

Academic Rankings
between
the “Republic of Science” and “New Public Management”

by

Margit Osterloh^{a,c}, and Bruno S. Frey^{b,c}

^aUniversity of Zurich

Institute of Organization and Administrative Science

^bUniversity of Zurich

Institute for Empirical Research in Economics

^cUniversity of Warwick, Warwick Business School

and

CREMA - Center for Research in Management, Economics and the Arts, Zurich

Abstract

Academic rankings mediate between the differing goals of the “republic of science” and of “new public management”. Nevertheless, rankings recently have come under scrutiny. We discuss the advantages and disadvantages of academic rankings, in particular, their unintended negative consequences on the research process, and conclude that academic rankings do not fulfil their intended goal. We discuss input control, that is, rigorous selection and socialization, supplemented by periodic self-evaluation and awards as a way to mitigate the tensions between the concept of the “republic of science” and “new public management”.

(89 words)

Keywords: peer reviews, rankings, research governance, psychological economics, new public management, economics of science, control theory

JEL Classification No: H52, H83, I23, J44, L38

In the current environment, the academic rankings of individuals or institutions, which are usually based on metrics like the number of publications and citations, are considered the backbone of research governance in academia. They serve as a basis for assessing the performance and impact of scholars, faculties, and universities for two purposes.

First, they are widely used for the allocation of resources to universities. Therefore they affect decisions on hiring, tenure, and salary of scholars. In many countries, such as Germany and Spain, recent reforms have linked access to tenure, promotion, and a higher salary more closely to publications in international journals. In some countries, universities provide cash bonuses for publications in key journals, for example, Australia, China, and Korea (Fuyuno & Cyranoski, 2006; Franzoni, Scellato, & Stephan, 2010) in order to raise their position in rankings.

Second, some believe that academic rankings 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, p. 119) for journalists as well as for deans, administrators, and politicians who have no special knowledge of the field.

However, in recent times, academic rankings have come under scrutiny (Butler, 2007; Donovan, 2007; Adler & Harzing, 2009; Albers, 2009; Van Noorden, 2010). This discussion focuses mainly on methods and on how to improve them. It is taken for granted that more and better indicators are needed to enhance the quality of rankings (Starbuck, 2009; Lane, 2010), and it is less important whether the advantage of controlling research activities from outside can produce unintended negative side effects, even if the indicators for research quality were perfect. In fact, nearly no discussion exists on whether there are viable alternatives to academic rankings as an instrument for academic governance (for an exception, see Gillies, 2008).

In this article, we discuss four issues. First, we analyze the theoretical basis of the present research governance, namely, on the one side “new public management” and on the other side the concept of the “republic of science.” Second, we compare advantages and disadvantages of peer reviews and rankings on the background of empirically based findings. Third, we discuss an aspect often disregarded, namely, the behavioral reactions to rankings that may overcompensate their advantages. Fourth, we ask whether there exist alternatives to rankings as the dominant instruments of research governance and we draw policy conclusions.

Conceptual Issues: New Public Management versus Republic of Science

Over the past years, universities have increasingly adopted the idea of new public management; namely, the idea that universities, like other public services, such as hospitals, schools, or public transport, should be subjected to a similar governance as for-profit enterprises. “More market” and “strong leadership” have become the key words (Schimank, 2005). This is reflected in procedures transferred from private companies such as management by objectives or pay-for-performance for scholars. Overall, the reforms are aimed at the establishment of an “enterprise university” (Clark, 1998; Marginson & Considine, 2000; Bok, 2003; Willmott, 2003; Khurana, 2007; Donoghue, 2008). A number of processes have been identified as the drivers behind this development (Bleiklie & Kogan, 2007; Schimank, 2005).

First, the rise of mass education during the 1980s and 1990s made higher education more expensive and visible to the public. This fact contributed to pressure for efficiency and accountability by taxpayers.¹ Second, there has been criticism about the traditional system of self-governance in universities, which may have impeded the necessary reforms toward mass education. New public management was seen as a way of breaking the “reform blockade.” Third, a growing demand for the relevance of research became influential in the public debate. In their book “The New Production of Science,” Gibbons et al. (1994) claimed that science has been transformed from a traditional university- and discipline-centered “Mode 1” knowledge production to a so-called transdisciplinary “Mode 2” knowledge production in which stakeholders from outside the university are involved. Therefore, academic peers alone no longer determine the criteria of quality. Research comes under pressure to legitimize its outcomes to people outside academia. Fourth, “economics has won the battle for theoretical hegemony in academia and society as a whole” (Ferraro, Pfeffer, & Sutton, 2005, p. 10). As a consequence, standard economics, in particular the principal agent view, has gained

¹ For an overview of the transition to mass higher education in various countries, see Teichler (1988).

dominance not only in corporate governance (Daily, Dalton, & Cannella, 2003), but also in public and academic governance.

According to standard economics, scholars have to be monitored and sanctioned in the same way as other employees. The underlying assumption is that control and correctly administered pay-for-performance schemes positively impact motivation and lead to an efficient allocation of resources (Propper, 2006). Taken together, the ideals about the governance of universities have changed from a “republic of scholars” to a “stakeholder organization” in which the voice of scholars is but one among several stakeholders and professorial autonomy is curtailed (Willmott, 2003; Bleiklie, & Kogan, 2007; Speckbacher, Wentges, & Bischof, 2008).

At first glance, this view stands in stark contrast to the ideal of self-governance of the scientific community.² This ideal was undisputed for a long time. Over three hundred years ago, Gottfried Leibniz, a seventeenth century philosopher and mathematician, promoted the “republic of letters”—an independent, self-defining network of scholars that transcends national and religious boundaries (Leibniz, 1931).³ Polanyi (1962/2002, p. 479) contends: “The soil of academic science must be exterritorial in order to secure its rule by scientific opinion.” His republic of science is based on the self-coordination of independent scientists. Authority “is established between scientists, not above them.” (Polanyi, p. 471). Authors like Bush (1945), Merton (1973), and Stokes (1997) warn that outside actors are tempted to shape science according to their own value systems and thus jeopardize the mission of science. This view is supported by the economics of science (Arrow, 1962; Nelson, 1959, 2004; Dasgupta & David, 1994; Stephan, 1996). According to this view, in academia, the evaluation by peers has to substitute for the evaluation by the market because of two fundamental characteristics of science: its public nature and high uncertainty that lead to market failures.

The public nature of scientific discoveries has been intensively discussed by Arrow (1962) and Nelson (1959, 2006). The fundamental uncertainty of scientific endeavors exists because success in academia is reflected by success in the market often only after a long delay or sometimes not at all (Bush, 1945; Nelson, 1959, 2004, 2006). In addition, research often produces serendipity effects; that is, it provides answers to unasked questions (Stephan, 1996; Simonton, 2004). As it is often not predictable which usefulness a particular research endeavor produces and whether it ever will be marketable, peers instead of the market have to evaluate whether a piece of research represents an advance. Peers have the opportunity to

² As Lawrence (2003, p. 259) puts it “Managers are stealing power from scientists.”

³ For a discussion, see Ultee (1987).

identify possible errors and risks; they can profit themselves from the innovation to push forward their own research, redundancies are avoided, and the new knowledge can quickly be used for new and cheaper technologies. Instead of market prices there is a special “currency” that governs the republic of science—the priority rule (Merton, 1973; Dasgupta & David, 1994; Stephan, 1996; Gittelman & Kogut, 2003). This rule attributes success to the person who first makes an invention, and who the scientific community recognizes to be first. The priority rule serves two purposes, hastening discoveries and their disclosure (Dasgupta & David, 1994, p. 499): A discovery must be communicated as quickly as possible to the community of peers in order to gain their recognition.

