Preface

Does Science Go Wrong?

Recently, the Economist (2013) stated on its front page "How science goes wrong," thereby calling attention to the discovery that much research is of poor quality. According to the Economist (2013, p. 11), "[a] rule of thumb among biotechnology venture-capitalists is that half of published research cannot be replicated." The Economist (2013) concludes that modern scholars trust published research too much and do not put enough effort into the verification of results. Nobel laureate Randy Schekman bemoaned the "toxic influence" (cf. Sample 2013) of the Impact Factor. Specifically, he emphasized that a high citation count does not necessarily indicate high-quality research; rather the citations might display an eye-catching, provocative topic or simply be wrong (Sample 2013). Consequently, Schekman declared that his lab would boycott the most highly ranked journals. In a similar vein, it has been criticized that "[t]ypically, a measure found to be ill-conceived, unreliable, and invalid will fall into disrepute and disuse among the members of a scientific community, remarkably this has not been the case with the impact factor [. . .]" (Baum 2011, p. 464).

Moreover, research often does not produce what is needed because of artificially created "competition without markets" (Binswanger 2011). In research, some key characteristics of markets are absent. In particular, there is (1) no unlimited market entry for suppliers and consumers, (2) low transparency, (3) no price that determines buying and selling decisions, and (4) no maximization of profits or benefits. Instead, research is characterized by production of public goods and fundamental uncertainty concerning outcomes. There exist no clear-cut measures for determining what research is good or bad.

Because measuring research quality is difficult, publications or citations are frequently used as a proxy for research quality. However, such quantitative indicators ultimately lead to a performance paradox (Frost and Brockmann 2014; Meyer 2005; Meyer and Gupta 1994; Osterloh 2010). Performance indicators lose

their ability to distinguish good and bad performance and lead to unintended side effects. Examples of such side effects include decreases in intrinsic work motivation, slicing of articles into as many publications as possible, and scientific misconduct (e.g., data fabrication). In addition, "[a]ny evaluation system in which the mere number of a researcher's publications increases his or her score creates a strong disincentive to pursue risky and potentially groundbreaking work [. . .]" (Alberts 2013, p. 787). Setting wrong incentives using indicators leads to useless outcomes and thus to deformation of research organizations. Consequently, research that has no relevance to the economy, society, and state is conducted. The Lancet reports that 85 % of research investment is misallocated and asks, "[. . .] how should the entire scientific enterprise change to produce reliable and accessible evidence that addresses the challenges faced by society and the individuals who make up their societies?" (Kleinert and Horton 2014, p. 2).

In summary, today, research organizations face many challenges: (1) false allocation of funding; (2) low reliability and reproducibility of results; (3) a focus on research quantity instead of quality, which leads to a performance paradox; (4) irrelevance of research; and (5) mainstream non-risky research.

The background of these challenges are the "audit explosion" (Power 1999, 2005) and the fundamental reform in the governance of research organizations that has occurred since 1990. A shift to so-called entrepreneurial universities and New Public Management was introduced, fueled by the idea to strengthen competition among scholars and to make researchers more accountable to the taxpayers (Stensaker and Benner 2013). In addition, a shift of the form of knowledge production from mode one to mode two was observed. Mode one knowledge production is defined as a disciplinary matrix with stable institutions, whereas mode two is characterized by blurring distinctions between different disciplines and also between science and technology (Gibbons et al. 1994).

In response to recent developments, and in particular to "Impact Factor games," in 2012, the American Society for Cell Biology and a group of editors and publishers of scientific journals initiated The San Francisco Declaration on Research Assessment (DORA). As of 2014, the DORA has been endorsed by 10,963 individuals and 484 organizations, including the Swiss National Foundation.[1] The central recommendations proposed by DORA are the following: (1) eliminate the use of journal-based metrics, such as the Impact Factor, (2) evaluate research content-based with own means, and (3) use the possibilities that arise from online publications and social media.

To conclude, a simple transfer of performance measurement methods used in private, for-profit organizations into the academic context—as suggested by New Public Management—is *not* suitable (Osterloh 2010).

The German Federal Ministry of Education and Research (BMBF) has realized the need for research concerning the reform of higher education and research governance at an early stage. BMBF launched a comprehensive research program

[1] Looked up on July 5th 2014.

titled "Research Regarding the Higher Education System" ("Hochschulforschung"). The overall goal of this research program is to provide usable and scientifically grounded knowledge to politicians and managers of higher education institutions (http://www.hochschulforschung-bmbf.de/1256.php). To achieve this overall goal, as part of this research program, BMBF launched different subprograms, such as "New Governance of Science," "Economics of Science,"[2] and, recently, the program "Performance Evaluations in Academia." In the wake of the program "New Governance of Science," Grande et al. (2013) published an insightful book titled "Neue Governance der Wissenschaft," which we consider to be an excellent starting point for our book. The authors address the changing relationship between government control and societal expectations regarding the science system as a whole. They state that the traditional governance system was characterized by a combination of high autonomy for professors and a high degree of academic self-organization within universities; however, it was also characterized by a detailed regulation by the state from outside. In contrast, New Public Management is characterized by hierarchical structures within the universities on the one hand and by comprehensive governance mechanisms from the outside (e.g., target agreements, university councils, and accreditation agencies) on the other hand. Our book complements their insights by focusing on internal governance, i.e., we primarily consider the opportunities and threats of New Public Management within research organizations.

