Isabell M. Welpe · Jutta Wollersheim
Stefanie Ringelhan · Margit Osterloh
Editors

Incentives and Performance

Governance of Research Organizations

Springer

Incentives and Performance

Isabell M. Welpe • Jutta Wollersheim
Stefanie Ringelhan • Margit Osterloh
Editors

Incentives and Performance

Governance of Research Organizations

Springer

Editors
Isabell M. Welpe
Jutta Wollersheim
Stefanie Ringelhan
TUM School of Management
Technische Universität München
Germany

Margit Osterloh
Zeppelin University, Friedrichshafen, Germany, and University of Zurich
Switzerland

SPONSORED BY THE
Federal Ministry of Education and Research

Sponsored by the German Federal Ministry of Education and Research (BMBF) (grant no. 01PW11006). Any opinions expressed here are those of the authors.

ISBN 978-3-319-09784-8 ISBN 978-3-319-09785-5 (eBook)
DOI 10.1007/978-3-319-09785-5
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014955363

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

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

Contents

Part I
Performance Management

Issues and Trends in Publishing Behavioral Science: A Quarrelsome Crew Struggling with a Disintegrating Boat on a Stormy Sea

William H. Starbuck

Abstract Changes in societies and communication technologies are forcing universities and related post-secondary institutions to change quite significantly. Some of these institutions face threats to their existence; many institutions will have to adopt new structures and find new rationales to attract students and faculty. Universities have long relied on academic publishers to provide evaluations of research quality via editorial review, but the publishing industry has gone through tremendous changes and no longer resembles the industry that once provided these evaluations. Neither publishers nor universities have found a clear path toward future relationships. A central issue for researchers is their inability to agree about the quality of research. This disagreement arises partly from the complexity of research and research reports and partly from humans' limitations, and it creates great ambiguity for editorial reviews and personnel reviews. Recognizing the unreliability of evaluations, however, can free researchers to take more control over their professional lives and can make science work better.

1 Introduction

Academic researchers in the behavioral sciences resemble the crew of a disintegrating boat in a storm at night. However, the situation is even worse than it appears because the crew does not have a compass and its members disagree about where they want to go. To clarify this simile: the sea is composed of contemporary universities and high-level technical institutes, which are struggling to adapt to changes—both rapid and slow—in the societal demands on higher education and in communications technology. These changes may well cause many universities to dry up. The boat is the industry that publishes academic research, which has been losing money and consolidating into a very small number of surviving publishers that may soon deteriorate into mere websites. The crew is composed of researchers in economics, management, political science, psychology,

W.H. Starbuck (✉)
Lundquist College of Business, University of Oregon, Eugene, OR 97403-1208, USA
e-mail: starbuck@uoregon.edu

I.M. Welpe et al. (eds.), *Incentives and Performance*,
DOI 10.1007/978-3-319-09785-5_1

and sociology, who engage in endless debates among themselves about what to do, how to do it, and whether it is worth doing.

Should we despair over the sorry state of affairs, or should we rejoice that so many ritualistic impediments to good scholarship are falling away? The answers are in our hands.

The next three sections consider factors that are making the sea turbulent, causing the boat to fall apart, and impeding agreement among the crew. Then, two concluding sections propose actions by researchers, universities, and governments.

2 The Storm-Tossed Sea: Universities Trying to Adapt to Social and Technological Revolutions

Universities and schools are contending with social ferment and technological change, as they have been doing for many centuries. Idealistic students experiment with novel ideas and challenge traditions. Changes in sources of funding and in political environments ceaselessly perturb academic governance and policies. However, recent decades have forced universities to confront new kinds of economic, societal, and technological challenges and opportunities. As some universities respond to these challenges and opportunities more quickly or effectively than others do, some universities have grown stronger and richer while others have lost.

One of the largest swells in this sea has been the long-term rise of business schools (Starbuck and Baumard 2009). Before 1800, predecessors of modern universities were finishing schools for sons of the wealthy, who studied history, languages, philosophy and classic literature. The industrial revolution brought increasing demand for education in accounting and economics. However, elites viewed these subjects as vulgarly practical, so the first business schools appeared outside of universities. For example, business people and academic economists created a business school in Paris (l'Ecole Supérieure de Commerce) as early as 1819, but this school was a lonely innovator for over 60 years. When a more elite business school (l'Ecole des Hautes Etudes Commerciales) appeared in 1881, it was not created by an academic institution but by the Chambre de Commerce et d'Industrie de Paris. Elsewhere, a few universities began to offer business education soon after 1880, but this innovation spread slowly. Elite universities in the US did not offer business education until 1908 and they limited this topic to postgraduate masters programs.

The end of World War II brought a new flow of students, many of them veterans, who demanded courses with practical content, and the numbers of business courses and students began an exponential growth. In many, many universities, business education transitioned from a somewhat disreputable, marginal activity to a major source of revenues. By 1956, nearly 43,000 Americans per year were graduating from collegiate business programs, but by 2011, this figure had multiplied more than eight times to 365,000 per year, who constituted 21 % of the graduates (http://nces.ed.gov/fastfacts/). Nearly half of all American undergraduates take some business courses. Growth at the graduate level has been even more explosive. In 1956, only 3000 Americans per year were graduating from MBA programs, and by 2011, this figure had soared to more than 187,000 per year. To accommodate these students, American universities had to offer higher salaries to professors who would teach business subjects. This policy has subsequently expanded to other subject matters and has made salary differentials a significant issue for intramural politics. Yet, departments in the arts and humanities depend on subsidies from the revenues attributable to business students.

Business studies started to grow later across Europe. For example, graduate education in business did not appear in French public universities until 1955. French public universities have continued to pay equal salaries across fields, but this policy has meant that the most prestigious business programs are outside the public universities. Some of the private business schools have raised professorial wages to levels that are internationally competitive and they are striving for high international reputations.

Another current roiling the sea of universities has been the declining importance of tenure. Although the granting of tenure has been a very important symbol of achievement for many years, promises of lifetime employment greatly restrict the flexibility of universities. Economic fluctuations alter the numbers of students who want to study; sometimes these fluctuations are large and sudden. Economic and social trends alter the topics that students want to study. Tenured professors have been prone to teach topics that interest them personally even when these topics no longer attract the interest of other people. Thus, universities have gradually been decreasing the fractions of their faculties who receive tenure, or they have been hiring teachers with the explicit understanding that tenure is not possible. In some American universities, many professors now have fixed-term contracts—for, say, 5 years.

Countries that supervise universities via national governmental agencies have generally installed various types of evaluation systems. A few of these systems assign ratings to individual professors based on their research publications or students' evaluations of teaching. More of these evaluations focus on entire departments and schools, with serious implications for future budgets. These departmental evaluations take account of student demands as well as the professors' research and teaching.

Another cause of turbulence in the sea of universities has been the development of rating systems, especially for business schools, but also for universities more generally. Before general-circulation magazines began to publish ratings, the unclear quality differences between universities and degree programs received little discussion. Then, in the late 1980s, *Business Week* published a list of top business schools. A year later, *US News and World Report* published a rating of colleges and universities. Today, many periodicals produce ratings, as do some national governments.

These ratings have become powerful forces. Students prefer to attend more highly rated schools. Donors prefer to give money to more highly rated schools. Schools with higher ratings charge higher prices; their facilities are newer and more elegant; they pay higher salaries to professors; they compete more effectively for professors.

Academic administrators cannot ignore these publicized ratings and the flows of funds they influence. Senior administrators pressure deans and department heads to raise ratings. In the mid 1990s, deans and department heads began urging researchers to publish in the most prestigious, most visible journals. One reason was a perception that researchers from highly rated universities dominated publication in these journals; universities with lower ratings sought to imitate universities with higher ratings (Starbuck 2005). Another reason was that papers in prestigious journals draw more citations, which raise visibility and have value as components of rating systems. As employing universities have placed more emphasis on papers in prestigious journals, they have given researchers less credit for publishing books.

These pressures to publish in highly prestigious journals have democratized academic publishing. Researchers in the most prestigious universities have been losing their grip on the most prestigious journals, thus undermining to some degree the status system that administrators sought to imitate. In particular, Certo, Sirmon, and Brymer (2010) observed that the number of different researchers who published in the most prestigious Management journals nearly doubled in one decade, from 600 in 1988 to 1,000 in 2008. This democratization has also been making citations in academic journals less valuable as a way to burnish universities' reputations. Currently, academic administrators appear to be urging professors to engage in activities that draw the attention of mass media.