Consequently, the peer review system is taken to be the founding stone of academic research evaluation. Indicators are awards, honorary doctorates, or membership in prestigious academies (Stephan, 1996; Frey & Neckermann, 2008).⁴ Its main form for the majority of scholars consists of publications and citations in professional journals with high impact factors. Such indicators are provided by academic rankings, based on peer-reviewed publications, citations, and the impact factors of journals like Thomson Reuters’s Impact Factor (JIF) (see Garfield, 2006, for a historical review) and the relatively recent h-index (Hirsch, 2005).⁵

In that view, a well-designed governance system based on academic rankings seems to combine perfectly an output-oriented evaluation of researchers, as postulated by new public management with the requirements of a peer-based evaluation system, as postulated by the concept of the republic of science. On the one side, it seems to be an easy to understand measure for nonexperts like politicians, administrators, and other stakeholders to evaluate the quality of research from outside. On the other side, it is based on the evaluations of peers who are able to assess the quality of research from inside the scientific world. Therefore, today these measures are adopted almost universally in academia for most things that matter as part of the present research governance system: tenure, salary, grants, and budget decisions. This has led to an ever-growing evaluation industry and actively marketed tools like the ISI Web of Science.

⁴ Zuckerman (1992) estimates that by the beginning of the 1990s around 3,000 different scientific awards existed in North America.

⁵ Examples of prominent rankings are the ISI Web of Knowledge Journal Citation Report (The Thomson Corporation, 2008b), ISI Web of Knowledge Essential Science Indicators (The Thomson Corporation, 2008a), IDEAS Ranking (IDEAS, 2008), Academic Ranking of World Universities (Shanghai Jiao Tong University, 2007); or Handelsblatt Ranking (Handelsblatt, 2010).

Peer Reviews as the Basis of Academic Rankings

Academic rankings are based on peer reviews because they are an aggregation of a great number of individual peer reviews leading to publications in refereed journals that in turn lead to citations and impact factors. However, peer reviews are faced with serious problems that have recently been discussed (e.g., Armstrong, 1997; Wenneras & Wold, 1999; Brook, 2003; Frey, 2003; Bedeian, 2004; Starbuck, 2005, 2006; Tsang & Frey, 2007; Gillies, 2005, 2008; Abramo, Angelo, & Caprasecca, 2009; Bornmann & Daniel, 2009).⁶

- *Low inter-rater reliability.* There is an extensive literature on the low extent to which reviewing reports conform to each other (Miner & MacDonald, 1981; Cole, 1992; Weller, 2001; Miller, 2006). 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, p. 1964).⁷ Peters and Ceci (1982) conducted a study of peer reviewing that was the subject of much discussion. They resubmitted 12 articles to the top-tier journals that had published them only 18 to 32 months earlier, giving the articles fictitious authors at obscure institutions. Only three of 38 editors and reviewers recognized that the articles had already been published. Of the remaining nine articles, eight were rejected. 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; that is, it is easier to identify papers that do not meet minimum quality standards than those that are a result of excellent research (Lindsey, 1991; Moed, 2007).
- *Low prognostic quality.* The reviewers' rating of manuscript quality has been found to correlate only 0.24 with later citations (Gottfredson, 1978). According to Starbuck (2006, pp. 83–84), the correlation of a particular reviewer's evaluation with the actual quality as measured by later citations of the manuscript reviewed is between 0.25 and 0.30. This correlation rarely rises above 0.37, although there is evidence that higher prestige journals publish more high-value articles (Judge, Cable, Colbert, & Rynes, 2007). Because of some

⁶ 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*, June 8, 2008.

⁷ According to Fletcher & Fletcher 2003, p. 66 it needs "to have at least six reviewers, all favouring publication or rejection, for their votes to yield a statistical significant conclusion"

randomness in editorial selections (Starbuck, 2005),⁸ one editor even advises rejected authors to “Just Try, Try Again” (Durso, 1997).⁹

- *Low consistency over time.* 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; Horrobin, 1996; Lawrence, 2003). This means that, in the case of radical innovations or paradigm shifts (Kuhn, 1962), peer reviews often fail.
- *Confirmation biases.* Reviewers find methodological shortcomings in 71 percent of papers contradicting the mainstream, compared to only 25 percent of papers supporting the mainstream (Mahoney, 1977).

Although the shortcomings of peer reviews are substantial, they may be 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 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.¹⁰ In addition, it is expected that some of the problems of peer reviews can be avoided.

⁸ See also the “Social Text”-Affair, which deals with the malfunction of editors: 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 (see Sokal, 1996).

⁹ However, this strategy overburdens reviewers and may lower the quality of reviews. For example, they have neither enough time nor the incentive to check the quality of the data and of the statistical methods employed, as some striking examples in economics demonstrate (Hamermesh, 2007).

¹⁰ 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 has used mostly bibliometrics in the past, plans in the future to strengthen qualitative peer review methods (Donovan, 2007).

Strengths and Weaknesses of Academic Rankings

Rankings have several advantages compared to qualitative peer reviews that help to explain why they have become so popular in the last years (Abramo et al., 2009).

- Rankings are more objective because they are based on more than the three or four evaluations typical for qualitative approaches. Through statistical aggregation, individual reviewers' biases may be balanced (Weingart, 2005).
- The influence of the old boys' network may be avoided. An instrument is provided to dismantle unfounded claims to fame. Rankings can serve as fruitful, exogenous shocks to some schools and make them care more about the reactions of the public (Khurana, 2007, p. 337).
- Rankings are cheaper than pure qualitative reviews, at least in terms of time. They admit updates and rapid intertemporal comparisons.
- Rankings facilitate the comparison between a large numbers of scholars or institutions.
- They give research administrators, politicians, journalists, and students an easy to use device to evaluate the standing of the research. As a consequence, attention for research outcomes and the willingness to spend money might arise.

However, 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). Until now, mainly technical and methodological problems were highlighted (van Raan, 2005).

Technical problems consist of errors in the citing-cited matching process, leading to a loss of citations to a specific publication. First, it is estimated that this loss amounts on average to 7 percent of the citations. In specific situations, this percentage may even be as high as 30 percent (Moed, 2002). Second, there are many errors made in attributing publications and citations to the source, for example, institutes, departments, or universities. In the popular ranking of the Shanghai Jiao Tong University, these errors led to differences of possibly five to 10 positions in the European list and about 25 to 50 positions in the world list (Moed, 2002). The impact factor of Thomson's ISI Web of Science is accused of having many faults (Monastersky, 2005; Taylor, Perakakis, & Trachana, 2008). It is unlikely that the errors are distributed equally. Kotiaho, Tomkin, and Simmons (1999) find that names from unfamiliar languages lead to a geographical bias against non-English speaking countries. Third, 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 and consistent indices to measure scientific output have been widely discussed recently (Lawrence, 2002, 2003; Frey, 2003, 2009; Adler et al., 2008; Adler & Harzing, 2009).

First, there are selection problems. Often only journal articles are selected for incorporation in the rankings, although books, proceedings, or blogs contribute considerably to scholarly work. Other difficulties include the low representation of small research fields, non-English papers, regional journals, and journals from other disciplines even if they are highly ranked in their respective disciplines. Hence, collaboration across disciplinary boundaries is not furthered.

Second, citations can have a supportive or rejective meaning or merely a herding effect. The probability of being cited is a function of previous citations according to the “Matthew effect” in science (Merton, 1968). Simkin and Roychowdhury (2005) estimate that, according to an analysis of misprints turning up repeatedly in citations, about 70–90 percent of scientific citations are copied from the list of references used in other papers; that is, 70–90 percent of the papers cited have not been read. Consequently, incorrect citations are endemic. They are promoted by the increasing use of meta-analyses, which generally do not distinguish between high and low quality analyses (Todd & Ladle, 2008). In addition, citations may reflect fleeting references to fashionable “hot topics.”

Third, using the impact factor of a journal as a proxy for the quality of a single article leads to substantial misclassification. It has been found that many top articles are published in non-top journals, and many articles in top journals generate very few citations in management research (Starbuck, 2005; Singh, Haddad, & Chow, 2007), economics (Laband & Tollison, 2003; Oswald, 2007), and science (Campbell, 2008; Rinia et al., 1998). A study of the “International Mathematical Union” even concludes that the use of impact factors can be “breathtakingly naïve” (Adler et al., 2008, p. 14) because it leads to large error probabilities.