Conceptual Basis

We define "governance" as coordination and regulation of different activities and interests of actors, organizations, or systems (Whitley 2011). According to control theory, we differentiate among three ideal types of governance mechanisms (Frey et al. 2013; Osterloh and Frey 2011; Ouchi 1977, 1979): (1) output control (i.e., markets or competition based on indicators),[3] (2) process control,[4] and (3) input control (i.e., socialization and selection into self-organized communities or peers). In addition, we distinguish two types of process control: bureaucratic control (command and monitoring) and peer control (control according to agreed-upon mechanisms regarding adequate procedures). In reality, governance is usually characterized by a mixture of these governance types. In the context of research organizations, New Public Management led, on the one hand, to a shift towards output control (i.e., performance indicators) and an attenuation of peer control and,

[2] Note that the research project that led to the publication of this book was funded as part of the program "Economics of Science" ("Wissenschaftsökonomie").

[3] Goal-oriented program ("Zweckprogramm") in the terminology of Luhman (1977).

[4] Conditional program ("Konditionalprogramm") in the terminology of Luhman (1977).

on the other hand, to a shift toward bureaucratic control and to an attenuation of input control.

As soon as collegial bodies are replaced by professional managers, indicators that measure output become important. In contrast with peers, professional managers do not possess the expert knowledge that is required for both peer and input control. Therefore, professional managers need indicators to be able to execute their governance functions internally and to ensure accountability to external agencies. "Governance by numbers" sets in, which implies both opportunities and risks.

Such opportunities and risks concern outward or inward effects. Positive outward effects include the following: first, the ivory tower breaks down, and research performance is made more visible to external actors and the public. Second, the willingness to provide state and private funding for research and the higher education system might increase. Third, higher interest and participation of laypeople in science ("citizen science") might result. Positive inward effects might result from more-objective and standardized performance criteria, which lead to increased fairness. "Old boys' networks," "club government," and sometimes unfounded claims to fame are weakened (Power 2005; Wenneras and Wold 2000, 2001). Second, transparency is increased, which enables more autonomy for the research organizations in relation to state governments (Hicks 2012; Lange 2008; Wissenschaftsrat 2011).

Negative outward effects of risks include the following: first, a diminished understanding of the process of scientific work, which is characterized by uncertainty of scientific outcomes, serendipity effects, a need for autonomy, and a lack of efficiency in favor of higher effectiveness, for example. Second, promotion of a "winner-takes-all" or "hit parade" mentality is favored. Only cutting-edge research is considered; research that is fundamental but is necessary for breakthroughs suffers as a result. Third, marketing, public relations, and rankings management might become paramount concerns. Negative inward effects might result from multitasking and lock-in effects (Osterloh and Frey 2014). Researchers, faculties, and universities increasingly orientate themselves by first considering visibility in terms of numbers (e.g., publication and citation rankings) rather than research content. This might lead to inadequate hiring or funding decisions. Second, the multidimensionality of academic performance is neglected, and researchers may focus only on citation numbers, thereby neglecting other (especially long-term) performance dimensions, such as teaching performance and third-mission goals. Third, the inherent risk of research outcomes is not sufficiently considered. Fourth, researchers may experience crowding-out or over-justification effects (Deci 1971). In other words, an intrinsically motivated "taste for science" may be replaced by an extrinsically motivated "taste for publication" or even a "taste for rankings" (Binswanger 2010; Osterloh 2010).

The following contributions in this book give an overview of the risks and opportunities of the shift toward more "entrepreneurial" research organizations. First, we consider performance management according to New Public Management as a whole. Second, it is asked to which extent the emphasis that New Public

Management places on output control is based on reliable performance measurement methods. Third, we pose the following question: what motivates researchers, extrinsic or intrinsic motivation, or both? The fourth part examines which conditions on the organizational and state levels foster or impede creative research. The fifth part considers Internet and new social media and asks whether the shift from production and dissemination of paper-based research to online activities changes knowledge creation and evaluation. The sixth part considers knowledge production in the business world. Could it be that "entrepreneurial" universities can learn something from real enterprises that differs from what New Public Management teaches us? Part seven collects diverse applied contributions, including case and country studies that illustrate nicely the advantages and disadvantages of recent reforms in research governance. The final part consists of the fairy tale "Cinderella," which teaches us to never trust tailored measures.

Governance of Research Organizations: Contributions of Our Book

Part I of the book considers the recent shift in research and performance management in the direction of output-oriented criteria and of strengthening the power of university management. On the whole, the authors present a critical view of this shift.

William H. Starbuck opens with the metaphor of research organizations as a disintegrating boat on a stormy sea. He focuses on research organizations in the United States, which often seem to pioneer developments in other parts of the world, and on the behavioral sciences. The sea is composed of universities that struggle to adapt to market pressures concerning, for example, internationalization, rating systems, and the spread of the Internet. The boat consists of the publishing industry, which is threatened by electronic open-access journals. The crew is the researchers, who are characterized by great difficulties in distinguishing among good, mediocre, and bad research. This difficulty is demonstrated, for example, by low inter-rater reliabilities in peer review processes. Nicolai et al. (2015) demonstrate that this problem is not restricted to the behavioral sciences. Starbuck encourages scholars to use such inconsistencies to gain more freedom from reviewers, editors, and governments, which, he argues, can be achieved.