This increase in the numbers of researchers who try to publish in the most prestigious journals is also slowing academic careers. As more researchers have submitted papers to the most prestigious journals, the average time a Management researcher needs to publish five papers in the most prestigious journals has doubled, from 5.35 years in 1988 to 9.72 years in 2008 (Certo et al. 2010). At the same time, the significance of each citation has been declining. Everyone is pressing for more citations. Academic administrators want more citations. Researchers want more citations. Publishers and journal editors want more citations. Thus, citations have been multiplying. Average citations per paper have been rising more than 4 % per year, more than 50 % over the last decade. Some of this increase probably results from generally better communication about what papers exist, but researchers have also generated more citations by putting more references into their papers. Reference lists have been growing around 3 % per year. When an average paper cites more references, citations per paper increase.

The twin emphases on citations and ratings by mass media have increased the influence of wealth, the English language, and American research values. Universities need money to support purchases of books and journals and to build facilities that attract excellent students; fund raising has become a very important component of universities' operations. Many of the periodicals that issue ratings of universities appear in the English language, especially the periodicals that initiated the ratings. American universities have substantial budgets for university libraries and they employ many professors whose many papers account for a large proportion of the citations.

Yet another major swell in the sea has been internationalization of universities—both in actual locations and via the Internet. Universities began to establish transoceanic programs during the 1990s. These were often programs shared by two or three universities, so they allowed teachers to remain in their home locations; student groups traveled from site to site perhaps two or three times a year. Then, some universities created branch campuses outside their home nations and enrolled full-time students locally. This expansion has challenged and continues to challenge these universities to equalize the capabilities of students across the different campuses and to employ professors who are willing to teach in foreign lands, at least temporarily. Both issues have been fueling controversy within universities.

Internationalization has also been changing professional societies. Several new societies have sprung up, especially in Europe, and existing societies that did not have nationalistic names have expanded. In the early 1990s, over 90 % of members of the Academy of Management were Americans; the general perception at that time was that the Academy was an American organization although the Academy did not have "American" in its title. Then, new members began to join spontaneously from many nations. The Academy's organization did nothing explicit to attract these new members. Indeed, for several years, the Academy's leaders remained unaware that important changes were occurring; they did not even discuss the changing membership. Two decades later, the Academy has doubled in size, and its leaders are trying to accommodate issues posed by an international membership. Nearly half of the Academy's members are not Americans and they live in 115 nations. By contrast, several other professional societies that call themselves "American"—the American Accounting Association, the American Economic Association, American Finance Association—appear to have experienced no significant increases in nonAmerican membership.

Of course, the Internet has facilitated the internationalization of both universities and professional societies. It has also made it possible for researchers to collaborate over large distances and large time differences. It is actually practical for collaborators to work on a project nearly around the clock, with impressive effects on productivity. I know professors who reside in houses about 1,400 miles apart and who teach in two universities that are more than 1,200 miles from either house. They teach online, mainly at times of day that are convenient for them.

In 2012, a major event disrupted the sea of universities—Massive Online Open Courses (MOOCs). These activities appeared rather unexpectedly and spread very quickly. Many of the most prestigious universities in the world are participating. For a prestigious university, a MOOC is excellent advertising. The university places an outstanding professor on a world stage, simultaneously displaying the high quality of its teaching and demonstrating an altruistic commitment to learning, at small additional cost. Some other universities are trying to exploit the MOOCs by offering course credit based on examinations, even to students who are not enrolled at those universities. These activities enable the universities to attract potential students at absolutely no immediate cost and enable the universities to take advantage of the high-quality lectures offered online.

Most of this online education is experimental. Although some universities have been exploiting the Internet for a decade or longer, most universities have ignored these technological opportunities. They have been slow adopters, but now they face a serious competitive threat. There will be continuing demand for housing people between 18 and 22 who are driving their parents crazy, but is this demand large enough to sustain all of the existing universities? It seems that most universities are going to have to provide something better than lectures in large halls by professors of mediocre competence.

Cellular telephones are proving to be the most widespread, fastest disseminating technology in human history. Asia, Europe, Latin America, and the Middle East average more than one cellphone per capita. How will universities adapt to and exploit this technology? Very few universities have incorporated cell phones into their teaching activities.

3 The Disintegrating Boat: Academic Publishers Fighting to Survive in a World that No Longer Needs Them

Thirty years ago, academic publishing appeared to be a vibrant and lucrative industry. There were many publishers, ranging from large bureaucratic firms producing textbooks to small firms serving specialized niches. New startups emerged frequently, and the big firms often experimented with new imprints.

However, the industry was prone toward faddish imitation. Many, perhaps most, of the publishers' "editors" had little interest in or understanding of the content of their books. Rarely did they read the books they published. They concentrated on persuading authors to sign agreements that committed books to specific publishers but imposed no reciprocal obligations on the publishers. The editors' indifference to content disposed them to imitate each other. When one publisher seemed to have success with a particular type of book, other publishers rushed to sign contracts for similar books.

One of these fads turned the industry toward publishing more journals. During the first half of the 1980s, academic publishers launched many new journals, typically journals that focused on narrow market segments. Each of these journals had very few readers and attracted few citations. However, there were many new journals, so they collectively published many papers and attracted many readers. Academic authors and readers urged their university libraries to buy these journals. A few firms began to offer CD-ROMs that aggregated scores of journals. Later, these aggregated subscriptions shifted to the Internet, which allowed libraries to buy as many copies of each journal as would satisfy local demand. These bundles of journals have become extremely profitable, as university libraries have been willing to accept substantial price increases (Forgues and Liarte 2014). University libraries discovered that journals in electronic form are much cheaper to maintain than books on paper, so the libraries' budgets could cover more pages of text. Thus, the libraries cut their purchases of books to accommodate purchases of more journals.

I have been asking academic book publishers about their markets for over 30 years. In 1980, they said that about 900 libraries in the world had standing orders to buy every new academic book released by the highly prestigious publishers. By 2010, publishers were saying that the libraries with standing orders had dropped to 275. University libraries had transformed from providers of books to providers of journals, and since this reduced the demand for books, academic publishers published ever fewer books.

In 1960, the Institute for Scientific Information (ISI) began counting and publishing citation data. These data were taken from books and journals by humans, who made frequent data-entry errors; to find my citations, I had to examine half a dozen variations on my initials and family name, including erroneous ones. Then, in 1992, ISI was purchased by Thomson-Reuters Corporation. Thomson ISI automated the data entry, which eliminated nearly all of the data errors; and to facilitate this change, they stopped recording citations in books and began to count only citations in journals. Citations to books by journal articles remain in the data, but citations by books are no longer included. Thus, books lost a bit of their social influence.

Meanwhile printing technology was also developing, in a direction that has made it possible to produce new books in very small quantities without risking much money. However, most of these new books also yield small profits. Small publishers of printed academic books have vanished, and larger publishers have merged with each other in search of large sales volumes. Very few firms remain, at least as producers of books on paper.

Impacts of the Internet have been pervasive. Professors and university libraries no longer keep paper copies of journals on shelves. Retail distribution of books has gone strongly online. Textbooks remain very important sources of profit for commercial publishers, but printed textbooks have dim long-run prospects. Although some new textbooks are coming out as mix-and-match printed modules, other textbooks are merely numeric codes that give short-term access to online modules. Some universities are including online copies of textbooks in the enrollment fees for courses. More and more authors are publishing their textbooks through online sellers.

Many researchers are making their papers available through online databases that charge nothing (e.g., SSRN, ResearchGate). Commercial publishers are worried that researchers are distributing their published papers very freely via their universities' websites and online databases. In medicine, major US and UK funding sources are attempting to replace traditional journals with electronic open-access journals. Forgues and Liarte (2014) have reported that open-access journals currently number around 10,000 and that 68 % of them charge nothing.

The surviving publishers are searching for a place in the future. They are not struggling financially; they are continuing to make large profits from textbooks and journals (Forgues and Liarte 2014). However, publishers see pressures for change looming ominously; it is doubtful that the current streams of profit can continue. New journals are appearing daily and the total number of journals is exploding. Publishers and some professional societies are hoping to persuade researchers to pay for open-access publication, but it is doubtful that many academic researchers have money for such payments or are willing to make them. Since most of the benefit to academic researchers comes from the implied recognition from journal editors, researchers who pay for publication are damning their own papers. Because libraries are resisting further price increases, journal publishers are proposing to price based on usage; a journal that attracts few readers would cost libraries very little. If this policy becomes widespread, journals having large circulations will be much more valuable than those having small circulations. Commercial publishers say hopefully that they are planning for user growth in China, India, the Middle East, and South America.