Fourth, there are difficulties comparing citations and impact factors between disciplines and even between subdisciplines (Bornman, Mutz, Neuhaus, & Daniel, 2008).

Proposals Made to Overcome the Problems of Rankings

In recent times, a number of different suggestions have been made to deal with the technical and methodological problems of rankings. First, a temporary moratorium of rankings has been suggested “until more valid and reliable ways to assess scholarly contributions can be developed” (Adler & Harzing, 2009, p. 72). As is the case for most authors, they believe that the identification of particular shortcomings should serve as a stepping-stone to develop a more reliable research evaluation system (see also Abramo et al., 2009; Starbuck, 2009). In contrast, policy makers admit that indicators like rankings and grants are spurious. However, as long as scholars present no better data, they will use these measures because they believe that the present data are better than none (Schimank, 2005).

Second, it has been argued that rankings should not be used as ready-to-go indicators by non-experts (van Raan, 2005). Therefore, standards of good practice for the analysis, interpretation, and presentation of rankings should be developed and adhered to when assessing research performance. This needs a lot of expertise (Bornmann et al., 2008), which constrains considerably the responsible use of rankings as a handy instrument for politicians, administrators, and journalists to assess academic performance.

Third, it is suggested to use a number of rankings because their results differ markedly (e.g., Adler & Harzing, 2009), in particular with respect to the ranking of individuals (Frey & Rost, 2010). Again, this suggestion constrains rankings as easy to handle instruments for nonexperts. In addition, the universities are even more burdened by evaluation efforts, distracting from research.

Fourth, a combination of qualitative peer reviews and bibliometrics, so-called informed peer reviews, could be applied. It is argued that they can balance the advantages and disadvantages of these two methods (Butler, 2007; Moed, 2007).

Fifth, a holistic approach of evaluation has been suggested, which combines measures of research quality and impact with peer and user evaluations, taking into account the views of various stakeholders inside and outside academia (Donovan, 2007). However, this approach bears the danger of compromising on the smallest common denominator and of inhibiting research with unorthodox or uncertain outcomes.

These suggestions to some extent may mitigate the problems of rankings, but they make the use of rankings difficult for nonexperts; thus, they are not able to reconcile the aims of new public management with the republic of science as intended. Moreover, even if the rankings worked perfectly, they cannot overcome the problems of behavioral reactions to rankings (Osterloh & Frey, 2009).

Behavioral Reactions to Rankings

Even if over time the methodological and technical problems could be handled, severe problems remain, which are caused by the unintended side effects of rankings on the side of individuals and institutions. First, they consist of the so-called reactive measures (Campbell, 1957), caused by the fact that people change their behavior strategically in reaction to being observed or measured, in particular if the measurement is not accepted voluntarily (Espeland & Sauder, 2007). Reactivity threatens the validity of measures according to the saying: “When a measure becomes a target, it ceases to be a good measure” (Strathern, 1996, p. 4). Second, the unintended consequences consist of the danger of reducing the intrinsically motivated curiosity of researchers. Both problems, which are discussed only by a few authors in the research governance literature, have consequences on the level of individual scholars and institutions.

Level of Individual Scholars

Reactivity on the level of individual scholars may take the form of goal displacement or of counterstrategies to “beat the system.” Goal displacement (Perrin, 1998) means that people maximize indicators that are easy to measure and disregard features that are hard to measure. This problem is also discussed as the multiple-tasking effect (Holmstrom & Milgrom, 1991; Ethiraj & Levinthal, 2009). There is much evidence of this effect in laboratory experiments (Staw & Boettger, 1990; Gilliland & Landis, 1992; Schweitzer, Ordonez, & Douma, 2004; Ordonez, Schweitzer, Galinsky, & Bazerman 2009).¹¹ For example, Fehr and Schmidt (2004) show that output-dependent financial incentives lead to the neglect of noncontractible tasks.

In academia, examples of goal displacement can be found, for example, as “slicing strategy,” by breaking them into as many papers as possible to increase their publication list. 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.

Empirical field evidence of goal displacement in academia is shown in an Australian study (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

¹¹ Locke and Latham (2009) in a rejoinder provide counterevidence to Ordonez et al. (2009). They argue that goal setting has no negative effects. However, they disregard that goal setting may well work for simple but not for complex tasks within an organization. For the latter case, see Earley, Connolly, and Ekegren (1989) and Ethiraj and Levinthal (2009).

increased dramatically, but the quality as measured by relative citation rates decreased.¹² A recent study that examined how incentive systems affected submissions and publications to the journal *Science* during the last decade found that submissions per year increased significantly with incentives. However, there was no significant impact of incentives on publications (Franzoni et al., 2010).

Counterstrategies are more difficult to observe than goal displacement. They consist of altering research behavior itself (Moed, 2007). Numerous examples can be found in educational evaluation (e.g., Haney, 2002; Nichols, Glass, & Berliner, 2006; Heilig & Darling-Hammond, 2008). The following behaviors are of special relevance in academia.

Scholars 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 than approvingly cite their work, and these same reviewers tend to reject papers that threaten their previous work (Lawrence, 2003, p. 260).¹³ Some editors admit freely that they encourage authors to cite their respective journals in order to raise their impact rankings (Garfield, 1997; Smith, 1997; Monastersky, 2005).

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 effects of reactivity are enforced if the second kind of unintended consequences takes place, the decrease of intrinsically motivated curiosity which generally is acknowledged to be of decisive importance in academic research (Amabile, 1996, 1998; Spangenberg et al, 1990; Stephan, 1996; Simonton, 2004). In both psychology and psychological economics,¹⁴ 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

¹² It could be argued that a remedy to this problem consists of resorting to citation counts. Although this remedy overcomes some of the shortcomings of publication counts, it is subject to the technical and methodological problems mentioned.

¹³ Such problems of sabotage in tournaments have been extensively discussed in personnel economics, see Lazear and Shaw (2007).

¹⁴ We prefer the expression psychological economics instead of the more common expression behavioral economics for two reasons. First, economists had already examined human behavior before this new field emerged. Second, Simon (1985) points out that the term behavioral is misleading because it may be confounded with the behaviorist approach in psychology.

feedback and are perceived to be controlling.¹⁵ (Hennessey & Amabile, 1998; Frey, 1992, 1997; Deci, Koestner, & Ryan, 1999; Gagné & Deci, 2005; Falk & Kosfeld, 2006; Ordóñez et al., 2009.)¹⁶

From that point of view, rankings tend to crowd out intrinsically motivated curiosity. First, in contrast to qualitative peer reviews, 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. It is substituted by the position in the rankings (Kruglansky, 1975). 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.

Level of Institutions

Reactivity on the institutional level takes several forms. First, if rankings are used as a measure to allocate resources and positions, they create a *lock-in effect*. Even those scholars and academic institutions that are aware of the deficiencies of rankings do well not to oppose them. If they did so, they would not only be accused of being afraid of competition, but also of not contributing to the prestige and resources of their department or university. Therefore, it is a better strategy to follow the rules and to play the game. For example, in several countries, highly cited scientists are hired in order to raise publication and citation records. Such stars are highly paid although they often have little involvement with the respective university (Brook, 2003; Stephan, 2008)

Second, a negative *walling-off effect* sets in. Scholars themselves are inclined to apply rankings to evaluate candidates in order to gain more resources for their research group or department. In addition, it is easier to count the publications and citations of colleagues than to evaluate the content of their scholarly contributions. By doing this, the scholars delegate their own judgment to the counting exercise behind rankings, although by using such metrics they admit their incompetence in that subject (Browman & Stergiou, 2008). This practice is defended by arguing that specialization in science has increased so much that even within disciplines it is impossible to evaluate the research in neighboring fields (Swanson, 2004; van

¹⁵ A third precondition is social relatedness, see Gagne and Deci (2005).

¹⁶ The crowding-out effect sometimes is contested, for example, Eisenberger and Cameron (1996), Gerhart and Rynes (2003), Locke and Latham (2009). However, the empirical evidence for complex tasks and actors intrinsically motivated in the first place is strong (Deci et al., 1999; Weibel, Rost, & Osterloh, 2010). For a survey of the empirical evidence see Frey and Jegen (2001).