Mathias Binswanger highlights the problem that competition in research is different from competition in the market for goods or services. Research produces common goods whose market value is characterized by high uncertainty. Often these goods are not marketable at all. Therefore, evaluating the performance of research using measurable outputs or orchestrating competition in research artificially produces fake competition without markets and unintended negative side effects. Artificial "competitions without markets" distract from the content of research and crowd out intrinsic motivation. For example, this effect occurs in

the form of competition for excellence, competition for funds, competition to be published, and competition for top positions in rankings.

Richard Münch analyzes changes within university management in the wake of "entrepreneurial universities" by comparing the US and German systems. In both countries, there is a growing dominance of university management of departments. However, in the United States, management concentrates on fundraising from private sources and—at least in the case of rich universities—lets the departments decide about research issues. In contrast, most German universities are publicly funded. To gain visibility and legitimacy in the eyes of the public and to impress external university councils, indicators such as third-party funds have become prominent. According to Münch, in Germany, this change has fuelled the so-called Excellence Initiative, large research clusters, and large-scale research collaborations. To achieve this aim, strategic management of universities, including tight control of departments, is strengthened. Self-coordination within the departments is weakened. The author demonstrates such a development with the example of the "Technische Universität München" (which calls itself an entrepreneurial university). He strongly disagrees with this development because, in his view, it undermines diversity and trust, which are preconditions of creative research.

Amanda H. Goodall and Agnes Bäker demonstrate how such dangers can be mitigated: by choosing excellent researchers instead of managers as leaders of the university. In general, in knowledge-intensive organizations, "expert leaders" have a deeper understanding of the core tasks of such organizations and higher credibility among their subordinates than do managers. They also hire better coworkers, create a more appropriate work environment, and are able to give better constructive feedback than managers. These advantages lead to more trust and trustworthiness within the organization and consequently to better performance. These insights caution against adopting a managerial leadership model in research organizations. Instead, it is argued that "expert leaders" can combine what in the research governance literature is often characterized as a conflict between the goals of the scientific community and the goals of an organization. Such a conflict is addressed in the next contribution.

Otto Hüther and Georg Krücken argue that in Germany, professors have traditionally enjoyed high autonomy and have strong linkages to their scientific community. In contrast, their linkages to the university in which they are employed are weak. High autonomy of professors is bolstered by the low personnel, organizational, and resource powers of the universities. The personnel power of universities is low because there exists an external market for professorships due to the ban on internal calls ("*Hausberufungsverbot*"). The organizational power of universities vis-à-vis professors is low because of lifetime tenure. Resource power is low because of the fixed-term funding that professors enjoy, which has been confirmed by law. However, universities today have to fulfill diverse tasks that do not always fit into the interests of professors, including high teaching duties and integration of women, minorities, migrants, and scholars from lower social classes. To strengthen the power of universities compared with the

power of professors, reforms, such as the introduction of internal tenure tracks, monetary incentives, and more powerful university leaders, have been undertaken. However, these measures may have unintended side effects. Professorships might become less attractive compared with research positions outside universities.

The final contribution in Part I by *Stefanie Ringelhan, Jutta Wollersheim, and Isabell M. Welpe* presents a literature review of performance and incentive systems. They also present the results of their own empirical research. According to the differentiation between output, input, and process control, they indicate which tools and evaluation methods are applied and linked to monetary and nonmonetary incentives. According to their own study, which was conducted across German-speaking business and economics faculties, output control prevails. In contrast with this result, young scholars expressed that they consider qualitative (process-oriented) criteria more important. Consequently, the authors recommend putting more weight on input control and qualitative feedback.

To summarize the contributions to Part I, there is consent among the authors that the present shift to more "entrepreneurial" research organizations bears more disadvantages than advantages.

Part II of the book addresses performance measurement in research, which is the basis of performance management. How can performance in academia be measured? Markets do not work in research because of the low marketability of research; thus, peer review is decisive (Arrow 1962; Merton 1973; Nelson 1971). Peer review is the backbone of all other types of research evaluation, including the number of publications in scholarly journals, the chance of obtaining funds, and positions in rankings. All these indicators are fundamentally based on peer review. Because publication in a highly ranked journal is often taken as an indicator of the quality of a manuscript,[5] journal peer reviews serve as gatekeepers to career progress, recognition within the scientific community, and income. In many countries, recent reforms have linked scholars' salaries to their numbers of publications in highly ranked journals. Therefore, the quality of performance management in research organizations depends on the quality of peer review.

Today, the most important form of peer review is pre-publication double-blind peer review. This type of review determines whether a scientific contribution will be published. In this review procedure, the author(s) and the—usually two to four—reviewers do not know each other.[6] Recently, this method has come under scrutiny. It is argued that such peer review lacks reliability, transparency, prognostic quality, and consistency over time (e.g., Campanario 1998; Starbuck 2003, 2006). The most important criticism is the lack of reliability, i.e., the lack of agreement between two or more reviewers on the quality of a manuscript. To examine this criticism, the contribution of *Alexander T. Nicolai, Stanislaw Schmal*,

[5] This assumption is very questionable (Baum 2011; Osterloh and Frey 2014; Osterloh and Kieser 2015).