4 The Stricken Crew: Behavioral Researchers Debating What to Do, How to Do It, and Whether It Is Worth Doing

The actions of behavioral researchers imply that they have great difficulty distinguishing between higher quality research and lower quality research. The reviewers for journals disagree with each other about the quality of manuscripts. I have found just 16 instances in which editors studied the agreement between reviewers and were brave enough to publish their findings (Starbuck 2005). For these 16 studies, the correlation between reviewers ranged from 0.05 to 0.37, and their mean was 0.18. The reviewers for journals also disagree with long-term citation behavior. Gottfredson (1978) calculated the correlation between reviewers' evaluations and later citations of published papers as being about 0.21–0.27. However, these correlations arise entirely from the highest quality manuscripts. Reviewers' judgments about the lower-rated 70 % of the manuscripts do not correlate at all with later citations, and for these manuscripts, there is no correlation between two reviewers (Starbuck et al. 2008).

These very low correlations imply that one reviewer's judgment about a manuscript does not predict a second reviewer's judgment or the number of citations the published paper will attract. Listening to reviewers does help editors avoid publishing the very worst manuscripts that the journal receives, but the reviewers' opinions discriminate very poorly between manuscripts that are average quality or above. The correlations between reviewers are so low that adding a third or fourth reviewer does not improve editorial selection. Although many editors believe that they have superior judgment, I have not found evidence to support this belief. Some evidence suggests that editors are more prone to erroneous judgments or biased judgments than typical reviewers.

I cannot quote large-scale statistics, but I believe that personnel evaluations are roughly as unreliable as editorial decisions. I have joined colleagues in two efforts to assess error rates by comparing promotion decisions with professors' post-decision achievements. In the first effort, we estimated the error rate as 33 % in promotion decisions at a not-very-prestigious business school; the wrong decisions were equally likely to have been 'do not promote' as to have been 'promote'. In the second effort, we estimated the error rate as 60 % in promotion decisions at three prestigious management departments; these wrong decisions were much more likely to have been 'do not promote' than to have been 'promote'.

These statistics testify to the difficulty of executing behavioral science. Humans are very complex systems. Dyads, groups, organizations, societies, economies, international bodies are even more complex. Historical events never repeat themselves exactly. Uncontrollable factors make it impossible to replicate an experiment exactly. Every situation involves many variables, some of them unobserved, and most of these variables correlate with each other to some degree, which creates a background of low-level statistical noise. Thus, study after study ends with researchers pointing out that they had obtained only partial answers to the questions they had initially hoped to answer; the initial experimental design had left important factors uncontrolled; there is need for future research to answer the new questions that the study had exposed. Indeed, in many, many cases, random combinations of variables or random changes in variables explain the observed data as well as the theories proposed by researchers (Denrell et al. 2013; Peach and Webb 1983; Levinthal 1991; Schwab and Starbuck 2012). Webster and Starbuck (1988) investigated 15,000 correlations published in three prominent management journals. These correlations had very similar distributions in all three journals, with means and medians close to 0.09. Researchers obtained larger samples when they were finding smaller correlations, and 69 % of the correlations were positive. When studying such data, researchers have very high likelihoods of finding statistically significant correlations. Researchers who choose two variables at random, taking no account of causal relations between these variables, have a 67 % chance of finding a statistically significant correlation. If the researchers examine three pairs of randomly chosen variables, they have a 96 % chance of finding at least one statistically significant correlation.

Behavioral researchers make their work even more difficult for themselves by claiming to have created new theories when they actually relabeled the variables in old theories (Webster and Starbuck 1988). Journal editors publish these restatements without comment. For example, in the seventeen and eighteenth centuries, mathematicians and economists created a theory of behavior based on the idea that choices maximize some kind of expected value. This idea is central to contemporary microeconomics, but it has also appeared in many different areas of behavioral research with different labels on the variables. For instance, in the latter part of the twentieth century, at least three researchers published theories asserting that effective "leadership" maximizes sums and products of variables that look very much like probabilities and valuations. Each of these innovators claimed credit for having proposed a new theoretical formulation.

Behavioral researchers also make their tasks more difficult for themselves by misrepresenting how they performed studies. Evidence indicates that a high percentage of published papers reflect HARKing—Hypothesizing After Results are Known (Bones 2012; Kepes and McDaniel 2013). Bedeian et al. (2010) found that 92 % of business professors knew a researcher who had "developed hypotheses after results were known". Mazzola and Deuling (2013) examined papers in seven industrial-psychology journals that had investigated 1,684 hypotheses: Researchers had claimed that data fully supported 73 % of these hypotheses and had partially supported another 15 %. Mazzola and Deuling also noticed that many researchers had tested unstated hypotheses, had failed to test stated hypotheses, and had drawn inferences about hypotheses that they neither stated nor tested. Because ex post theories inevitably claim to explain relations between randomly selected variables, they offer researchers weak bases for future theorizing.

The low correlations between two reviewers and between reviewers and citations also testify to the fallibility of human judgments. Because humans are simple creatures, they have difficulty understanding their complex environments. The Ashby's (1960) Law of Requisite Variety points out that for humans to understand their environments, their comprehension abilities must be as complex and diverse as their environments. Yet, humans find it virtually impossible to comprehend relationships that involve four or more variables (Box and Draper 1969; Faust 1984). Human minds try to classify almost everything into binary categories, and thereby, to ignore fine distinctions and to imagine weak relations to be strong. Humans see nonexistent phenomena that their minds say ought to be there (Singer and Benassi 1981) and they remember events that never occurred (Kiesler 1971; Loftus 1979; Nisbett and Wilson 1977).

Journal editors and reviewers demonstrate such biases in their evaluations of manuscripts. They award higher ratings to papers by researchers who work in prestigious universities, to papers written in the English language, to papers that include algebraic equations (even when the equations are irrelevant), and to papers that endorse the reviewers' own public statements (Ellison 2002; Eriksson 2012; Mahoney 1977, 1979; Nylenna et al. 1994; Peters and Ceci 1982). Mahoney (1977) observed that reviewers disguise their biases by making comments about research methods rather than about the substance of findings. Reviewers praised the methods when studies reported findings consistent with the reviewers' own previous findings, and reviewers pointed out methodological defects when studies contradicted the reviewers' own previous findings.

Evaluating research pushes human capabilities toward their limits because the manuscripts that report research are very complex stimuli. Gottfredson (1978), Gottfredson and Gottfredson (1982), and Wolff (1970) found that reviewers for psychological journals agree rather strongly with each other about the qualitative properties manuscripts should possess. They agree that manuscripts should exhibit good writing style; they should discuss interesting topics; they should present elegant theories; they should analyze relevant data, their statistical methods should be appropriate; and so forth. However, agreement between reviewers disappears when they judge specific manuscripts. Judging an actual manuscript forces reviewers to decide which of many properties are more important and how that specific manuscript scores on each of many dimensions.

5 Implications for Universities, Schools, and Governments

Disagreements among scholars about the value of research, teaching, and even ideas might imply that there is little to gain from aggregation or organization. Some profound intellectual contributions have come from people who worked alone. Yet, supportive social contexts have benefited scholarly work. Scholars need contexts that provide them with necessities such as food and shelter, and thinking usually benefits from discussion. Both large-scale and small-scale social contexts have sometimes been remarkably stimulating (Dizikes 2011). During the nineteenth century, Vienna spawned significant developments in economics, mathematics, medicine, music, philosophy, physics, and psychology, at least; this outpouring of ideas involved numerous “private seminars” that brought small groups of scholars together for discussions in their homes. During the 1960s, two dozen faculty and doctoral students at Carnegie Institute of Technology (later renamed Carnegie Mellon University) were extremely influential in revolutionizing research about business by inventing several new fields of study and redefining several others; daily periods of discussion played important roles in their efforts (Starbuck 2014).

Because examples such as Vienna and Carnegie seem to have demonstrated the potential value of social interaction, it seems curious that universities and schools in general make so little effort to create social interaction. Universities use class schedules as the primary mechanisms for coordinating the activities of professors. As long as professors appear in their assigned classrooms regularly without evoking student protests, administrators assume that the organization is functioning correctly. In many universities, professors inhabit their offices only during formally announced office hours, so their encounters with their professorial colleagues are infrequent and brief. Groups of professors assemble infrequently, and most of these meetings are ritualized presentations of research by visiting speakers or job candidates. It is rare that a university’s formal managerial system attempts to draw faculty and students together for regular and frequent discussions (not presentations) of new ideas, work in progress, or potential projects. Most universities and schools lack comfortable spaces for casual discussions.