Fleet, McWilliams, & Siegel, 2000). However, this practice in turn reinforces specialization and furthers a walling-off effect between disciplines and subdisciplines. By using output indicators instead of communicating on the contents, the knowledge in the various fields becomes increasingly disconnected. This hampers the ability to create radical innovations that often cross disciplinary borders (Amabile, Conti, Coon, Lazenby, & Herron, 1996; Dogan, 1999).

Third, *research is in danger of being increasingly homogenized*. Research endeavors tend to lose the diversity that is necessary for a creative research environment. This consequence was pointed out for business schools by Gioia and Corley (2002). For economics, Great Britain provides an example: the share of heterodox, not strictly neoclassical economics, sank drastically since the ranking of departments became important. Heterodox journals have become less attractive for researchers due to their smaller impact factor when compared to mainstream journals (Lee, 2007; Holcombe, 2004).

Fourth, the *establishment of new research areas is inhibited*. It has been argued that in Great Britain, the Research Assessment Exercise has discouraged research with uncertain outcomes and has encouraged projects with quick payoffs (Hargreaves Heap, 2002).

Fifth, it is argued that a *positional competition* or a rent-seeking game takes place instead of an enhancement of research quality by the increased investment by universities and journals in evaluating research (Ehrenberg, 2000). It has been shown that the percentage of “dry holes” (i.e., articles in refereed journal that have never been cited) in economic research during 1974 to 1996 has remained constant (Laband & Tollison, 2003), although the resources to improve the screening of papers have risen substantially.

With respect to the *motivational aspects* of rankings on the institutional level, a negative selection effect is to be expected, in particular, when monetary rewards are linked to the position in rankings. In academia, a special incentive system called “taste for science” exists (Merton, 1973; Dasgupta & David, 1994; Stephan, 1996; Sorenson & Fleming, 2004; Roach & Sauermann, 2010). It is characterized by a relatively low importance of monetary incentives and a high importance of peer recognition and autonomy. People are attracted to research for which, at the margin, the autonomy to satisfy their curiosity and to gain peer recognition is more important than money. They value the possibility of following their own scientific goals more than financial rewards (Bhagwat et al, 2004). These scholars are prepared to trade-off autonomy against money, as empirically documented by Roach and Sauermann (2010) and Stern (2004): scientists pay to be scientists. The preference for autonomy to choose their own goals is important for innovative research in two ways. It leads

to a useful self-selection effect, and autonomy is the most important precondition for intrinsic motivation, which in turn is required for creative research (Amabile et al., 1996; Amabile, 1998; R. Mudambi, S. Mudambi, & Navarra, 2007).

Are there Alternatives to Academic Rankings?

As discussed, academic rankings have advantages and disadvantages as a basis of research governance. So far, it has not been decided whether the advantages of rankings outweigh the disadvantages. The intended advantages consist of more transparency and control of research by nonexperts, as expressed by the view of new public management. Moreover, rankings are able to avoid the pernicious cronyism that often prevailed in the past. The disadvantages consist firstly of the technical and methodological problems that might be overcome sometime in the future. Secondly, they consist of the behavioral reactions of reactivity and motivation disturbances that remain even if the indicators were perfect. As a consequence, there is the danger that “the very action of controlling universities and making them more accountable leads them to give a less good account” (Hargreaves Heap, 2002, p. 388). The question arises whether there is a third way for research governance, which makes some use of peer reviews and rankings, but limits its disadvantages.

To answer this question we refer to insights from managerial control theory (e.g., Thompson, 1967; Ouchi, 1977, 1979; Eisenhardt, 1985; Schreyögg & Steinmann, 1987; Simons, 1995). According to this approach, different kinds of controls are needed to achieve strategic goals. Three types of control systems may be distinguished: output control, process control, and input control. The type of control applied must fit the knowledge available to the controller with respect to outcome measurability and process relations (Turner & Makhija, 2006).

Output control is useful if well-defined unambiguous indicators are available to the evaluator, although knowledge of cause-and-effect or process relations is not necessary. Therefore, output controls are attractive to nonexperts. As we have discussed, rankings are far from delivering such unambiguous indicators to nonexperts and should therefore be used with the utmost care. *Process control* is useful when outputs are not easy to measure and to attribute, but when the controller is knowledgeable on process relations whose correctness is to be evaluated ex post. Therefore, process control is applicable only for peers who are familiar with the state of the art about processes and methodologies in the respective research field. As discussed, peer control has many shortcomings and is particularly questionable when

unorthodox contributions have to be evaluated. In such cases, well-established standards or methods often are challenged.

If neither output control nor process control works sufficiently, then *input control* has to be applied (Ouchi, 1977, 1979). Input control is an ex-ante form of control, based on careful selection and socialization. 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 measurement is questionable and procedural rules are in flux.

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 was recommended by the famous President of Harvard University James Bryan Conant: “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” (Renn, 2002).¹⁷ This view is still 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.”¹⁸ Input control is also applied with the selection of fellows at the Institutes of Advanced Studies.

A comparison between two Australian universities with similar research interests illustrates that input control is not only useful for top research institutions like Harvard or Institutes of Advanced Studies (Butler, 2003). In the late 1980s, the University of Western Australia distributed research funds according to publication counts as the main criterion. The University of Queensland followed a different strategy, recruiting bright young researchers and providing them with a strong resource base. Both universities succeeded in lifting their publications per researcher. However, only the University of Queensland was successful in improving the quality of its publications, whereas the University of Western Australia fell below the average Australian score.

¹⁷ Letter to the New York Times, August 13, 1945.

¹⁸ See <http://www.fas.harvard.edu/research/greybook/principles.html>

Input control has empirically proven to be successful also in R&D organizations of industrial companies (Abernethy & Brownell 1997). It is employed also in other professions characterized by a low degree of observable outputs. Examples are life-tenured judges (e.g., Benz & Frey, 2007; Posner, 2010), and executive search companies (Zehnder, 2001). To some extent, input control is applied by innovative companies like Google and 3M (Brand, 1998),¹⁹ which allow their researchers to spend 20 to 40 percent of their work time in pursuing self-chosen goals.²⁰ These ideas are 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 some similarities to the traditional concept of the republic of science with its emphasis on qualitative peer reviews. However, there are major differences. First, input control restricts itself to admission control in specific situations of status passage (Glaser & Strauss, 1971). This is a wise self-restriction in view of the numerous shortcomings of peer review discussed above. Second, input control includes not only qualitative peer reviews but may take into account various metrics in an informed way. Input control therefore is a comprehensive form of control which combines all other kinds of control. Third, the criteria of admission must be communicated in a transparent way to the outside. They are not of an “extraterritorial” nature for the republic of science as claimed by Polanyi (1962/2002) but must be comprehensible for the public, politicians, and university administrators. The quality of a university depends on how carefully and rigidly the criteria of admission are followed. For example, they must involve a broad and international selection of reviewing peers in order to avoid cronyism. This procedure corresponds to the characteristics of an acceptable legal process, where the judges make decisions according to transparent and comprehensible rules, although the content of a decision is sometimes hard to understand.

¹⁹ See <http://www.google.com/support/jobs/bin/static.py?page=about.html&about=eng>

²⁰ Another impressive example for how autonomy in knowledge productions furthers productivity is open source software production, see Osterloh and Rota (2007).