[6] More precisely, the reviewers and authors are supposed to not know each other, which might often not be the case.

and Charlotte L. Schuster considers inter-rater agreements for five management journals and compares them with inter-rater agreements in chemistry and physics. The results demonstrate that in all subject areas, inter-rater agreement is low, even in highly ranked journals. It is not the case that in chemistry and physics, which are characterized by a relatively high consistency of paradigms, inter-rater reliability is higher than in management research, which encompasses multiple paradigms. These results indicate a certain randomness of the peer review process. They explain why type 1 errors (rejection errors) and type 2 errors (acceptance errors) are common (Engwall 2014).

Can the stated problems with double-blind peer review be mitigated using bibliometrics? Does aggregation of independent judgments compensate for individual reviewers' biases through error compensation and the inclusion of broader perspectives? Do bibliometrics represent the "collective wisdom" (Laband and Tollison 2003) of a scientific community? The contribution of *Stefanie Haustein and Vincent Larivière* gives an overview of bibliometric indicators such as the Science Citation Index, the Journal Impact Factor, and the H-index. The authors then highlight the misuse and adverse effects of bibliometrics, such as honorary authorship, "salami publishing," "impact factor games," and personal impact factors. Consequently, possible error compensation through the use of bibliometrics is balanced by the unintended negative consequences of bibliometrics. In particular, the authors argue against the view that high-impact journals always publish high-quality papers. This view is also questioned by the extremely skewed citation counts in journals, including high-impact journals (Baum 2011).

The biases of bibliometrics are carried into international university rankings, which suffer from additional problems. Using the examples of the Shanghai Ranking, Times Higher Education Ranking, and Taiwan Ranking, *Ulrich Schmoch* demonstrates that the position in these rankings heavily depends on sub-indicators that are not visible (e.g., the size of the university, its research vs. teaching orientation, and the age of the institution). The ranking positions are further biased by language effects and by a bias concerning social sciences and humanities. For example, a high position in the Shanghai Ranking can be expected when universities are large, when they focus on medicine and natural sciences rather than humanities and social sciences, and when they are located in English-speaking areas. Performance management should never be legitimated with such ranking positions.

Are there any performance indicators that are useful for performance management? The contribution of *Stefano Nigsch and Andrea Schenker-Wicki* discusses an approach that has been widely applied in the industrial analysis and public service domains, called Frontier Efficiency Analysis, which relies on Data Envelope Analysis (DEA). In contrast with the university rankings mentioned above, Frontier Efficiency Analysis always relates outputs to inputs and does not favor large universities with certain foci. It accounts for existing diversity in teaching and research tasks, for example, for scale economies, and for changes in efficiency over time. The authors demonstrate the usefulness of this approach using studies from

the United Kingdom, Australia, Germany, and Italy, and cross-country studies. They recommend using this approach in combination with other performance measures, in particular for comparing subunits of medical and social sciences departments, for example.

Part III of the book asks whether researchers can be motivated by monetary or nonmonetary incentives and which is the impact of the shift of traditional universities to "entrepreneurial" universities. Are researchers primarily intrinsically motivated to solve puzzles? What role do extrinsic incentives such as money and reputation play? Three empirical papers are presented to answer these questions.

Alice Lam bases her work on self-determination theory to explain the mix of monetary and nonmonetary incentives of researchers. Using in-depth interviews and questionnaires performed at five UK research universities, she identifies three patterns of motivation. The "reluctant commercializers" are motivated mainly by recognition and prestige as extrinsic incentives. Commercialization is at odds with their goals. The "committed commercializers" have fully integrated the norm of entrepreneurialism, i.e., they strive for money because of extrinsic reasons. However, they also derive intrinsic satisfaction from participating in commercial ventures as well as from solving puzzles. The "strategic commercializers" are intrinsically motivated in their commercial activities and see them as an extension of their research that satisfies their intellectual curiosity. Lam demonstrates that intrinsic and extrinsic motivations do not exclude one another. Researchers are motivated by a complex mix of incentives. It would be a mistake to consider commercialization to be at odds with creativity. However, it would also be a mistake to focus only on financial rewards to motivate researchers.

Uwe Wilkesmann also draws on self-determination theory. He asks whether the application of New Public Management in universities results in a contradiction between profession and organization, as Hüther and Krücken (2015) discuss in this book. He also asks whether the introduction of monetary incentives strengthens conflict between intrinsic and extrinsic motivation. He answers these questions on the basis of two surveys of German professors at research universities and at universities of applied sciences. He concentrates on academic teaching. He finds that the new steering instruments such as pay for performance and performance-related budgets do not increase the conflict between professional identity and organizational commitment; in fact, quite the opposite is true. Concerning motivation, he presents a crowding-out effect, i.e., a contradiction between extrinsic and intrinsic motivation, except for awards. These results give insight into two usually under-researched topics, namely teaching, and universities of applied sciences.

The third contribution in this section by *René Krempkow* considers in detail awards as incentives in academia. Awards are interesting because they are clearly extrinsic incentives but do not crowd out intrinsic motivation, as Frey and Neckermann (2008) and Frey and Gallus (2014) have demonstrated. The author discusses the potential of awards as incentives in comparison with income, the possibility to realize one's ideas, and the reputation gained by publishing in highly ranked journals. Based on survey research performed among selected medical

faculties, he finds that publication aspects (such as Impact Factors) have the highest importance for scholars but that awards are also perceived as important. To evaluate how the potential of awards can be exploited, the author takes stock of awards in Germany. He presents a great variety of 400 different awards and fellowships in the medical field. With respect to prize money, these awards range from less than one thousand to as much as five million euros.