Of course, most universities and schools have little reason to encourage intellectual development. Only a small fraction of universities draws significant revenue from research grants and technological innovations. Some universities have been operating in traditional ways for hundreds of years; they expect their reputations and established clienteles to continue to supply them more than adequately; and they bask in the prestige of alumni in high positions. Many universities are controlled by governments that place very little importance on knowledge development; they see education as intergenerational transmission of established beliefs. These governments provide minimal funding and measure accomplishment by the numbers of students per professor, so their universities have few opportunities to attract nongovernmental funds and they make little effort to attract more students because they already have too many.

Relations between intellectuals and governments have been troubled and unstable throughout human history. To erase the lingering influence of traditions and to stifle dissent, the First August Emperor of China attempted to kill every literate person and to destroy every written document. A panel of Athenians voted for the death of Socrates, probably because they saw his probing questions and sarcasm as challenges to the current government's legitimacy; and Socrates, who had the option of leaving town, may have proceeded to commit suicide in order to demonstrate his disgust over Athens' current rulers.

Involvement by governments in universities has often indicated ignorance by politicians as well as their efforts to control or influence ideas. Recent examples include beliefs about evolution, genders, global warming, racial differences, religion, and warfare. Some governments see education as a way to preserve the high statuses of elites or to suppress dissidents. Other governments try to use education to promote and consolidate political changes. Other governments announce policies toward education to symbolize their promises to make changes in the future. Still other governments see education as a means to greater tax revenues in the future. All of these expectations and goals have shifted abruptly. The one consistency has been that politicians have tried to control or influence universities, and education more generally, in pursuit of political agendas, not to improve scholarship or intellectual development.

Many universities, schools, and governments have had conflicts with scholars about voice and territory. A significant fraction of scholars, although certainly not all, believe they should be able to make their own decisions about what to say. They criticize governments' policies, they reveal information that their employers want to keep secret, they travel to foreign lands and form friendships with foreign people, and they lead or participate in public protests. Whereas all governments and most universities define themselves geographically, many scholars act as if geographic boundaries are irrelevant or detrimental. Physical scientists especially act as it the same physical laws operate throughout the universe, and social scientists tend to see differences between people or societies as mysteries crying for investigation.

The Internet and widespread social media are raising issues about the permeability of territorial boundaries, the audiences who hear about administrative actions, and the possibility of containing secrets. People everywhere are becoming better informed about events everywhere else. Language differences and societal customs are becoming less and less important. Such evolution is challenging the fundamental premises on which most universities and governments have defined themselves and is creating new competitions among them. There will be winners and losers. In domains such as technological innovations in products or production, the most successful actors have usually been those who moved second. First-movers have paid the costs of adopting new technologies that still required major improvements, whereas second-movers have avoided such costs yet they have innovated early enough to gain market share or to obtain price advantages.

6 Implications for Researchers

Stormy seas seem likely to extend far ahead. Human societies are changing rapidly and give no sign of stabilizing; communications technologies are changing at least as rapidly as societies. To survive, universities will have to adapt to both kinds of changes. Universities are receiving less consistent funding from governments and they are having to take more responsibility for generating revenues, so curricula may have to change more often than in the past. Practices that assume stability—such as tenure or long-term employment contracts—may become utterly impractical, at least for universities that lack large capital endowments. Professors may face much more frequent evaluations of their performance.

Over recent decades, American universities have hired a large underclass of poorly paid "contingent faculty" who work on short-term contracts, often on a part-time basis (Nelson 2010). According to the American Association of University Professors (www.aaup.org), contingent faculty comprise nearly half of all faculty and they are providing roughly two-thirds of all instruction. They usually teach basic courses. The "contingent faculty" are unionizing and demanding better working conditions, higher pay, and higher status.

Some professors complain that they tire of teaching the routinized, low-level courses. These complaints may become much less frequent, as MOOCs make it unnecessary for most universities to teach such courses. Students can obtain the basic content over the Internet, via either programmed texts or highly skilled lecturers, so local instructors will be limited to tutorials. Although some students complain that lecturers are inferior to online sources, more students say they prefer lecturers. However, students in online and lecture courses perform similarly on examinations.

Taking advantage of cellular telephones may prove to be a challenge to professorial ingenuity. Cell phones do not tolerate verbose messages or complex interconnections between ideas, and most human minds welcome simplicity. How can professors and universities counteract these twin pressures? At some elite business schools (at least), groups of professors have begun to discuss ways to create useful cell-phone "apps".

The disintegrating academic publishing industry seems to be a challenge for publishers but a boon for professors and the universe of learners. Already, many researchers are building new boats to transport their ideas and research findings via the Internet. Many authors of textbooks are distributing and advertising their books without the aid of commercial publishers.

A central issue is how universities can satisfy their needs for evaluation without publishers. In the twentieth-century economy, academic publishers obtained manuscripts at very low cost from professors who received no significant revenues from their research reports. Universities paid the professors to write these documents, paid professors to edit academic journals, and then bought the published books and journals. Many professors engaged in these activities because they genuinely love doing research. Many professors were flattered to be chosen as editors or reviewers, and universities acknowledged the value of these activities. Many other professors disliked doing research or found the processes of editorial evaluation very unpleasant, and when such people minimized their publication efforts, they became largely invisible to evaluation processes that focused on research papers in prestigious journals. Although some of these abstainers may take more interest in publishing research if they do not have to conform to the preferences of reviewers and editors, they remain an elusive challenge that universities have not addressed.

The evidence cited above indicates that universities have been placing too much confidence in the accuracy of editorial evaluations and in the quality gradations across journals. The most prestigious journals publish fewer of the very worst papers than do the least prestigious journals, but low prestige journals publish some of the most cited papers (Starbuck 2005). The highest quality manuscripts may be rejected by as many as five journals before seeing publication. Ratings of journal quality are largely public-relations facades, and they reinforce similarly fuzzy ratings of the universities.

It does researchers no good to bemoan the demise of long traditions or the ambiguity of research evaluations. Social and technological changes are inevitable. By participating in the adaptions to these changes, researchers may be able to steer change. No one intends to evaluate manuscripts poorly. Unreliable evaluations result from the complexity of the research processes and research reports and from human limitations.

Merely recognizing that research evaluations are unreliable can be liberating. Reviewers' opinions are only their opinions and the probabilities are very high that other reviewers will have very different opinions. Yet, many researchers attribute more value to reviewers' opinions than they deserve. By recognizing the unreliability of editorial review, researchers gain freedom to exert more influence over the fates of their manuscripts.

A large majority of manuscripts elicit invitations to revise, some more encouraging and others more negative. As editor of Administrative Science Quarterly, I observed that only about half of the researchers who received invitations to revise actually submitted revised manuscripts that differed noticeably from their earlier manuscripts. The other half of these researchers either submitted very superficial revisions or never resubmitted. Thus, a significant fraction of researchers were withdrawing from editorial evaluation. Although some editorial demands are unreasonable or ignorant or unethical, editors and reviewers do not intend to appear unreasonable or ignorant or unethical; they are just fallible humans struggling with very difficult tasks. Researchers should also keep in mind that they too are fallible humans struggling with very difficult tasks. The challenges from editors and reviewers often disclose problems with study design, or data interpretation, or writing style, and attending to these problems makes science work better.

Peter and Olson (1983, 111) urged behavioral scientists to look upon research as "the marketing of ideas in the form of substantive and methodological theories". Researchers need to win audiences for their work—to induce potential readers to read and to convince actual readers that ideas and theories are credible and useful. To achieve these goals, researchers need to tailor their manuscripts to the perceptual frameworks of potential readers. Feedback from editors and reviewers can provide useful information about the perceptual frameworks of potential readers.

Although researchers should value the inputs from editors and reviewers as information about the potential consumers of their research, they dare not depend on editors, reviewers, or colleagues to tell them what is right. The ultimate decisions about what to do must come from inside researchers themselves, expressing their own expertise, ways of thinking, and ethics.