Advantages and Disadvantages of Input Control

Input control has advantages and disadvantages. The *advantages* first consist in downplaying the unfortunate behavioral reactions to rankings while inducing young scholars to learn the professional standards of their discipline under the supporting assistance of peers. This support allows them to balance the internal tension of scientific work between conformity and originality. “The professional standards of science must impose a framework of discipline and at the same time encourage rebellion against it” (Polanyi 1962/2002, p. 470). Second, although input control still requires peer evaluations, this applies during limited time periods, namely during situations of status passage. However, to escape the permanent pressure makes a great difference and provides much autonomy. 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 communication process. In contrast, rankings tend to impose a one dimensional order on scholarly work, in particular if one or few rankings dominate public opinion. Fourth, input control is better able than pure output control to use different indicators in an informed way by taking their relative strengths and weaknesses into account. Fifth, to the extent input control is accompanied by the provision of basic funds to those that have passed the entrance barriers, diversity of research approaches increases (Gillies, 2008). It helps to avoid inefficient “research empires” subject to a decreasing marginal effect of additional research resources (Horrobin, 1996; Viner, Powell, & Green, 2004). Although there exists some empirical work in this regard (Etzkowitz & Leydesdorff, 2000; Jansen, Wald, Frenke, Schmoch, & Schubert, 2007), this issue needs further research.

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. However, this is the price that has to be paid for potential high performers to flourish. It will be lower when the selection process is conducted rigorously. As a consequence, recruiting is by far the most important issue for academic self-governance. Insufficient recruiting efforts cannot be substituted for output control.

Second, input control is in danger of being submitted to groupthink (Janis, 1972). This danger can be overcome by fostering the diversity of scholarly approaches within the relevant peer group.²¹

²¹ Rost and Osterloh (in press) showed empirically, that in the Swiss banking industry during the recent financial market crisis companies with heterogeneous boards performed better than those with homogeneous boards.

Third, input control in the form of careful selection and socialization processes is more costly than pure output control in the form of counting publications and citations. However, if linked to an incentive system, output control may become very expensive. In the financial industry it became obvious that monetary incentives linked to short-term indicators could be disastrous. As a consequence, even in such industries today, input instead of output control is proposed (e.g., Schmidt, 2010).

Fourth, 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 can only control whether the commonly agreed upon criteria of admissions are met. They have to accept that to evaluate scholarly activities—as is sometimes the case for professional activities outside academia—there is no clear-cut quality criteria because disputes on heterogeneous views are the essence of the scholarly communication process.²² On the other hand, we have shown that single rankings fail to provide a reliable quality measure. A multiplicity of rankings and informed peer reviews may be an alternative, but they need a lot of expertise to be interpreted in a serious way and thus are not easy to handle instruments for the public.

To compensate for the disadvantages of input control, two measures are advisable that introduce some elements of process and output control. The first measure consists of periodic self-evaluation over considerable intervals of, say, six years (as done in some research universities in the United States). External members may be involved but should not dominate the process. The major goal is to induce self-reflection and feedback among the members of a research unit, comparable to the goals of organization development (e.g., Bradford & Burke, 2005). The second measure compensates to a certain extent for the limited visibility of input control to the public. Awards like prizes and titles as well as different kinds of professorships and fellowships (from assistant to distinguished) signal the recognition of peers to nonexperts (Frey & Osterloh, 2010). They consist of an overall evaluation, which avoids the issue that particular metrics can be manipulated (Frey & Neckermann, 2008).

As empirical evidence shows, both measures, though being partly extrinsic motivators, do not crowd out intrinsic motivation (Neckermann, Cueni, & Frey, 2010). They match the main motivational factors that conform to the “taste of science.” These factors consist in the first place of peer recognition and the granting of autonomy, while pay plays a secondary role

²² 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 university’s research performance is improved after an accomplished scholar has been hired as president.

(Jimenex-Contreras, de Moya Anegon, & Delgado Lopez-Cozar, 2003; Stern, 2004; Roach & Sauermann, 2010). They motivate well even for those who do not actually win such an award.²³

To conclude, simple and easy to comprehend criteria to evaluate scholarly activity and to motivate researchers are available only at a high cost. In our view, input control is a better method than qualitative peer reviews and rankings, insofar as it takes into account all kinds of evaluation methods but limits itself to specific situations. In particular, it refrains from recruitment systems, resource allocation systems, and incentive systems that are linked mainly to output indicators like rankings.

The discussion suggests that it is questionable to combine new public management with the concept of the republic of science. Research governance must acknowledge that the “secrets of the world of research” (Weingart, 2005, p.119) can be unlocked to the public to a limited extent only—this is the price that must be paid for a flourishing scholarly enterprise.

Policy Implications

This paper argues that research governance mainly based on academic rankings has major disadvantages that tend to be disregarded or downplayed both in the literature and in practice. For scholarly work, instead of output control, the emphasis should be put on input control. Rigorous selection and socialization should play a major role in research governance, supplemented by periodic self-evaluation and awards. This kind of control encompasses process control like peer reviews and output control like rankings in a comprehensive and informed way but only during limited time periods. In contrast, rankings should be attributed lesser importance. All this does not signal a return to the old system of “academic oligarchy” as long as the heterogeneity of scholarly approaches is maintained.

Because of the lock-in effect, the change in academic governance cannot be achieved by individual scholars or single institutions. It needs more far-reaching institutional changes. In particular, the steering bodies overseeing the research system should place more emphasis on the selection and socialization process that provides the basis of academic excellence.

²³ The money attached to awards is less important than the reputation of the award-giving institution, see Frey and Neckermann (2008).

References

- Abernethy, M.A., & Brownell, P. (1997). Management Control Systems in Research and Development Organizations: The role of Accounting, Behavior and Personnel Controls. *Accounting, Organizations and Society*, 22 (3/4), 233-248.
- Abramo, G. D., Angelo, C. A., & Caprasecca, A. (2009). Allocative efficiency in public research funding: Can bibliometrics help? *Research Policy*, 38, 206–215.
- Adler, N. J., & Harzing, A-W. (2009). When knowledge wins: Transcending the sense and nonsense of academic rankings. *Academy of Management Learning*, 8, 72–95.
- Adler, R., Ewing, J., & Taylor, P. (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).
- Albers, S. (2009). Misleading rankings of research in business. *German Economic Review*, 3, 352–363.
- Amabile, T. M. (1996). *Creativity in context: Update to the social psychology of creativity*. Boulder, CO: Westview Press.
- Amabile, T. M. (1998). How to kill creativity. *Harvard Business Review*, 76, 76–87.
- Amabile, T. M., Conti, R., Coon, H., Lazenby, J., & Herron, M. (1996). Assessing the work environment for creativity. *Academy of Management Journal*, 39, 1154–1184.
- Armstrong, J. S. (1997). Peer review for journals: Evidence on quality control, fairness, and innovation. *Science and Engineering Ethics*, 3, 63–84.
- Arrow, K. (1962). Economic welfare and the allocation of resources for invention. In R. Nelson (Ed.), *The rate and direction of inventive activity: Economic and social factors* (pp. 609–626). Princeton, NJ: Princeton 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. (2004). Peer review and the social construction of knowledge in the management discipline. *Academy of Management Learning and Education*, 3, 198–216.
- Benz, M., & Frey, B. S. (2007). Corporate governance: What can we learn from public governance? *Academy of Management Review*, 32(1), 92–104.

- Bhagwat, J.G., Ondategui-Parra, S., Zou, K.H., Gogate, A., Intriere, L.A., Kelly, P. Seltze, S. E., & Ros, P.R. (2004). Motivation and compensation in Academic Radiology. *Journal of the American College of Radiology*, 1 (7), 493-496.
- Bleiklie, I., & Kogan, M. (2007). Organization and Governance of Universities. *Higher Education Policy*, 20, 477-493.
- Bok, D. (2003). *Universities in the marketplace. The commercialization of higher education*. Princeton, NJ: Princeton University Press.
- Bornmann, L., & Daniel, H.D. (2009). The luck of the referee draw: The effect of exchanging reviews. *Learned publishing*, 22 (2), 117-125.
- Bornmann, L., Mutz, R., Neuhaus, C., & Daniel H. D. (2008, June). Citation counts for research evaluation: standards of good practice for analyzing bibliometric data and presenting and interpreting results. *Ethics in Science and Environmental Politics*, 8, 93-102.
- Bradford, D. L., & Burke, W. W. (Eds.). (2005). *Organization Development*. San Francisco, CA: Pfeiffer.
- Brand, A. (1998). Knowledge Management and Innovation at 3M. *Journal of Knowledge Management*, 2(1), 17-22.
- Brook, R. (2003). Research survival in the age of evaluation. In *Science between evaluation and innovation: A conference on peer review* (pp. 61-66). München: Max-Planck-Gesellschaft.
- Browman, H. I., & Stergiou, K. I. (2008). Factors and indices are one thing, deciding who is scholarly, why they are scholarly, and the relative value of their scholarship is something else entirely. *Ethics in Science and Environmental Politics*, 8, 1-3.
- Bush, V. (1945, July). *Science: The endless frontier. A report to the president by Vannevar Bush, Director of the Office of Scientific Research and Development*. Washington, D.C.: United States Government Printing Office. Retrieved July 1, 2010, from <http://www.nsf.gov/od/lpa/nsf50/vbush1945.htm>
- Butler, L. (2003). Explaining Australia's increased share of ISI publications—the effects of a funding formula based on publication counts. *Research Policy*, 32, 143-155.
- Butler, L. (2007). Assessing university research: a plea for a balanced approach. *Science and Public Policy*, 34, 565-574.
- Campanario, J. M. (1996). Using citation classics to study the incidence of serendipity in scientific discovery. *Scientometrics*, 37, 3-24.

- Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, *54*, 297–312.
- Campbell, P. (2008, June). Escape from the impact factor. *Ethics in Science and Environmental Politics*, *8*, 5–7.
- Cicchetti, D. V. (1991). The reliability of peer review for manuscript and grant submissions: A cross-disciplinary investigation. *Behavioral and Brain Sciences*, *14*, 119–135.
- Clark, B. R. (1998). *Creating entrepreneurial universities. Organizational pathways of transformation*. Surrey: Pergamon Press.
- Cole, S. (1992). *Making science. Between nature and society*. Cambridge, MA: Harvard University Press.
- Daily, C. M., Dalton, D. R., & Cannella, A. A. (2003). Corporate Governance: Decades of Dialogue and Data. *Academy of Management Review*, *28*, 371–382.
- Dasgupta, P., & David, P. A. (1994). Toward a new economics of science. *Research Policy*, *23*, 487–521.
- 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*, 627–668.
- Dogan, M. (1999). Marginality. *Encyclopedia of Creativity*, *2*, 179–184.
- Donoghue, F. (2008). *The last professors. The corporate university and the fate of the humanities*. New York, NY: Fordham University Press.
- Donovan, C. (2007). The qualitative future of research evaluation. *Science and Public Policy*, *34*, 585–597.
- Durso, T. W. (1997). Editor's advice to reject authors: Just try, try again. *The Scientist*, *11*, 13.
- Earley, P. C., Connolly, T., & Ekegren, G. (1989). Goals, strategy development, and task performance: Some limits on the efficacy of goal setting. *Journal of Applied Psychology*, *74*, 24–33.
- Ehrenberg, R. G. (2000). *Tuition rising: Why college costs so much*. Cambridge, MA: Harvard University Press.
- Eisenberger, R., & Cameron, J. (1996). Detrimental effects of reward: Reality or myth? *American Psychologist*, *51*(11), 1153–1166.
- Eisenhardt, K. M. (1985). Control: Organizational and economic approaches. *Management Science*, *31*(2), 134–149.
- Espeland, W.N., & Sauder, M. (2007). Rankings and Reactivity: How Public measures Recreate Social Worlds. *American Journal of Sociology*, *113*(1), 1–40.

- Ethiraj, S. K., & Levinthal, D. (2009). Hoping for A to Z while rewarding only A: Complex organizations and multiple goals. *Organization Science*, *20*, 4–21.
- Etzkowitz, H., & Leydesdorff, L. (2000). The dynamics of innovation: From national systems and “mode 2” to a triple helix of university–industry–government relations. *Research Policy*, *29*, 109–123.
- Falk, A., & Kosfeld, M. (2006). The hidden cost of control. *American Economic Review*, *96*, 1611–1630.
- Fehr, E., & Schmidt, K. M. (2004). Fairness and incentives in a multi-task principal-agent model. *Scandinavian Journal of Economics*, *106*, 453–474.
- Ferraro, F., Pfeffer, J., & Sutton, R. I. (2005). Economics language and assumptions: How theories can become self-fulfilling. *Academy of Management Review*, *30*, 8–24.
- Fong, E. A., & Tosi Jr., H. L. (2007). Effort, performance, and conscientiousness: An agency theory perspective. *Journal of Management*, *33*, 161–179.
- Franzoni, C., Scellato, G., & Stephan, P. E. (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. *Kyklos*, *45*, 161–185.
- Frey, B. S. (1997). *Not just for the money: An economic theory of personal motivation*. Cheltenham, UK: Edward Elgar Publishing.
- Frey, B. S. (2003). Publishing as prostitution? – Choosing between one’s own ideas and academic success. *Public Choice*, *116*, 205–223.
- Frey, B. S. (2009). Economists in the PITS. *International Review of Economics*, Vol 56, Number 4. Heidelberg, Germany: Springer Berlin.
- Frey, B. S., & Jegen, R. (2001). Motivation crowding theory. *Journal of Economic Surveys*, *15*(5), 589–611.
- Frey, B. S., & Neckermann, S. (2008). Awards – A view from psychological economics. *Journal of Psychology*, *216*, 198–208.
- Frey, B. S., & Osterloh, M. (2010). Motivate People with Prizes, *Nature*, *465*, 871.
- Frey, B. S., & Rost, K. (2010). Do Rankings Reflect Research Quality? *Journal of Applied Economics*, *13*(1), 1–38.
- Frost, J., Osterloh, M., & Weibel, A. (2010). Governing Knowledge Work: Transactional and Transformational Solutions. *Organizational Dynamics*, *39*, 126–136.
- Fuyuno, I., & Cyranoski, D. (2006). Cash for papers: putting a premium on publication. *Nature*, *441*, 792.

- Gagné, M., & Deci, E. L. (2005). Self-determination theory and work motivation. *Journal of Organizational Behavior*, 26, 331–362.
- 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, 1765.
- Garfield, E. (2006). The history and meaning of the journal impact factor. *The Journal of the American Medical Association (JAMA)*, 295(1), 90–93.
- Gerhart, B., & Rynes, S. L. (2003). *Compensation: theory, evidence, and strategic implications*. Thousand Oaks, CA: Sage Publications, Inc.
- Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., & Trow, M. (1994). *The New Production of Knowledge. The Dynamics of Science and Research in Contemporary Societies*. Thousand Oaks, CA: Sage Publications, Inc.
- Gillies, D. (2005). 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. (2008). *How should research be organised?* London, UK: College Publication King's College.
- Gilliland, S. W., & Landis, R. S. (1992). Quality and quantity goals in a complex decision task: strategies and outcomes. *Journal of Applied Psychology*, 77(5), 672–681.
- Gioia, D. A., & 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, 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.
- Glaser, B. G., & Strauss, A. (1971). *Status Passage*. Rutgers, NJ: Transaction Publishers.
- Gneezy, U., & Rustichini, A. (2000). Pay enough or don't pay at all. *Quarterly Journal of Economics*, 115, 791–810.
- Goodall, A.H. (2009). Highly cited leaders and the performance of research universities. *Research Policy*, 38, 1070-1092.
- Gottfredson, S. D. (1978). Evaluating psychological research reports: Dimensions, reliability, and correlates of quality judgments. *American Psychologist*, 33, 920–934.