Part IV analyzes under which conditions creativity in scientific research is fostered or impeded at the university and national levels.

Martin Quack considers the university level. He presents strong support for the importance of what we have called input control. He argues that numbers such as citation counts, Impact Factors, the Hirsch-index, and amount of research funding acquired should not be applied as criteria for hiring researchers. Rather, the most important criteria are that people are good department citizens, are good teachers, and have the potential to become great researchers. To determine the potential of a researcher, a group of competent and trustworthy experts should thoroughly examine every person and every research proposal. Such a time-consuming and costly evaluation minimizes the risk of severe errors. It allows academic freedom by enabling generous appointment contracts, which are the main precondition for creative research.

Gunnar Öquist and Mats Benner consider the national level. They compare the research governance conditions of two countries—Denmark and Sweden—that have otherwise similar conditions. Both countries are small and have predominantly publicly funded research systems and similar levels of public spending. However, the countries have different scientific impacts. According to relative citation rates, Denmark performs much better than Sweden. Although such indicators should be handled with care, they can be used to compare performance on a highly aggregated level. The authors discuss potential reasons for these differences. First, they find that Denmark is much more strongly oriented toward academic excellence on an international level, whereas Sweden has concentrated on local industrial and labor market needs. Its higher education institutions have developed such that all professional education is integrated into the university system. Second, universities in Denmark are centrally governed, whereas in Sweden, the autonomy of universities and departments has increased, but they have become dependent on external money. Third, in Denmark, recruitment is based on international competition. In Sweden, recruitment and career systems have been relaxed in terms of the required academic merits. Fourth, in Denmark, faculty positions are financed by budgets that are under the control of universities. In Sweden, budgets are oriented primarily toward research questions that are of strategic importance for the country.

Both contributions in Part IV demonstrate that on both the organizational and state levels, the university system needs, first, a high degree of internal autonomy versus external influences, and second, rigorous and thorough selection of scholars to foster creativity.

In Part V, we again consider the issue of performance measurement. It is asked which novel approaches to evaluate the quality of research exist. In particular, novel approaches that are closely intertwined with the Internet are discussed.

Katrin Weller presents an overview of webometrics or altmetrics. These are metrics that are based on user activities in social environments such as Twitter, ResearchGate, blogs, and other social media. Altmetrics can help locate interesting research and can tell something about the attention that such research has gained by counting bookmarks or downloads, for example. Altmetrics may uncover the impact of work that is not published in peer-reviewed journals. However, as soon as altmetrics are used for evaluations, "gaming of altmetrics" may occur. Therefore, altmetrics open new possibilities to attract attention, but whether they can be used to measure scholarly quality remains an open question.

Sascha Friesike and Thomas Schildhauer consider "Open Science," which concentrates on content rather than metrics. They trace the change in publication behavior from paper-based toward open and collaborative Internet-based behavior. In this manner, knowledge generation has become accessible to everyone. This change may constitute a cultural shift in how knowledge is created and disseminated. However, the incentive systems in science are still based on paper-based research. Making one's research public causes a social dilemma. For individuals, it would be rational to not share one's knowledge publicly until it is published in a highly ranked journal or is patented. There have been attempts to put incentive structures in place to promote Open Science. Such attempts include data sharing policies of journals and requirements that accompany project funding, such as Horizon 2020. However, as long as academic reputation is linked to publications in highly ranked journals, this social dilemma remains.

Christian P. Hoffmann analyzes social media in a communication framework and highlights that success measurement in universities is implemented primarily by analyzing scientific communication. The author asks how success measurement is altered by the use of social media. It is argued that new media broaden performance measurement opportunities, such as citations, bookmarks, and downloads. Although online metrics may also have their shortcomings, such as a filtered or selective view, social media provide insights into conversations and personal networks and thus enable a richer and more differentiated understanding of the university's communication success.

Margit Osterloh and Alfred Kieser consider the problems with peer reviews, in particular the problems with double-blind peer review that are addressed by Starbuck (2015) and Nicolai et al. (2015) in this book. They ask how use of the Internet could improve the peer review process. Peer review is—despite its problems—at the heart of scholarly performance evaluation and scholarly discourse. Today, double-blind peer review is considered a sacred cow, although there is considerable evidence that such reviews suffer from low reliability, low prognostic quality, low constituency over time, a lack of transparency, and other problems. The authors argue that inconsistency of peer reviews is not a problem but rather a source of valuable insight, as soon as there is an open review process that fuels scholarly discourse. Such discourse can be enabled by the Internet

in the form of open post-publication peer review. Such review strengthens open scholarly debate, enhances the quality of peer review, and makes peer review a part of the scholarly reputation system. Moreover, it impedes the opaque power of reviewers and ranking games.

Part VI shifts to the question of to what extent the governance of research organization can learn from other knowledge-intensive organizations. *Nancy R. Edwards and Berthold U. Wigger* draw insights from private businesses that New Public Management has disregarded: for-profit firms in the form of professional partnerships. These firms are not governed in a command-and-control manner. Instead, they are governed in the form of self-regulated communities that are characterized by input control (socialization and rigorous selection) and peer control (mutual process control). Traditional governance in academia shares a similar approach, whereas the "entrepreneurial university" disregards such insights by strengthening output instead of input and peer control.