References

Ashby WR (1960) Design for a brain, 2nd edn. Wiley, New York

Bedeian AG, Taylor SG, Miller AN (2010) Management science on the credibility bubble: cardinal sins and various misdemeanors. Acad Manage Learn Edu 9:715–725

Bones AK (2012) We knew the future all along: scientific hypothesizing is much more accurate than other forms of precognition—A satire in one part. Perspect Psychol Sci 7(3):307–309

Box GEP, Draper NR (1969) Evolutionary operation. Wiley, New York

Certo ST, Sirmon DG, Brymer R (2010) Competition and knowledge creation in management: investigating changes in scholarship from 1988 to 2007. Acad Manage Learn Edu 9:591–606

Denrell J, Fang C, Liu C (2013) The random school of thought in management. New York University; Warwick University, New York

Dizikes P (2011) The office next door. Technol Rev 114(6):12–15

Ellison G (2002) The slowdown of the economics publishing process. J Polit Econ 110:947–993

Eriksson K (2012) The nonsense math effect. Judgm Decis Mak 7(6):746–749

Faust D (1984) The limits of scientific reasoning. University of Minnesota Press, Minneapolis

Forgues B, Liarte S (2014) Academic publishing: past and future. M@n@gement 16(5):739–756

Gottfredson SD (1978) Evaluating psychological research reports: dimensions, reliability, and correlates of quality judgments. Am Psychol 33(10):920–934

Gottfredson DM, Gottfredson SD (1982) Criminal justice and (reviewer) behavior: how to get papers published. Crim Just Behav 9(3):259–272

Kepes S, McDaniel MA (2013) How trustworthy is the scientific literature in industrial and organizational psychology? Ind Organ Psychol 6(3):252–268

Kiesler CA (1971) The psychology of commitment. Academic, New York

Levinthal D (1991) Random walks and organizational mortality. Adm Sci Q 36:397–420

Loftus EF (1979) The malleability of human memory. Am Sci 67:312–320

Mahoney MJ (1977) Publication prejudices: an experimental study of confirmatory bias in the peer review system. Cogn Ther Res 1:161–175

Mahoney MJ (1979) Psychology of the scientist: an evaluative review. Soc Stud Sci 9(3):349–375

Mazzola JJ, Deuling JK (2013) Forgetting what we learned as graduate students: HARKing and selective outcome reporting in I–O journal articles. Ind Organ Psychol 6(3):279–284

Nelson C (2010) No university is an island: saving academic freedom. New York University Press, New York

Nisbett RE, Wilson TD (1977) Telling more than we can know: verbal reports on mental processes. Psychol Rev 84:231–259

Nylenna M, Riis P, Karlsson Y (1994) Multiple blinded reviews of the same two manuscripts: effects of referee characteristics and publication language. JAMA 272:149–151

Peach JT, Webb JL (1983) Randomly specified macroeconomic models: some implications for model selection. J Econ Issues 17:697–720

Peter JP, Olson JC (1983) Is science marketing? J Mark 47:111–125

Peters DP, Ceci SJ (1982) Peer-review practices of psychological journals: the fate of published articles, submitted again. Behav Brain Sci 5:187–255, includes 50 pages of comments by others and a response by Peters & Ceci

Schwab A, Starbuck WH (2012) Using baseline models to improve theories about emerging markets. In: Wang C, Bergh D, Ketchen D (eds) Research methodology in strategy and management, vol 7, East meets west. Emerald, Bingley, UK, pp 3–33

Singer B, Benassi VA (1981) Occult beliefs. Am Sci 69(1):49–55

Starbuck WH (2005) How much better are the most prestigious journals? The statistics of academic publication. Organ Sci 16:180–200

Starbuck WH (2014) Five stories that illustrate three generalizations about radical innovations. Indus Market Manag (forthcoming, published online August 23, 2014)

Starbuck WH, Baumard P (2009) Les semailles, la longue floraison et les rares fruits de la théorie de l'organisation. In: Rojot J, Roussel P, Vandenberghe C (eds) Comportement organisationnel, Tome III, chapitre 1: Théorie des organisations, motivation au travail. Engagement Organisationnel, De Boeck, Bruxelles, pp 15–58

Starbuck WH, Aguinis H, Konrad AM, Baruch Y (2008) Tradeoffs among editorial goals in complex publishing environments. In: Baruch Y, Konrad AM, Aguinis H, Starbuck WH (eds) Opening the black box of editorship. Macmillan, Basingstoke, pp 250–270

Webster EJ, Starbuck WH (1988) Theory building in industrial and organizational psychology. In: Cooper CL, Robertson I (eds) International review of industrial and organizational psychology 1988. Wiley, London, pp 93–138

Wolff WM (1970) A study of criteria for journal manuscripts. Am Psychol 25:636–639

How Nonsense Became Excellence: Forcing Professors to Publish

Mathias Binswanger

Abstract In current academic systems professors are forced to publish as much as possible because they are evaluated and ranked according to the number of their publications and citations in scientific journals. This "publish or perish"-principle results in the publication of more and more nonsense. This tendency can only be stopped by abolishing the currently pervasive competition for publication. In the past, researchers who had nothing to say were not incentivized to publish but nowadays they also have to publish continually. Non-performance has been replaced by the performance of nonsense. This is worse because it results in an increasing difficulty to find truly interesting research among the mass of irrelevant publications.

A number of perverse incentives are associated with the competition for publication. This includes strategic citing and praising, endless variation of already existing models and theories, and emphasizing formal and mathematical skills, while deemphasizing the content of a paper. Furthermore, in order to maximize the number of publications, scientists also try to squeeze out as many publications as possible from minor ideas (salami tactics), increase the number of co-authors, try to become ever more specialized in already highly specialized scientific disciplines and, in the most extreme case, just fake experiments and results. Engaging in all these activities is basically a waste of time as it does not foster the advancement of science. Instead, it crowds out the intrinsic motivation of professors and other scientists, which is essential for creativity.

M. Binswanger (✉)
School of Business, University of Applied Sciences and Arts of Northwestern Switzerland, Olten, Switzerland
e-mail: mathias.binswanger@fhnw.ch

I.M. Welpe et al. (eds.), *Incentives and Performance*,
DOI 10.1007/978-3-319-09785-5_2

1 Introduction: The Illusion of Promoting Efficiency by Setting up Competitions

Once upon a time it was believed that professors and scientists mainly engage in research because they are interested in understanding the world they live in, and because they are motivated to come up with new explanations and solutions. This was not always true but it was accepted not to tell a country's best academics which kind of research they should do (Kohler 2007; Schatz 2001). Academic work was not assessed systematically, as it was tacitly assumed that academics would strive for excellence without having to be forced to do so.

Today we live in a different world. To ensure the efficient use of scarce funds (which nevertheless are growing in the EU since 2001!),[1] the government forces professors and their academic staff, to continually take part in competitions, which are set up in order to promote "academic excellence" (Binswanger 2010, 2013). What caused such a drastic change in science policy? Why did universities forget about their noble purpose of increasing knowledge and degenerated instead into ranking-minded fundraisers and publication factories?

Ironically this degeneration is rooted in the now-fashionable and omnipresent search for excellence, where universities are supposed to outperform each other. Germany started an Excellence Initiative in order to boost its international competitiveness. Switzerland aimed to become one of the top five countries for innovation by supporting excellence (Gassmann and Minsch 2009). And the European Union, with the so-called Lisbon-strategy of 2000, had hoped to turn the EU into the most dynamic knowledge-based economy by 2010 (Lisbon European Council 2000). Amongst this childish race for excellence, it was overlooked that not everybody can be more excellent than everybody else. The fallacy of composition applies once more. Instead, the term 'excellence' became a meaningless catchword. The German philosopher Jürgen Mittelstrass (2007, p. 4, translated by the author) writes:

> Until now, no one took offence at the labeling of excellent cuisine, excellent performance, excellent academics or excellent scientists. [. . .] In the case of science this changed since science policy has occupied this term and talks about excellent research, clusters of excellence and excellence initiatives, in endless and almost unbearable repetitions.

Yet, how do we actually distinguish between an excellent and a not so excellent professor? In reality no one really knows, least of all the politicians who enthusiastically launch such excellence initiatives. But setting up competitions is supposed to solve the problem. It is assumed that competitions will automatically reveal the best researchers so it will not be necessary to care about neither content nor purpose of research. This illusion became prominent under the Thatcher government in England in the 1980s and then soon spread to other countries as well (Binswanger

[1] See Research and development expenditure, by sectors of performance (Eurostat Code: tsc00001).