- Hamermesh, D. S. (2007). Viewpoint: replication in economics. *Canadian Journal of Economics*, 40(3), 715–733.
- Handelsblatt. (2010). Handelsblatt-Ranking schlägt hohe Wellen. Retrieved July 1, 2010, from <http://www.handelsblatt.com/politik/vwl-ranking/>
- Haney, W. M. (2002, July 10). Ensuring failure: How a state's achievement test may be designed to do just that. *Education Week*, 56–58.
- Hargreaves Heap, S. P. (2002). Making British universities accountable. In P. Mirowski & E-M. Sent (Eds.), *Science Bought and Sold: Essays in the Economics of Science* (pp. 387–411). Chicago, IL: University of Chicago Press.
- Heilig, J. V., & Darling-Hammond, L. (2008). Accountability Texas-style: The progress and learning of urban minority students in high-stakes testing context. *Educational Evaluation and Policy Analysis*, 30(2), 75–110.
- Hennessey, B. A. & Amabile, T. (1998). Reward, Intrinsic Motivation and Creativity. *American Psychologist*, 53(6), 647-675.
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 16569–16572. Retrieved July 1, 2010, from <http://dx.doi.org/10.1073/pnas.0507655102>
- 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, P. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *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? *Lancet*, 348, 1293–1295.
- IDEAS. (2008). *IDEAS rankings*. Retrieved July 2, 2010, from <http://ideas.repec.org/top/>
- Janis, I. L. (1972). *Victims of groupthink: A psychological study of foreign-policy decisions and fiascos*. Boston, MA: Houghton Mifflin.
- Jansen, D., Wald, A., Frenke, K., Schmoch, U., & Schubert, T. (2007). Drittmittel als Performanzindikator der wissenschaftlichen Forschung. Zum Einfluss von Rahmenbedingungen auf Forschungsleistung. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 59, 125–149.

- Jimenez-Contreras, E., de Moya Anegon, F., & Delgado Lopez-Cozar, E. (2003). The Evolution of Research Activity in Spain: The Impact of the National Commission for the Evaluation of Research Activity (CNEAI). *Research Policy*, 32, 123–142.
- Judge, T. A., Cable, D. M., Colbert, A. E., & Rynes, S. L. (2007). What causes a management article to be cited – article, author, or journal? *Academy of Management Journal*, 50, 491–506.
- 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, NJ: Princeton University Press.
- Kotiaho, J. S., Tomkin, J. L., & Simmons, L. W. (1999, July). Unfamiliar citations breed mistakes. *Nature*, 400, 307.
- Kruglanski, A. W. (1975). The endogenous-exogenous partition in attribution theory. *Psychological Review*, 82, 387–406.
- Kuhn, T. (1962). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.
- Laband, D. N., & Tollison, R. D. (2003). Dry holes in economic research. *Kyklos*, 56, 161–174.
- Lane, J. (2010, March). Let's make science metrics more scientific, *Nature*, 464, 488–489.
- Lawrence, P. A. (2002, February 21). Rank injustice: the misallocation of credit is endemic in science. *Nature*, 415, 835–836.
- Lawrence, P. A. (2003, March 20). The politics of publication – authors, reviewers, and editors must act to protect the quality of research. *Nature*, 422, 259–261.
- Lazear, E. P., & Shaw, K. L. (2007). Personnel economics: The economist's view of human resources. *Journal of Economic Perspectives*, 21, 91–114.
- Lee, F. S. (2007). The research assessment exercise, the state and the dominance of mainstream economics in British universities. *Cambridge Journal of Economics*, 31, 309–325.
- Leibniz, G. W. (1931). Relation de l'état présent de la république des lettres. *Sämtliche Schriften und Briefe (Akademie-Ausgabe)*, 4. Reihe, Politische Schriften, Band 1. Darmstadt, 569–571.
- Lindsey, D. (1991). Precision in the manuscript review process: Hargens and Herting revisited. *Scientometrics*, 22(2), 313–325.
- Locke, E. A., & Latham, G. P. (2009). Has goal setting gone wild, or have its attackers abandoned good scholarship? *Academy of Management Perspectives*, 21, 17–23.

- Mahoney, M. J. (1977). Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy Research, 1*, 161–175.
- Marginson, S., & Considine, M. (2000). *The enterprise university: Power, governance and reinvention in Australia*. Cambridge, MA: Cambridge University Press.
- Merton, R. K. (1968). The Matthew effect in science. *Science, 159*, 56–63.
- Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigation*. Chicago, IL: University of Chicago Press.
- Miller, C. C. (2006). Peer Review in the organizational and management sciences: Prevalence and effects of reviewer hostility, bias, and dissensus. *Academy of Management Journal, 49*, 425–431.
- Miner, L., & McDonald, S. (1981). Reliability of peer review. *Journal of the Society of Research Administrators, 12*, 21–25.
- Moed, H. F. (2002). The impact factors debate; the ISI's uses and limits. *Nature, 415*, 731–732.
- Moed, H. F. (2007). The future of research evaluation rests with an intelligent combination of advanced metrics and transparent peer review. *Science and Public Policy, 34*, 575–583.
- Monastersky, R. (2005). The number that's devouring science. *Chronicle of Higher Education, 52* (8), A12.
- Mudambi, R., Mudambi, S., and Navarra, P. (2007). Global innovation in MNCs: The effects of subsidiary self-determination and teamwork. *Journal of Product Innovation Management, 24*, 442–455.
- Neckermann, S., Cueni, R., & Frey, B. S. (2010, May). Awards at work. (Working Paper Nr. 411, Institute for Empirical Research in Economics, University of Zurich).
- Nelson, R. R. (1959). The simple economics of basic scientific research. *Journal of Political Economy, 67*, 297–306.
- Nelson, R. R. (2004). The market economy, and the scientific commons. *Research Policy, 33*, 455–471.
- Nelson, R. R. (2006). Reflections on “The simple economics of basic scientific research”: Looking back and looking forward. *Industrial and Corporate Change, 15*, 903–917.
- Nichols, S. L., Glass, G. V., & Berliner, D. C. (2006). High-stakes testing and student achievement: Does accountability pressure increase student learning? *Education Policy Analysis Archives, 14*(1), 1–175.

- Ordonez, L. D., Schweitzer, M. E., Galinsky, A. D., & Bazerman, M. H. (2009). Goals gone wild: the systematic side effects of overprescribing goal setting. *Academy of Management Perspectives*, 23, 6–16.
- Osterloh, M., & Frey, B. S. (2000). Motivation, knowledge transfer, and organizational forms. *Organization Science*, 11, 538–550.
- Osterloh, M., & Frey, B. S. (2009). Are more and better indicators the solution? Comment to William Starbuck. *Scandinavian Journal of Management*, 25(2), 225–227.
- Osterloh, M., & Rota, S. (2007). Open Source Software Development - Just another Case of Collective Invention? *Research Policy* 36(2), 157–171.
- Oswald, A. J. (2007). An examination of the reliability of prestigious scholarly journals: Evidence and implications for decision-makers. *Economica*, 74, 21–31.
- Ouchi, W. G. (1977). The relationship between organizational structure and 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, 833–848.
- Perrin, B. (1998). Effective use and misuse of performance measurement. *American Journal of Evaluation*, 19, 367–379.
- Peters, D., & Ceci, S. J. (1982). Peer review practices of psychological journals: The fate of published articles, submitted again. *The Behavioral and Brain Sciences*, 5, 187–195.
- Polanyi, M. (1962). The republic of science: Its political and economic theory. *Minerva*, 1, 54–73. (Reprinted in Polanyi, M. (1969). *From Knowing and Being*. Chicago, IL: University of Chicago Press, pp. 49–72. Re-reprinted in Mirowski, P., & Sent, E. M. (2002). *Science Bought and Sold. Essays in the Economics of Science* (pp. 465–485). Chicago, IL: The University of Chicago Press.)
- Posner, R.A. (2010). From the new institutional economics to organization economics: Four applications to corporate governance, government agencies, and legal institutions. *Journal of Institutional and Theoretical Economics*, 6, 1–37. doi: 10.1017/S1744137409990270
- Prichard, C., & Willmott, H. (1997). Just how managed is the McUniversity? *Organization Studies*, 18, 287–316.
- Propper, C. (2006, March). *Are public sector workers motivated by money?* (CHERE distinguished lecture). Retrieved July 2, 2010, from <http://www.chere.uts.edu.au/pdf/propper.pdf>