Sven Grzebeta also considers knowledge-intensive companies. Based on an extensive literature review, the author addresses the following question: "What is the best approach to designing an effective incentive program for innovation in a knowledge-intensive business context?" He argues that a "one-size-fits-all" approach for incentivizing innovation is misleading and provides the following general recommendations for achieving a fit: (1) clearly define a program's goal, (2) strive for consistency between the program and existing policies, (3) implement fair and transparent rules, and (4) communicate intensively and provide feedback. He stresses that innovation programs should be designed in a manner that fosters creativity and enables support throughout the organization. Additionally, the author calls for adequate rewards that reflect the value added to the firm by submitting an idea as part of an innovation program.

Part VII of the book encompasses country and case studies and applied contributions regarding performance management and measurement.

The first contribution by *Thomas W. Guenther and Ulrike Schmidt* complements the conceptual contributions in this book by empirically investigating management control systems of 176 higher education institutions in Germany, Austria, and Switzerland. They consider management control systems as "formal, routine-based systems which help to maintain or alter organizational activities to increase efficiency and effectiveness." Based on a survey of heads of university management, the authors observe that management control systems and measures are used interactively and with medium intensity. Comparing the different countries, the authors find that the use of management control instruments is more intense in Austria and Switzerland than in Germany. Based on their empirical findings, the authors discuss potential implications for different stakeholders. They state that higher education institutions and their top management should focus on the question of how to use management control instruments. Institutions should develop procedures and techniques regarding how to supervise and consult higher education institutions to determine the optimal design and use of management control systems.

Focusing on one of the German-speaking countries, Austria, *Otto A. Altenburger and Michaela M. Schaffhauser-Linzatti* provide deeper insights into performance measurement at universities. The authors compare studies that apply ABIV (i.e., "Arbeitsbericht des Institutsvorstands") with studies that apply the ICR (i.e., Intellectual Capital Report) as reporting tools in an academic context. ICR-based studies differ from ABIV-based studies primarily in terms of the quantitative approaches that they apply. Based on the comparison, the authors state that the ICR includes "too many indicators and enforces excessive bureaucracy." They recommend that the ICR applied at universities in Austria should be either carefully revised or questioned. Both tools suffer from an inadequate data base for applying quantitative approaches (such as the DEA or fuzzy logic approaches). Overall, based on their Austrian case study, they conclude that performance indicators are not able to exactly measure performance. Information provided by such reporting tools can only serve as an approximation and has to be interpreted with care. The authors recommend not replacing established strategies with new ones unless there is an urgent need to do so.

Christoph Biester and Tim Flink conduct their case study in Germany. They provide interesting insights into professors' perceptions of organizational performance management techniques at a large German university. The authors combine interviews with an online survey. In general, they observe that the performance measurement system is perceived as positive. However, they highlight that the degree to which measurement (especially across different disciplines) is possible is often questioned. While acknowledging that most professors financially profit from the performance measurement system, the authors conclude that performance measurement has a "disciplining rather than a motivating effect on professors."

Also using a case study, *Ana I. Melo and Cláudia S. Sarrico* investigate performance management and measurement outside of the German-speaking region, namely in a high-performing Portuguese university. The authors ask the following questions: (1) How is performance measured? (2) How is performance information used and who uses performance information? (3) Where do pressures to measure occur, and in which way have performance management systems influenced the roles, influences, and accountabilities of key actors? The authors find that a fully developed performance management system is currently nonexistent. Regarding their first research question, they observe that the degree of performance measurement varies considerably among the different types of activities (i.e., teaching and learning, research and scholarship, and third mission) and among the different members of the university's governing bodies (i.e., students, academics, nonacademic staff, and external representatives). Concerning their second research question, they find considerable differences in how performance information is used. With regard to their third research question, they state that the pressure to implement performance management systems primarily comes from external stakeholders, in particular, from European policies and the Portuguese state. The authors highlight that scholars have to invest much time in performing bureaucratic tasks.

Matthias Klumpp introduces Data Envelope Analysis (DEA) as a potential approach to compare performance across disciplines.[7] DEA bears the advantage that multiple input and output indicators are taken into account without the need to determine the specific weightings of the indicators prior to the analyses. Based on a sample of 88 German universities and universities of applied sciences that include four distinct disciplines, the author demonstrates that DEA is a promising tool. It is able to compare performance across disciplines that differ, for example, with regard to the importance that is ascribed to journal vs. book publications. He finds that taking the disciplinary details into account is advantageous for small- and mid-size universities, for specialized universities, and for universities of applied sciences.

Also focusing on alternative research evaluation approaches, *Harald F. O. von Kortzfleisch, Matthias Bertram, Dorothée Zerwas, and Manfred Arndt* broaden the discussion of performance measurement by posing the following question: "to what extent do existing methods of research evaluation take into account specific Knowledge and Technology Transfer characteristics?" The authors stress that not only research output but also the "third mission" is worth considering. The authors start by providing two examples, one successful and one unsuccessful Knowledge and Technology Transfer. They demonstrate that evaluation by the market has to be complemented by evaluations by peers as soon as complex technologies are considered. Based on an extensive literature review, the authors first identify specific characteristics that should be considered in transfer-oriented research evaluations. Second, they analyze different research evaluation approaches for Knowledge and Technology Transfer. They come to the conclusion that the existing research evaluation approaches mainly neglect the importance of Knowledge and Technology Transfer.

