) are enforced because they are not constrained by intrinsic preferences. The inducement to 'game the system' in an instrumental way may get the upper hand.

To sum up the answers to the question about the behavioral reactions to output control: It has been shown that also the second and the third precondition of efficient output control in research has to be put into question. That is the existence of clear-cut and stable output criteria, and the promise that researchers are motivated in a way that possible performance increases exceed unintended side effects.

Can output measurements be used as basis for efficient resource allocation?

As mentioned, output indicators like academic rankings are used to allocate resources be it in the form of budgeting or of competitive grants. It is overlooked that it is problematic to derive forward oriented strategies from backward oriented data. The actors and organizations do not know and control their production function that transforms efforts into results in the future. In particular also in research there is a decreasing marginal effect of additional research resources (Jansen et al. 2007). This means, that providing more resources to high-performing researchers or research groups according to their past output might create inefficient 'research empires' due to a decreasing marginal productivity.

There are numerous approaches to combine backward oriented performance indicators with forward oriented strategic planning (e.g. Schreyögg/Steinmann 1987; Simons 1995; Meyer 2009). They differ in some respect. But they conform with the following ideas. First, backward oriented numbers are an important, but by far not the only basis for strategic resource allocation. Second, if such numbers are used, they should serve to foster exploration and learning. They should inspire discussion and mutual consultation. For this purpose one should abstain to use them as pay-for-performance incentives. Third, there must be an ongoing control of premises of strategic planning. This has to be done in an in-

[7] A third precondition is social relatedness, see Gagne/Deci 2005.

teractive way in which the researchers themselves (and not only administrators) should have a say.

To summarize the findings about the four preconditions that must be fulfilled for an efficient output control: I conclude that this kind of control should be applied in research with utmost care. It should be used only either in an exploratory way or as a complement to peer reviews in the form of so-called 'informed peer review'. However, this constrains the use of output measures as a handy instrument for non-experts so assess scholarly performance.

5. The Quality of Input Control in Research

If neither output control nor process control work sufficiently well, then *input control* has to be applied (Ouchi 1979). The aim is to make candidates members of a community in which aligned norms and values are internalized and are part of their intrinsic motivation. If input control is successful, mutual tolerance for ambiguity is possible, which is important when output and process control is questionable.

What does input control mean in the case of research governance? Aspiring scholars should be carefully socialized and selected by peers to prove that they have mastered the state of the art, have preferences according to the "taste for science" (Merton 1973), and are able to direct themselves. Those passing a rigorous input control should be given much autonomy to foster their creativity and intrinsic motivated curiosity. This includes the provision of basic funds to provide a certain degree of independence after having passed the entrance barriers (Gillies 2008; Horrobin 1996).

Input control is part of the *Principles Governing Research at Harvard*, which states: "The primary means for controlling the quality of scholarly activities of this Faculty is through the rigorous academic standards applied in selecting its members."[8] Input control has empirically proven to be successful also in R&D organizations of industrial companies (Abernethy/Brownell 1997). This is in accordance with empirical findings in psychological economics. They show that on average intrinsically motivated people do not shirk when they are given autonomy (Frey 1992; Gneezy/Rustichini 2000; Fong/Tosi 2007). Instead, they raise their efforts when they perceive that they are trusted (Falk/Kosfeld 2006; Osterloh/Frey 2000; Frost/Osterloh/Weibel 2010).

Input control has advantages and disadvantages. The *advantages* first consist in downplaying the unfortunate 'performance paradox' while inducing young scholars to learn the professional standards of their discipline under the assistance of peers. Second, although input control still requires process control in the form of peer evaluations, this applies during limited time periods, namely during situations of status passage. Third, input control is a decentralized form of peer evaluation, for example, when submitting papers or applying for jobs. It supports the heterogeneity of scholarly views central to the scientific communi-

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

cation process. Fourth, input control is a kind of 'informed peer review' that is able to use output indicators in an exploratory way.