2010, p. 44). The Thatcher government, guided by its faith in markets and competition, would have loved to privatize all institutions engaged in academic research and to let markets decide who is doing excellent research. However, this proved to be impossible. Basic research constitutes, for the most part, a common good which cannot be sold profitably on markets. Privatization would therefore crowd out basic research. Thus, as a second-best solution, competitions without markets were promoted, which nevertheless were termed markets (e.g., internal markets, pseudo-markets), even though this is a false labeling.

Related to the euphoria about markets and competition, there was also a constant suspicion regarding independent research taking place within "ivory towers", where scientists engage in such obscure activities as the search for truth. Consequently, the former British minister of education Charles Clarke characterized "the medieval search for truth" as obsolete and unnecessary (cited from Thorpe 2008). Modern universities are supposed to produce applicable knowledge, which has the potential to increase economic growth. Universities should think "entrepreneurially" and adjust to economic needs (see Maasen and Weingart 2008). For this reason, governments in many countries, and particularly in the EU, started to establish gigantic national or supranational research-funding programs. Instead of making funds directly available to universities, they have to compete for these funds, so that only the "best" get a chance to do more research. This is supposed to ensure that practice-oriented and applicable knowledge is created and that government funds are not wasted. Hence universities are forced to construct illusionary worlds of applicability and to pretend that all research serves a practical purpose (Körner 2007, p. 174).

Therefore, the major challenge is the question how to impress research commissions responsible for the distribution of funds in order to get additional funding. Mostly researchers try to impress by showing how many successful projects they did in the past, how many articles they already published and how much they were networking with other important scientists in the particular field. In this way, measurable "excellence" is demonstrated, which increases the probability of getting more funds as well. The assumption seems to be that our knowledge increases proportionally to the amount of scientific projects, of publications, and of networking activities, which in turn is supposed to lead to progress and growth. This naïve ton ideology is widespread among politicians and bureaucrats.

Consequently, modern universities are not focused any more on gaining knowledge. On the one hand, they became fundraising institutions determined to receive as much money as possible from government research-funding programs or private institutions. And on the other hand, they became publication factories, bound to maximize their publication output. Hence, the ideal professor is a mixture of fundraiser, project manager, and mass publisher (mostly as co-author of publications written by his or her assistants as he or she has no more time to do research), whose main concern is a measurable contribution to scientific excellence rather than increasing our knowledge. Moreover, in order to make sure that professors will deliver their contribution to excellence, faculty managers have been recruited for each department in addition to traditional deans. They act like CEOs in private companies and they are supposed to implement new strategies for becoming increasingly excellent. Research becomes a means in the battle for "market shares" of universities and research institutions (Münch 2009a, pp. 148–164).

Universities which on the surface expose themselves as temples of scientific excellence, constantly have to participate in project- and publication-contests, where instead of medals, winners are rewarded with elite or excellence status and, as far as professors a concerned, with exemption from teaching duties, and sometimes also with higher salaries (pay for performance). This is the case, notwithstanding the fact that many projects and publications do not have the slightest importance for people outside and often even inside the academic system. But these measurable outputs play a central role in today's research rankings, such as, for example, the German CHE Research Ranking of German universities (see Berghoff et al. 2009).

In this contribution we focus on the competition for publication. In fact this competition consists of two closely connected competitions, which are of crucial importance in the current scientific system:

The competition among scientists to get published in scientific journals, in which peer-reviews play a major role.

The competition for rankings based on publications and citations, which are important for individual scientists as well as for research institutes and universities.

Both kinds of competitions will be analyzed in more detail. It will be shown how they result in perverse incentives, which incentivize scientists to strive for excellence by engaging in nonsense activities.

2 Competing to Get Published: The Peer-Review Process

In almost every academic discipline, publications are the most important and often the only measurable output. Therefore, it seems to be straightforward to measure a scientist's output or productivity by the number of his publications. For is it not the case that many publications are the result of a lot of research, consequently increasing our relevant knowledge? Should not every scientist be driven to publish as much as possible in order to achieve maximum "scientific productivity"? The answer to these questions will be a clear "no", if you are familiar with the academic world. Indeed, more publications increase the amount of printed sheets of paper, but this number tells us as little about the relevance of a scientist's research activity than the number of notes played by a musician tells us about the quality of a piece of music.

Of course, measurements of scientific output are more sophisticated than just counting the written pages published by a scientist. Relevant publications are published in professional journals, where submitted papers have to go through the so-called "peer-review process". This should ensure that only "qualitatively superior" papers are published, which then can be considered to be "true scientific publications". Thus, strictly speaking, the competition among scientists is to publish as many articles as possible in peer-reviewed scientific journals.

However, there exist strict hierarchies among scientific journals, which are supposed to represent the average "quality" of articles published in these journals. In almost every scientific discipline there are a few awe-inspiring top-journals (A-journals), and then there are various groups of less respected journals (B- and C-journals), where it is easier to place an article, but where the publication does not have the same relevance as a publication in an A-journal. Publishing one's work in an A-journal is therefore the most important and often also the only aim of modern scientists, which allows them to ascend to the "Champions League" of their discipline. Belonging to this illustrious club makes it easier to publish further articles in A-journals, to secure more research funds, to conduct even more expensive experiments, and, therefore, to become even more "excellent". In this fashion, the "Taste for Science", described by Merton (1973), which is based on intrinsic motivation and which was supposed to guide scientists, is replaced by the extrinsically motivated "Taste for Publications" (Binswanger 2010, p. 150).

Looking at the development of the number of scientific publications, it seems that scientists are actually accomplishing more and more. Worldwide, the number of scientific articles has increased enormously. The number of scientific publications in professional journals increased from approximately 3,965,000 in the years from 1981 to 1985 to about 10,573,000 in the years from 2005 to 2009 (SBF 2011, p. 10), which corresponds to an increase by 270 %. The annual growth rate calculated on this basis was around 4.2 %. In the decade from 2000 to 2009 this growth rate even increased to 5.6 %. Therefore, the number of scientific publications grows faster than the global economy and significantly faster than the production of goods and services in North America and Europe, where the majority of publications is coming from SBF (2011, p. 11).

Once we begin to examine the background of this increasing flood of publications it quickly loses its appeal. A closer look reveals that the peer-review process is highly problematic. This supposedly objective procedure for assessing the quality of articles in reality often resembles a random process (Osterloh and Frey 2008). A critical investigation discovers a number of facts that fundamentally question the peer-review process as a quality-ensuring procedure (cf. Atkinson 2001; Osterloh and Frey 2008; Starbuck 2006). It generally appears that expert judgments are highly subjective, since the consensus of several expert judgments is usually low. One reason is that many peers, who are mostly busy with their own publications, will not actually read, let alone understand, the papers they are supposed to evaluate. There is not enough time for reviewing and usually there are also more rewarding things to do. Therefore, peers quite often pass the articles on to their assistants, who have to draft reviews as ghostwriters (Frey et al. 2009). No wonder that under such conditions important scientific contributions will often be rejected. Top-journals repeatedly rejected articles that later on turned out to be scientific breakthroughs and even won the Nobel Prize. Conversely, plagiarism, fraud and deception are hardly ever discovered in the peer review process (Fröhlich 2007,

p. 339). In addition, unsurprisingly, reviewers assess those articles that are in accordance with their own work more favorably, and on the other hand, are more likely to reject articles that question their own research (Lawrence 2003, p. 260).

Due to the just-described peer-review process, the competition for publication in scientific journals results in a number of perverse incentives (see also Anderson et al. 2007). To please reviewers, a potential author makes all kinds of efforts. To describe this behavior Frey (2003) rightly coined the term "academic prostitution", which—in contrast to traditional prostitution—is not the result of a naturally existing demand, but is induced by the forced competition for publications. In particular, the peer-review process is associated with the following perverse effects.

2.1 Strategic Citing and Praising

When submitting an article to a journal, the peer-review process induces authors to think about possible reviewers who have already published articles dealing with the same or similar topics. And they know that editors often consult the bibliography at the end of an article when looking for possible reviewers. Therefore, it is quite easy to guess, who the potential reviewers will be. To flatter them the author will preferably quote as many as possible and praise their work (for instance as a seminal contribution or an ingenious idea). Moreover, an additional citation is also useful for the potential reviewer himself because it improves his or her own standing as a scientist. Conversely, an author will avoid criticizing the work of potential reviewers, as this is likely to lead to rejection. Accordingly, this attitude prevents the criticism and questioning of commonly accepted approaches. Instead, it leads to replication of established knowledge through endless variations of already existing models and tests.