The section ends with the contribution of *Tina Seidel,* which contributes to the topic of performance management from an educational science perspective. She presents different conceptual teaching and learning models and provides insights into the risks and benefits that result from assessments of teaching quality by students and instructors. Based on research findings, she concludes that in general, considering different perspectives for evaluating instructional performance (including the perspective of external experts) is beneficial. The author elaborates on professional development of university instructors (e.g., via the program "Learning to Teach") and discusses the importance of systematically training novices instead of utilizing "learning-by-doing" approaches.

In the final part of the book, Part VIII, *Rolf Wunderer* nicely demonstrates how the academic discussion regarding incentives and performance can be stimulated by fairy tales. Cinderella was chosen for marriage by the prince according to just one criterion: her shoe size. What would have happened if the doves did not give the signal "rucke di guck, rucke di guck, blood is in the shoe, the

[7] See also Nigsch and Schenker-Wicki (2015) and Altenburger and Schaffhauser-Linzatti (2015) in this book.

shoe is too small, the right bride is still at home"? The prince would have married a swindler. Rolf Wunderer and the entire book intend to play the role of the doves.

Outlook

Several fundamental changes have occurred in the last decade(s); thus, it is currently not possible to investigate the long-term effects. Failures of governance of research organizations might not surface immediately because research projects and their outcomes can take years to be publicized to their peers and other stakeholders and it can take years for (un)intended effects to become visible. Hence, our book aims to consider fundamental changes in the governance of research organizations and to portray evolving articles that reflect on New Public Management-style performance management. At best, we pursue an indirect, oblique approach in achieving our goal (Kay 2010) to analyze appropriate governance mechanisms in research organizations.

Open questions that arise from the chapters of this book include the following: (1) How can we improve performance evaluation and performance management in teaching? (2) What are the effects of New Public Management on applied universities? (3) What are the selection criteria for third-party funds? (4) Which new performance measurement opportunities might arise through the Internet and social media? What we know for certain is that there are many grievances in research governance and evaluation that remain a challenge in our future. Although innovative suggestions, such as open-access publication and review, are gaining popularity, quality assurance and acknowledgment remains an important topic for individual scholars and for research governance as a whole.

Acknowledgments

First, we would like to express our sincere gratitude to the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung). This book was published as part of a BMBF-funded project in a research program regarding the topic of "Wissenschaftsökonomie." The grant provided by BMBF facilitated editing the book and thus international publication of current research findings in the field.

Second, we would like to give special thanks to the project management agency in the German Aerospace Centre (DLR). From the project management agency in the German Aerospace Centre, Michael Kindt, Friedrich Neumann, and Michael Saager, in particular, facilitated our work and supported us in making our project a success.

Third, we thank all of the authors who contributed to this book for devoting their time and effort in preparing their book chapters. The authors thoroughly considered the comments that we sent out based on the editorial review and responded to them

within a very short time frame. Without the interesting and inspiring contributions that we received and without the willingness of our authors to cooperate during the publication process, the publication of this book would not have been possible.

Last, we would like to thank our student assistants, Carina Schmid and Stefan Körner, who were involved in, for example, formatting the book chapters.

Margit Osterloh
Jutta Wollersheim
Stefanie Ringelhan
Isabell M. Welpe

References

Alberts B (2013) Impact factor distortions. Science 340(6134):787

Altenburger OA, Schaffhauser-Linzatti MM (2015) The suitability of intellectual capital reports for the quantitative measurement of overall university performance. In: Welpe IM, Wollersheim J, Ringelhan S, Osterloh M (eds) Incentives and performance - governance of research organizations. Springer, Cham

Arrow K (1962) Economic welfare and the allocation of resources for invention. In: Nelson R (ed) The rate and direction of inventive activity: economic and social factors. Princeton University Press, Princeton, NJ, pp 609–626

Baum JAC (2011) Free-riding on power laws: questioning the validity of the impact factor as a measure of research quality in organization studies. Organization 18(4):449–466

Binswanger M (2010) Sinnlose Wettbewerbe: Warum wir immer mehr Unsinn produzieren. Freiburg, Herder

Binswanger M (2011) Sinnlose Wettbewerbe behindern Wissenschaft und Forschung. Cicero. Available via http://www.cicero.de/kapital/sinnlose-wettbewerbe-behindern-wissenschaft-und-forschung/41572. Accessed 6 June 2014

Campanario JM (1998) Peer review for journals as it stands today, part 1. Sci Commun 19(3):181–211

Deci EL (1971) Effects of externally mediated rewards on intrinsic motivation. J Pers Soc Psychol 18(1):105–115

Economist (2013) How science goes wrong. http://www.economist.com/news/leaders/21588069-scientific-research-has-changed-world-now-it-needs-change-itself-how-science-goes-wrong?spc=scode&spv=xm&ah=9d7f7ab945510a56fa6d37c30b6f1709. Accessed 31 Oct 2013

Engwall L (2014) On the quality of quality assessments. In: Blockmans W, Engwall L, Wearie D (eds) Bibliometrics: use and abuse in the review of research performance. Portland Press, London, pp 93–104