The *disadvantages* consist first in the danger that some scholars who have passed the selection might misuse their autonomy, reduce their work effort, and waste their funds. This disadvantage will be lower when the selection process is conducted rigorously.

Second, input control is in danger of being submitted to groupthink (Janis 1972) and cronyism. This danger can be mitigated by fostering the diversity of scholarly approaches within the relevant peer group.

Third, the public as well as university administrators do not get an easy to comprehend picture of scholarly activities as it is intended with output control based on rankings. People outside the scholarly community have to accept that to evaluate scholarly activities there is no easy to understand criteria comparable to rankings of football leagues. As a consequence, universities leaders like presidents, vice chancellors and deans should consist of accomplished scholars. In contrast to pure managers top scholars have a better understanding of the research process. Goodall (2009) shows for a panel of 55 research universities that a universityt's research performance is improved after an accomplished scholar has been hired as president.

To compensate for the disadvantages of input control, periodic self-evaluation including external evaluators should be applied. The major goal is to induce self-reflection and feedback among the members of a research unit.

6. Conclusion

'Governance by numbers' as applied by New Public Management to research activities comes at a high cost. Though all kinds of control have advantages and disadvantages, I have argued that in research input control is the most adequate form of control. Input control includes elements of output control as well as of process control during a thorough selection and socialization process conducted in the form of 'informed peer reviews'. The disadvantages of output and process control mentioned with input control are limited for several reasons. First, input control takes place in few situations only. Second, input control still leaves open much heterogeneity of different scholarly views. Third, it helps scholars internalize professional norms and standards during their socialization process to a high degree, so that they can be granted much autonomy and are able to participate in the highly controversial and ambiguous debate that is at the heart of scholarly activity. The price to pay is that the public does not get an easy to comprehend picture of scholarly activity. But it should be part of the accountability of research and research administration to the public to communicate that scholarly work has to be evaluated in a different way than football games or hit parades.

Bibliography

Abernethy, M. A./P. Brownell (1997), Management Control Systems in Research and Development Organizations: The Role of Accounting, Behavior and Personnel Controls, in: *Accouning, Organizations and Society* 22(3/4), 233–248

Adler, N. J./A.-W. Harzing (2009), When Knowledge Wins: Transcending the Sense and Nonsense of Academic Rankings, in: *Academy of Management Learning* 8, 72–95

Adler, R./J. Ewing/P. Taylor (2008), *Citation Statistics, a Report from the Joint Committee on Quantitative Assessment of Research* (IMU, ICIAM, IMS). Report from the International Mathematical Union (IMU) in cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics (IMS)

Abramo, G./C. A. D'Angelo/A. Caprasecca (2009), Allocative Efficiency in Public Research Funding: Can Bibliometrics Help?, in: *Research Policy* 38, 206–215

Amabile, T. M. (1996), *Creativity in Context: Update to the Social Psychology of Creativity*, Boulder

— (1998), How to Kill Creativity, in: *Harvard Business Review* 76, 76–87

Armstrong, J. S. (1997), Peer Review for Journals: Evidence on Quality Control, Fairness, and Innovation, in: *Science and Engineering Ethics* 3, 63–84

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

— (2004), Peer Review and the Social Construction of Knowledge in the Management Discipline, in: *Academy of Management Learning and Education* 3, 198–216

Benz, M./B. S. Frey (2007), Corporate Governance: What Can We Learn from Public Governance?, in: *Academy of Management Review* 32(1), 92–104

Bok, D. (2003), *Universities in the Marketplace. The Commercialization of Higher Education*, Princeton

Bornmann, L./H.-D. Daniel (2009), The Luck of the Referee Draw: The Effect of Exchanging Reviews, in: *Learned Publishing* 22(2), 117–125

Butler, L. (2003), Explaining Australia's Increased Share of ISI Publications—The Effects of a Funding Formula Based on Publication Counts, in: *Research Policy* 32, 143–155