2.2 Sticking to Established Theories

In any scientific discipline there are some leading scientists who dominate their field and who often are also the editors of top journals. This in turn allows them to avoid publication of approaches or theories that question their own research. Usually this is not difficult, since most authors already try to adapt to currently prevailing mainstream theories in their own interest. The majority of the authors simply want to publish articles in top journals, and this makes them flexible in terms of content. They present traditional or fashionable approaches that evoke little protest (Osterloh and Frey 2008, p. 14). In this way, some disciplines (e.g., economics) have degenerated into a kind of theology where heresy (questioning the core assumptions of mainstream theories) is no longer tolerated in established journals. Heresy takes place in a few marginal journals specializing in divergent theories. But these publications rarely contribute to the reputation of a scientist.

Therefore Gerhard Fröhlich observes that (2003, p. 33) similar conditions prevail as in the Catholic Church: censorship, opportunism and adaptation to the mainstream of research. As a consequence, scientific breakthroughs rarely take place in peer-reviewed journals nowadays.

2.3 Impressing by Technicalities Instead of by Content

Since it does not pay off to question commonly accepted theories and research methods, authors have shifted their creativity to the development of increasingly sophisticated models and research methods. Simple ideas are blown up into highly complex formal models which demonstrate the technical and mathematical expertise of the authors and signal importance to the reader. In many cases, the reviewers are not able to evaluate these models because they have neither the time nor the motivation to carefully check them over many hours. Therefore, formal brilliance is often interpreted as a signal for quality and it helps to immunize authors from criticism. Peers, who are not working within the same narrowly defined research field just need to believe what insiders "prove" to be right in their complicated models.

By emphasizing formal aspects instead of content, sciences increasingly move away from reality Precision in fictitious models is more important than actual relevance. The biologist Körner writes (2007, p. 171, translated by the author): *"The more precise the statement [of a model], the less it usually reflects the scope of the real conditions which are of interest to or available for the general public and which leads to scientific progress."*

2.4 Undermining Anonymity by Building Expert Networks

In theory, the peer-review process should make sure that publication opportunities are the same for all potential authors. Both the anonymity of the authors and the reviewers are supposed to be guaranteed by the double-blind principle. But competition under such conditions would be too much of a hassle for established professors at top universities. After all, why did they work hard in the past and made a scientific career, if at the end they are treated like a newcomer, whenever they submit a paper to a journal? The critical debate on the peer-reviewed process discussed in the journal Nature in (2007) however showed that established scientists are "less anonymous" than other potential authors in the peer-review process. They know each other and are informed in advance which papers of colleagues and their co-authors will be submitted. In research seminars or informal meetings, they present new papers to each other, which successfully undermines anonymity of the peer-review process.

3 Competing for Top-Rankings by Maximizing Publications and Citations

Despite the great difficulties of publishing articles in professional journals, the number of publications is constantly growing and the number of journals is also increasing. Publications are important for rankings of individual scientists as well as of research institutions and universities. Furthermore, if young scientists apply for the post of a professor, the list of publications is usually the most important criterion in order to decide who will get the post. No wonder that scientists do everything to publish as much as possible despite the onerous peer-review process. The question what to publish, where to publish, and with whom to publish dominates the professional life of a modern scientist. Publication problems cause sleepless nights and the acceptance of an article in a top journal makes their heart sing.

But how does the number of publications actually get into the evaluation and ranking process of scientists and their institutions? At first glance, this seems quite simple: count all articles published by a scientist in scientific journals (or the number of pages) and then you will get a measure for the publication output. However, there is a problem. As was already mentioned, journals differ dramatically in terms of their scientific reputation, and an article in an A journal is supposed to be considerably "more excellent" than an article in a B or C journal. So we must somehow take into account the varying quality of the journals in order to achieve a "fairly" assessed publication output. To this end, an entirely new science has emerged, which is called scientometrics or bibliometrics. The only topic of this science is measurement and evaluation of publication output in other sciences. By now this new discipline has its own professors and its own journals, and consequently the measurements are also becoming more complex and less transparent, which in turn justifies even more bibliometric research.

A measure which has become particularly popular is the so-called "Impact Factor" (Alberts 2013). Nowadays this factor is commonly used in order to assess the "quality" of a journal. The Impact Factor of a particular journal is a quotient where the numerator represents the number of citations of articles published in that particular journal during previous years (mostly over the last 2 years) in a series of selected journals in a given year. The denominator represents the total number of articles published in that journal within the same period of time. For example, if a journal has an Impact Factor of 1.5 in 2013, this tells us that papers published in this journal in 2011 and 2012 were cited 1.5 times on average in the selected journals in 2013.

The Impact Factors used in science today are calculated annually by the American company Thomson Scientific and get published in the Journal Citation Reports. Thomson Scientific in fact became a monopolist in the calculation of impact factors, although the exact method of calculation is not revealed, which has been criticized repeatedly (see, e.g., Rossner et al. 2007). *"Scientists have allowed Thomson Scientific to dominate them"* (Winiwarter and Luhmann 2009, p. 1). This monopoly enables Thomson Scientific to sell its almost secretly fabricated Impact Factors to academic institutions at a high price, although in many sciences less than 50 % of today's existing scientific journals are included in the calculation (Winiwarter and Luhmann 2009, p. 1).

The competition for top rankings by maximization of publications and citations leads to additional perverse incentives, which can be experienced at almost every university and research organization. They are described in more detail below.

3.1 Using Salami Tactics

Knowing that the ultimate goal in current science is the maximization of relevant publications, researchers often apply the principle of "doing more with less", which has also been termed "salami tactics" (Weingart 2005). New ideas or records are cut as thin as salami slices in order to maximize the number of publications. Minor ideas are presented in complex models or approaches in order to qualify for an entire article. As a consequence, further publications can be written by varying these models and approaches. No wonder that the content of such articles gets increasingly irrelevant, meaningless, and redundant. Hence, it is becoming difficult to find new and really interesting ideas in the mass of irrelevant publications.

The most extreme form of a Salami tactic is to publish the same result twice or more. Such duplication of one's own research output is of course not allowed, but in reality often proves to be an effective way to increase one's "research productivity". As we have seen above, the peer-review process frequently fails to discover such double publications. Therefore, an anonymous survey of 3,000 American scientists from the year 2002 shows that at least 4.7 % of the participating scientists admitted to have published the same result several times (Six 2008). And in reality this percentage will probably be even higher as not all scientists will admit their misbehavior in a survey even if it is anonymous.

3.2 Increasing the Number of Authors per Article

It can be observed that the number of authors publishing articles in scientific journals has substantially increased over recent decades. For example, in the "Deutsche Ärzteblatt" the average number of authors per article has risen from 1 author per article in 1957 to 3.5 in 2008 (see Baethge 2008). This is, on the one hand, due to the fact that experiments have become increasingly complex and that they are no longer carried out by single scientists, but rather by a team. An evaluation of international journals showed that today's average number of authors per article in modern medicine is 4.4, which is the highest number of all disciplines. This is followed by physics with 4.1 authors per article. In psychology, the average is 2.6 authors per article, while in philosophy, still free of experiments, the average number of authors of an article is 1.1 (Wuchty et al. 2007).

However, the increase in team research is not the only reason for the constant increase in the number of authors per article. There is also the incentive to publish as much as possible and to be cited as often as possible. So, especially those who have some power in the academic hierarchy (professors or project leaders) try to use their power by forcing other team members to include them as authors in all publications of the research team. And the larger a team is, the more publications with this kind of "honorary authorship" are possible. Conversely, it is often advisable to a young scientist to include a well-known professor as a co-author because—also thanks to the lack of anonymity in the peer-review process (see Sect. 2)—this improves the chances of publication (see above).

The growing number of co-authors not only increases the publication list of the participating authors themselves, but also the number of direct and indirect "self-citations" (Fröhlich 2006), which triggers a snowball effect. The more authors an article has, the more all participating authors will also quote this article. "*I publish an article with five co-authors and we have six times as many friends who quote us*." (Fröhlich 2007).