Frey BS, Gallus J (2014) Awards are a special kind of signal. Working paper no. 2014–04. CREMA – Center for Research in Economics, Management and the Arts, Zürich

Frey BS, Neckermann S (2008) Awards – a view from psychological economics. J Psychol 216:198–208

Frey BS, Homberg F, Osterloh M (2013) Organizational control systems and pay-for-performance in the public service. Org Stud:1–24

Frost J, Brockmann J (2014) When quality is equated with quantitative productivity – Scholars caught in a performance paradox. Zeitschrift für Erziehungswissenschaft (in press)

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. Sage, Thousand Oaks, CA

Grande E, Jansen D, Jarren O, Schimank U, Weingart P (2013) Die neue Governance der Wissenschaft. Zur Einleitung. In: Grande E, Jansen D, Jarren O, Rip A, Schimank U Weingart P (eds) Neue Governance der Wissenschaft - Reorganisation - externe Anforderungen - Medialisierung. transcript, Bielefeld, pp 15–45

Hicks D (2012) Performance-based university research funding systems. Res Policy 41(2):251–261

Hüther O, Krücken G (2015) Incentives and power: an organizational perspective. In: Welpe IM, Wollersheim J, Ringelhan S, Osterloh M (eds) Incentives and performance - governance of research organizations. Springer, Cham

Kay J (2010) Obliquity: why our goals are best achieved indirectly. Profile Books, London

Kleinert S, Horton R (2014) How should medical science change? Lancet 1–2

Laband DN, Tollison RD (2003) Dry holes in economic research. Kyklos 56:161–174

Lange S (2008) New Public Management und die Governance der Universitäten. Zeitschrift für Public Policy, Recht und Management 1:235–248

Luhman N (1977) Zweckbegriff und Systemrationalität: Über die Funktionvon Zwecken in sozialen Systemen. suhrkamp taschenbuch Wissenschaft, Frankfurt am Main

Merton RK (1973) The sociology of science: theoretical and empirical investigations. University of Chicago Press, Chicago, IL

Meyer MW (2005) Can performance studies create actionable knowledge if we can't measure the performance of the firm? J Manag Inq 14(3):287–291

Meyer MW, Gupta V (1994) The performance paradox. Res Organ Behav 16:309–369

Nelson RR (1971) The simple economics of basic scientific research. J Polit Econ 67:297–306

Nicolai AT, Schmal S, Schuster CL (2015) Interrater reliability of the peer review process in management journals. In: Welpe IM, Wollersheim J, Ringelhan S, Osterloh M (eds) Incentives and performance - governance of research organizations. Springer, Cham

Nigsch S, Schenker-Wicki A (2015) Frontier efficiency analysis in higher education. In: Welpe IM, Wollersheim J, Ringelhan S, Osterloh M (eds) Incentives and performance - governance of research organizations. Springer, Cham

Osterloh M (2010) Governance by numbers. Does it really work in research? Analyse und Kritik 2:267–283

Osterloh M, Frey BS (2011) Input control and random choice improving the selection process for journal articles. 4 Sept 2011. University of Zurich, Zurich, pp 1–34

Osterloh M, Frey BS (2014) Ranking games. Eval Rev (in press)

Osterloh M, Kieser A (2015) Double-blind peer review; how to slaughter a sacred cow. In: Welpe I, Wollersheim J, Ringelhan S, Osterloh M (eds) Incentives and performance - governance of research organizations. Springer, Cham

Ouchi WG (1977) The relationship between organizational structure and organizational control. Adm Sci Q 22(1):95–113

Ouchi WG (1979) A conceptual framework for the design of organizational control mechanisms. Manag Sci 25(9):833–848

Power M (1999) The audit society: rituals of verification. Oxford University Press, Oxford

Power M (2005) The theory of the audit explosion. Oxford University Press, Oxford

Sample I (2013) Nobel winner declares boycott of top science journals. The Guardian

Starbuck WH (2003) Turning lemons into lemonade. J Manag Inquiry 12(4):344–351

Starbuck WH (2006) The production of knowledge: the challenge of social science research. Oxford University Press, Oxford

Starbuck WH (2015) Issues and trends in publishing behavioral science: a quarrelsome crew struggling with a disintegrating boat on a stormy sea. In: Welpe IM, Wollersheim J, Ringelhan S, Osterloh M (eds) Incentives and performance - governance of research organizations. Springer, Cham

Stensaker B, Benner M (2013) Doomed to be entrepreneurial: institutional transformation or institutional lock-ins of 'new' universities? Minerva 51(4):399–416

Wenneras C, Wold A (2000) A chair of one's own. Nature 408(6813):647

Wenneras C, Wold A (2001) Nepotism and sexism in peer-review. Women, science, and technology. Routledge, New York, pp 46–52

Whitley R (2011) Changing governance and authority relations in the public sciences. Minerva 49(4):359–385

Wissenschaftsrat (2011) Empfehlungen zur Bewertung und Steuerung von Forschungsleistung. Halle

http://www.springer.com/978-3-319-09784-8

Incentives and Performance
Governance of Research Organizations
Welpe, I.M.; Wollersheim, J.; Ringelhan, S.; Osterloh, M. (Eds.)
2015, XXV, 488 p. 36 illus., 24 illus. in color., Hardcover
ISBN: 978-3-319-09784-8