Campanario, J. M. (1996), Using Citation Classics to Study the Incidence of Serendipity in Scientific Discovery, in: *Scientometrics* 37, 3–24

Cicchetti, D. V. (1991), The Reliability of Peer Review for Manuscript and Grant Submissions: A Cross-disciplinary Investigation, in: *Behavioral and Brain Sciences* 14, 119–135

Clark, B. R. (1998), *Creating Entrepreneurial Universities. Organizational Pathways of Transformation*, Surrey

Dasgupta, P./P. A. David (1994), Toward a New Economics of Science, in: *Research Policy* 23, 487–521

Donoghue, F. (2008), *The Last Professors. The Corporate University and the Fate of the Humanities*, New York

Donovan, C. (2007), The Qualitative Future of Research Evaluation, in: *Science and Public Policy* 34, 585–597

Eisenhardt, K. M. (1985), Control: Organizational and Economic Approaches, in: *Management Science* 31(2), 134–149

Espeland, W. N./M. Sauder (2007), Rankings and Reactivity: How Public Measures Recreate Social Worlds, in: *American Journal of Sociology* 113(1), 1–40

Ethiraj, S. K./D. Levinthal (2009), Hoping for A to Z while Rewarding Only A: Complex Organizations and Multiple Goals, in: *Organization Science* 20, 4–21

Falk, A./M. Kosfeld (2006), The Hidden Cost of Control, in: *American Economic Review* 96, 1611–1630

Fase, M. M. G. (2007), For Examples of a Trompe-l'Oeil in Economics, in: *De Economist* 155(2), 221–238

Fehr, E./K. M. Schmidt (2004), Fairness and Incentives in a Multi-task Principal-agent Model, in: *Scandinavian Journal of Economics* 106, 453–474

Figlio, D. N./L. S. Getzler (2002), Accountability, Ability and Disability: Gaming the System, *NBER Working Paper No. W9307*, URL: http://bear.warrington.ufl.edu/figlio/w9307.pdf

Fong, E. A./H. L. Tosi Jr. (2007), Effort, Performance, and Conscientiousness: An Agency Theory Perspective, in: *Journal of Management* 33, 161–179

Franzoni, C./G. Scellato/P. Stephan (2010), Changing Incentives to Publish and the Consequences for Submission Patterns, *Working paper*, IP Finance Institute

Frey, B. S. (1992), Tertium Datur: Pricing, Regulating and Intrinsic Motivation, in: *Kyklos* 45, 161–185

— (1997), *Not Just for the Money: An Economic Theory of Personal Motivation*, Cheltenham

— (2003), Publishing as Prostitution?—Choosing between One's Own Ideas and Academic Success, in: *Public Choice* 116, 205–223

— (2009), Economists in the PITS, in: *International Review of Economics* 56(4), 335–346

— /K. Rost (2010), Do Rankings Reflect Research Quality?, in: *Journal of Applied Economics* 13(1), 1–38

Frost, J./M. Osterloh/A. Weibel (2010), Governing Knowledge Work: Transactional and Transformational Solutions, in: *Organizational Dynamics* 39, 126–136

Fuyuno, I./D. Cyranoski (2006), Cash for Papers: Putting a Premium on Publication, in: *Nature* 441, 792

Gagné, M./E. L. Deci (2005), Self-determination Theory and Work Motivation, in: *Journal of Organizational Behavior* 26, 331–362

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

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

— (2008), *How Should Research Be Organised?*, London

Gioia, D. A./K. G. Corley (2002), Being Good versus Looking Good: Business School Rankings and the Circean Transformation from Substance to Image, in: *Academy of Management Learning and Education* 1, 107–120

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

Goodall, A. H. (2009), Highly Cited Leaders and the Performance of Research Universities, in: *Research Policy* 38, 1070–1092

Heintz, B. (2008), Governance by Numbers. Zum Zusammenhang von Quantifizierung und Globalisierung am Beispiel der Hochschulpolitik, in: Schuppert, G. F./A. Voßkuhl (eds.), *Governance von und durch Wissen*, Baden-Baden, 110–304