- Renn, J. (2002). *Challenges from the past. Innovative structures for science and the contribution of the history of science.* (Paper presented at the Max Planck Forum 5, Innovative structures in basic decision research, Ringberg Symposium, München.)
- Rinia, E.J., Leeuwen, Th.N., van Vuren, H.G., van Raan, A.F.J. (1998). Comparative analysis of a set of bibliometric indicators and central peer review criteria. Evaluation of condensed matter physics in the Netherlands. *Research Policy*, 27, 95-107.
- Roach, M., & Sauermann, H. (2010). A taste for science? PhD scientists' academic orientation and self-selection into research career in industry. *Research Policy*, 39, 422–434.
- Rost, K., & Osterloh, M. (in press). Opening the Black Box of Upper Echelons: Expertise and Gender as Drivers of Poor Information Processing. *Corporate Governance. An International Review*.
- Rothwell, P. M., & Martyn, C. N. (2000). Reproducibility of peer review in clinical neuroscience. Is agreement between reviewers any greater than would be expected by chance alone? *Brain*, 123, 1964–1969.
- Schimank, U. (2005). New public management and the academic profession: Reflections on the German situation. *Minerva*, 43, 361–376.
- Schmidt, R. (2010). Are Incentives the Bricks or the Building? *Journal of Applied Corporate Finance*, 22, 129- 136.
- Schreyögg, G., & Steinmann, H. (1987). Strategic control: A new perspective. *Academy of Management Review*, 12, 91–103.
- Schweitzer, M. E., Ordóñez, L., and Douma, B. 2004. Goal setting as a motivator of unethical behavior. *Academy of Management Journal*, 47, 422–432.
- Shanghai Jiao Tong University. (2007). Academic Ranking of World Universities. Retrieved July 2, 2010, from <http://www.arwu.org>.
- Simkin, M. V., & Roychowdhury, V. P. (2005). Copied citations create renowned papers? *Annals Improb. Res.*, 11(1), 24–27.
- Simon, H. A. (1985). Human nature in politics – the dialogue of psychology with political science. *American Political Science Review*, 79(2), 293–304.
- Simons, R. (1995). Control in an age of empowerment. *Harvard Business Review*, 73(2), 80–88.
- Simonton, D. K. (2004). *Creativity in Science. Chance, Logic, Genius, and Zeitgeist.* Cambridge, UK: Cambridge University Press.
- Singh, G., Haddad, K. M., & Chow, S. (2007). Are articles in “top” management journals necessarily of higher quality? *Journal of Management Inquiry*, 16, 319–331.

- Smith, R. (1997). Journal accused of manipulating impact factor. *British Medical Journal*, *314*, 463.
- Sokal, A. D. (1996). A physicist experiments with cultural studies. Retrieved July 2, 2010, from http://www.physics.nyu.edu/faculty/sokal/lingua_franca_v4/lingua_franca_v4.html
- Sorenson, O., & Fleming, L. (2004). Science and the diffusion of knowledge. *Research Policy*, *33*, 1615–1634.
- Spangenberg, J.F.A., Starmans, R., Bally, Y.W., Breemhaar, B., Nijhuis, F. J.N., van Dorp, C.A.F. (1990). Prediction of scientific performance in clinical medicine. *Research Policy*, *19*, 239-255.
- Speckbacher, G., Wentges, P., & Bischof, J. (2008). Führung nicht-erwerbswirtschaftlicher Organisationen: Ökonomische Überlegungen und Folgerungen für das Hochschulmanagement. *Betriebswirtschaftliche Forschung und Praxis*, *60*, 43–64.
- Starbuck, W. H. (2005). How much better are the most prestigious journals? The statistics of academic publication. *Organization Science*, *16*, 180–200.
- Starbuck, W. H. (2006). *The production of knowledge. The challenge of social science research*. Cambridge, UK: Oxford University Press.
- Starbuck, W. H. (2009). The constant causes of never-ending faddishness in the behavioural and social sciences. *Scandinavian Journal of Management*, *25*, 225–227.
- Staw, B. M., & Boettger, R. D. (1990). Task revision: A neglected form of work performance. *Academy of Management Journal*, *33*, 534–559.
- Stephan, P. E. (1996). The economics of science. *Journal of Economic Literature*, *34*, 1199–1235.
- Stephan, P. E. (2008). Science and the university: Challenges for future research. *CESifo Economic Studies*, *54*, 313–324.
- Stern, S. (2004). Do scientists pay to be scientists? *Management Science*, *50*, 835–853.
- Stokes, D. E. (1997). *Pasteur's Quadrant: Basic Science and Technological Innovation*. Washington, D.C.: Brookings Institution Press.
- Strathern, M. (1966). From Improvement to Enhancement: An Anthropological Comment on the Audit Culture. *Cambridge Anthropology*, *19*, 1–21.
- Swanson, E. (2004). Publishing in the majors: A comparison of accounting, finance, management, and marketing. *Contemporary Accounting Research*, *21*(1), 223–225.
- Taylor, M., Perakakis, P., & Trachana, V. (2008, June). The siege of science. *Ethics in Science and Environmental Politics*, *8*, 17–40.

- Teichler, U. (1988). *Changing Patterns of the Higher Education System. The Experience of Three Decades*. London, UK: Jessica Kingsley.
- The Thomson Corporation. (2008a). ISI Web of Knowledge Essential Science Indicators. Retrieved July 3, 2010, from <http://esi.isiknowledge.com/home.cgi>
- The Thomson Corporation. (2008b). ISI Web of Knowledge Journal Citation Report. Retrieved July 3, 2010, from <http://admin-apps.isiknowledge.com/JCR/>
- Thompson, J. D. (1967). *Organizations in action: social science bases of administrative theory*. New York, NY: McGraw-Hill.
- Todd, P. A., & Ladle, R. J. (2008, June). Hidden dangers of a “citation culture.” *Ethics in Science and Environmental Politics*, 8, 13–16.
- Tsang, E. W. K., & Frey, B. S. (2007). The as-is journal review process: Let authors own their ideas. *Academy of Management Learning and Education*, 6, 128–136.
- Turner, K. L., & Makhija, M. V. (2006). The role of organizational controls in managing knowledge, *Academy of Management Review*, 31, 197–217.
- Ultee, M. 1987. The Republic of Letter: Learned Correspondence, 1680–1720. *Seventeenth Century*, 2(1), 95–112.
- Ursprung, H. W., & Zimmer, M. (2006). Who is the “Platz–Hirsch” of the German economics profession? A citation analysis. *Jahrbücher für Nationalökonomie und Statistik*, 227, 187–202.
- van Fleet, D., McWilliams, A., & Siegel, D. S. (2000). A theoretical and empirical analysis of journal rankings: the case of formal lists. *Journal of Management*, 26(5), 839–861.
- Van Noorden, R. (2010, June 17). A profusion of measures. *Nature* 7300, 864–866.
- van Raan, A. F. J. (2005). Fatal attraction: Conceptual and methodological problems in the rankings of universities by bibliometric methods. *Scientometrics*, 62, 133–143.
- Viner, N., Powell, P., & Green, R. (2004). Institutionalized biases in the award of research grants: A preliminary analysis revisiting the principle of accumulative advantage. *Research Policy*, 33, 443–454.
- Weibel, A., Rost, K., & Osterloh, M. (2010). Pay for performance in the public sector – benefits and (hidden) costs. *Journal of Public Administration Research and Theory*, 20(2), 387–412.
- Weingart, P. (2005). Impact of bibliometrics upon the science system: Inadvertent consequences? *Scientometrics*, 62, 117–131.
- Weller, A. C. (2001). *Editorial peer review. Its strength and weaknesses*. Medford, NJ: American Society for Information Science and Technology.

- Wenneras, C., & Wold, A. (1999). Bias in peer review of research proposals in peer reviews in health sciences. In F. Godlee & T. Jefferson (Eds.), *Peer review in health sciences* (pp. 79–89). London, UK: BMJ Books.
- Willmott, H. (2003). Commercializing Higher Education in the UK: the state, industry and peer review. *Studies in Higher Education*, 28 (2), 129-141.
- Zehnder, E. (2001, April). A simpler way to pay. *Harvard Business Review*, 53–60.
- Zuckerman, H. A. (1992). The proliferation of prizes: Nobel complements and Nobel surrogates in the reward system of science. *Theoretical Medicine*, 13, 217–231.