Hennessey, B. A./T. M. Amabile (1998), Reward, Intrinsic Motivation and Creativity, in: *American Psychologist* 53(6), 647–675

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

Horrobin, D. F. (1996), Peer Review of Grant Applications: A Harbinger for Mediocrity in Clinical Research?, in: *Lancet* 348, 1293–1295

Janis, I. L. (1972), *Victims of Groupthink: A Psychological Study of Foreign-policy Decisions and Fiascos*, Boston

Jansen, D./A. Wald/K. Franke/U. Schmoch/T. Schubert (2007), Drittmittel als Performanzindikator der wissenschaftlichen Forschung. Zum Einfluss von Rahmenbedingungen auf Forschungsleistung, in: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 59, 125–149

Kerr, S. (1975), On the Folly of Rewarding A While Hoping for B, in: *Academy of Management Journal* 18(4), 769–783

Kirsch, L. J. (1996), The Management of Complex Tasks in Organizations: Controlling the Systems Development Process, in: *Organization Science* 7(1), 1–21

Khurana, R. (2007), *From Higher Aims to Hired Hands: The Social Transformation of Business Schools and the Unfulfilled Promise of Management as a Profession*, Princeton

Kuhn, T. S. (1962), *The Structure of Scientific Revolutions*, Chicago

Laband, D. N./R. D. Tollison (2003), Dry Holes in Economic Research, in: *Kyklos* 56, 161–174

Lawrence, P. A. (2003), The Politics of Publication—Authors, Reviewers, and Editors Must Act to Protect the Quality of Research, in: *Nature* 422, 259–261

Locke, E. A./G. P. Latham (2009), Has Goal Setting Gone Wild, or Have Its Attackers Abandoned Good Scholarship?, in: *Academy of Management Perspectives* 21, 17–23

Marginson, S./M. Considine (2000), *The Enterprise University: Power, Governance and Reinvention in Australia*, Cambridge/MA

Mahoney, M. J. (1977), Publication Prejudices: An Experimental Study of Confirmatory Bias in the Peer Review System, in: *Cognitive Therapy Research* 1, 161–175

Merton, R. K (1940), Bureaucratic Structure and Personality, in: *Social Forces* 18, 560–568

— (1961), Singletons and Multiples in Scientific Discovery: A Chapter in the Sociology of Science, in: *Proceedings of the American Philosophical Society* 105(5), 470–486

— (1968), The Matthew Effect in Science, in: *Science* 159, 56–63

— (1973), *The Sociology of Science: Theoretical and Empirical Investigation*, Chicago

Merchant, K./W. Van der Stede (2003), *Management Control Systems. Performance Measurement, Evaluation and Incentives*, Prentice Hall

Meyer, M. W. (2005), Can Performance Studies Create Actionable Knowledge If We Can't Measure the Performance of the Firm?, in: *Journal of Management Inquiry* 14(3), 287–291

— (2009), *Rethinking Performance Management. Beyond the Balanced Scorecard*, Cambridge

Meyer, M. W./V. Gupta (1994), The Performance Paradox, in: *Research in Organizational Behavior* 16, 309–369

Moed, H. F. (2002), The Impact Factors Debate; The ISI's Uses and Limits, in: *Nature* 415, 731–732

— (2007), The Future of Research Evaluation Rests with an Intelligent Combination of Advanced Metrics and Transparent Peer Review, in: *Science and Public Policy* 34, 575–583

Ordonez, L. D./M. E. Schweitzer/A. D. Galinsky/M. H. Bazerman (2009), Goals Gone Wild: The Systematic Side Effects of Overprescribing Goal Setting, in: *Academy of Management Perspectives* 23, 6–16

Ouchi, W. G. (1979), A Conceptual Framework for the Design of Organizational Control Mechanisms, in: *Management Science* 25, 833–848

Osterloh, M./B. S. Frey (2000), Motivation, Knowledge Transfer, and Organizational Forms, in: *Organization Science* 11, 538–550

Oswald, A. J. (2007), An Examination of the Reliability of Prestigious Scholarly Journals: Evidence and Implications for Decision-makers, in: *Economica* 74, 21–31

Perrin, B. (1998), Effective Use and Misuse of Performance Measurement, in: *American Journal of Evaluation* 19, 367–379

Posner, R. A. (2010), From the New Institutional Economics to Organization Economics: With Applications to Corporate Governance, Government Agencies, and Legal Institutions, in: *Journal of Institutional Economics* 6(1), 1–37

Prichard, C./H. Willmott (1997), Just How Managed is the McUniversity?, in: *Organization Studies* 18, 287–316

Rothwell, P. M./N. Martyn (2000), Reproducibility of Peer Review in Clinical Neuroscience. Is Agreement between Reviewers any Greater Than Would Be Expected by Chance Alone?, in: *Brain* 123, 1964–1969

Schreyögg, G./H. Steinmann (1987), Strategic Control: A New Perspective, in: *Academy of Management Review* 12, 91–103

Simkin, M. V./V. P. Roychowdhury (2005), Copied Citations Create Renowned Papers?, in: *Annals Improb. Res* 11(1), 24–27

Simons, R. (1995), Control in an Age of Empowerment, in: *Harvard Business Review* 73(2), 80–88

Simonton, D. K. (2004), *Creativity in Science. Chance, Logic, Genius, and Zeitgeist*, Cambridge

Snell, S. A. (1992), Control Theory in Strategic Human Resource Management—The Mediating Effect of Administrative Information, in: *Academy of Management Journal* 35(2), 292–327

Sokal, A. D. (1996), *A Physicist Experiments with Cultural Studies*, URL: http://www.physics.nyu.edu/faculty/sokal/lingua_franca_v4/lingua_franca_v4.html

Spangenberg, J. F. A./R. Starmans/Y. W. Bally/B. Breemhaar/F. J. N. Nijhuis (1990), Prediction of Scientific Performance in Clinical Medicine, in: *Research Policy* 19, 239–255

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

— (2006), *The Production of Knowledge. The Challenge of Social Science Research*, Cambridge

Stephan, P. E. (1996), The Economics of Science, in: *Journal of Economic Literature* 34, 1199–1235

Thompson, J. D. (1967), *Organizations in Action: Social Science Bases of Administrative Theory*, New York

Thurner, S./R. Hanel (2010), Peer-review in a World with Rational Scientists: Towards Selection of the Average, arXiv:1008.432v1 (*physics.soc-ph*)

Tsang, E. W. K./B. S. Frey (2007), The as-is Journal Review Process: Let Authors Own Their Ideas, in: *Academy of Management Learning and Education* 6, 128–136

Turner, K. L./M. V. Makhija (2006), The Role of Organizational Controls in Managing Knowledge, in: *Academy of Management Review* 31, 197–217

Ursprung, H. W./M. Zimmer (2006), Who Is the 'Platz-Hirsch' of the German Economics Profession? A Citation Analysis, in: *Jahrbücher für Nationalökonomie und Statistik* 227, 187–202

Van Thiel, S./F. L. Leeuw (2002), The Performance Paradox in the Public Sector, in: *Public Performance & Management Review* 25(3), 267–281

Weingart, P. (2005), Impact of Bibliometrics upon the Science System: Inadverted Consequences?, in: *Scientometrics* 62, 117–131

Wenneras, C./A. Wold (1999), Bias in Peer Review of Research Proposals in Peer Reviews in Health Sciences, in: Godlee, F./T. Jefferson (eds.), *Peer Review in Health Sciences*, London, 79–89

Willmott, H. (2003), Commercializing Higher Education in the UK: The State, Industry and Peer Review, in: *Studies in Higher Ecucation* 28(2), 129–141

Zehnder, E. (2001), A Simpler Way to Pay, in: *Harvard Business Review* April 2001, 53–60