3.3 Becoming More and More Specialized

To meet the enormous demand for publication, new journals for ever more narrowly defined sub-categories of a research discipline are constantly launched. Thus, the total number of worldwide existing scientific journals is estimated between 100,000 and 130,000 (Mocikat 2009), and each year there are more (Ware and Mabe 2012). By becoming increasingly focused on highly specialized topics chances for publication are improved (Frey et al. 2009). It often pays off to specialize in an exotic but important-sounding topic, which is understood only by very few insiders, and then to establish a scientific journal for this topic. Consequently, the few specialists within this field can promote their chances of publication by writing positive reviews for each other so that they will all end up with more publications.

Let us just take the topic of "wine" as an example: There is the "Journal of Wine Economics", the "International Journal of Wine Business Research", "Journal of Wine Research", the "International Journal of Wine Marketing," and so on. These are all scientific journals, which cover the topic of wine on a "highly scientific" level dealing with wine economics or wine marketing. It would not be surprising if soon we will also have specialized journals for red-wine economics and white-wine economics.

3.4 *Engaging in Fraud*

Last but not least, the competition for publications and citations also leads to fraud. "*The higher the pressure to increase productivity, the more likely it is to resort to doubtful means*" (Fröhlich 2006). The assumption that universities are committed to the search for truth (Wehrli 2009) becomes more and more fictitious. Modern universities are exclusively committed to measurable excellence and the search for truth often does not help much in this respect. No wonder that quite a few cases of fraud have recently been discovered.

A good example is the former German physicist Jan-Hendrik Schoen, born 1970, who was celebrated as a German prodigy in the beginning of the new millennium. For some time it was believed that he had discovered the first organic laser and the first light-emitting transistor, and, accordingly, he was highly praised and received a number of scientific awards. At the peak of his career, as a 31-year-old rising star at Bell Laboratories in the United States, he published an article every week, of which 17 were even published in highly respected journals such as "Nature" or "Science". No one seemed to notice that this is simply impossible if you do proper research. Instead the scientific community of Germany was proud to present such a top performer. But after some time, co-researchers doubted his results and it turned out that they mostly had been simulated on the computer. An interesting fact is, as Reich (2009) writes in her book "Plastic Fantastic" that these frauds would probably never have even been discovered, had Schoen not exaggerated so much with his publications. Otherwise, he would probably be a respected professor at a top university by now.

Cases of fraud such as the example of Jan Hendrik Schoen mainly occur in natural sciences, where the results of experiments are corrected or simply made up. In social sciences, however, empirical research has often reached a degree of irrelevance, where it does not matter anymore, whether results are faked or whether they are the "true outcome" of an experiment or a survey. They are irrelevant in one way or the other.

Conclusion

Forcing professors to publish by setting up competitions for publication, results in the production of more and more nonsense, which does not add to scientific progress. This is true in spite of the fact that the number of citations of articles also increases along with the number of publications. But the increase in citations of useless publications is not a sign of increased dispersion of scientific knowledge. Presumably, many articles even get quoted unread. This has been shown by research that documents how mistakes from the cited papers are also included in the articles which cite them (Simkin and Roychowdhury 2005). Therefore, more and more articles are published but they are read less and less.

(continued)

The whole process represents a vicious circle that can only be escaped by stopping competitions for publication in their current form. In the past, researchers who had nothing to say were not incentivized to publish. But today even uninspired and mediocre scientists are forced to publish all the time even if they do not add anything to our knowledge. Non-performance has been replaced by the performance of nonsense. This is worse because it increases the difficulty to find truly interesting research among the mass of irrelevant publications.

But the nonsense does not only concern the publications themselves. It also involves many corresponding activities, which are the result of the perverse incentives, set by the competition for publication. This includes strategic citing and praising, endless variation of already existing models and theories, and using mathematical artistry while not caring too much about the content of a paper. Furthermore, in order to maximize the number of publications, scientists also try to squeeze out as many publications as possible from minor ideas (salami tactics), increase the number of co-authors, try to become ever more specialized in already highly specialized scientific disciplines and, in the most extreme case, just fake experiments and results. Engaging in all these activities is basically nonsense, which does not foster the advancement of science.

Another serious consequence of the permanent competition for publication is the demotivating effect on professors and other scientists. Their intrinsic motivation is increasingly crowded out by the principle of "publish or perish". This principle replaced the "Taste for Science" (Merton 1973), which however is indispensable for scientific progress. A scientist, who is not truly interested in his work, will never be a great scientist. Yet exactly those scientists, who are intrinsically motivated, are the ones whose motivation is usually crowded out the most rapidly. They are often rather unconventional people who do not perform well in standardized competitions, and they do not feel like constantly being forced to publish just to attain high scores. Therefore, a lot of potentially valuable research is crowded out along with intrinsic motivation. But intrinsic motivation (Merton's Taste for Science) is a necessary condition for true excellence.

References

Alberts B (2013) Editorial: impact factor distortions. Science 340:787

Anderson MS, Ronning EA, De Vries R, Martinson BC (2007) The perverse effects of competition on scientists' work and relationships. Sci Eng Ethics 13:437–461

Atkinson M (2001) "Peer review" culture. Sci Eng Ethics 7(2):193–204

Baethge C (2008) Gemeinsam veröffentlichen oder untergehen. Deutsches Ärzteblatt 105:380–383

Berghoff S et al. (2009) CHE-Hochschulranking. Vorgehensweise und Indikatoren. In: Arbeitspapier Nr. 119. Available via http://www.che.de/downloads/CHE_AP119_Methode_Hochschulranking_2009.pdf
Binswanger M (2010) Sinnlose Wettbewerbe – Warum wir immer mehr Unsinn produzieren. Herder, Freiburg
Binswanger M (2013) More nonsense and less happiness: the unintended effects of artificial competitions. In: Brockmann H, Delhey J (eds) Human happiness and the pursuit of maximization. Springer, Heidelberg, pp 27–40
Staatssekretariat für Bildung und Forschung (2011) Bibliometrische Untersuchung zur Forschung in der Schweiz 1981-2009. Bern
Frey BS (2003) Publishing as prostitution? Choosing between one's own ideas and academic success. Public Choice 116:205–223
Frey BS, Eichenberger R, Frey RL (2009) Editorial ruminations: publishing Kyklos. Kyklos 62 (2):151–160
Fröhlich G (2003) Anonyme Kritik: Peer Review auf dem Prüfstand der Wissenschaftsforschung. Medizin Bibliothek Information 3(2):33–39
Fröhlich G (2006) Evaluation wissenschaftlicher Leistungen: 10 Fragen von Bruno Bauer an Gerhard Fröhlich. Schweizerische Gesellschaft für Strahlenbiologie und Medizinische Physik: SGSMP Bulletin. Available via http://www.sgsmp.ch/bullA62.pdf
Fröhlich G (2007) Peer Review und Abweisungsraten_Prestigeschmuck wissenschaftlicher Journale. Forschung und Lehre 6:338–339
Gassmann O, Minsch R (2009) Die Schweiz im Wettbewerb der Wissensgesellschaft: Was sind die Wirkungen des HFKG? Die Volkswirtschaft 9:30–32, Staatssekretariat für Wirtschaft (SECO), Bern
Kohler G (2007) Über das Management der Universität. Anmerkung zu einer aktuellen Debatte. Neue Zürcher Zeitung. Available via http://www.nzz.ch/aktuell/feuilleton/uebersicht/ueber-das-management-der-universitaet-1.538892
Körner C (2007) Die Naturwissenschaft im Spannungsfeld zwischen individueller Kreativität und institutionellen Netzen. In: Berka W, Schmidinger W (eds) Vom Nutzen der Wissenschaften. Böhlau, Wien, pp 169–181
Lawrence PA (2003) The politics of publication. Nature 422(6929):259–261
Lisbon European Council (2000) Presidency conclusions. Available via http://www.europarl.europa.eu/summits/lis1_en.htm
Maasen S, Weingart P (2008) Unternehmerische Universität und neue Wissenschaftskultur. In: Matthies H, Simon D (eds) Wissenschaft unter Beobachtung. VS Verlag für Sozialwissenschaften, Wiesbaden, pp 141–160
Merton RK (1973) The normative structure of science. In: Merton RK (ed) The sociology of science: theoretical and empirical investigations. University of Chicago Press, Chicago
Mittelstrass J (2007) Begegnungen mit Exzellenz. Jahreskongress der Schweizerischen Akademie für Naturwissenschaften. Available via https://www.yumpu.com/de/document/view/19550833/konstanz-begegnungen-mit-exzellenz-euler-2007
Mocikat R (2009) Die Diktatur der Zitatenindizes: Folgen für die Wissenskultur. Gaia 2(18):100–103
Münch R (2009a) Entkopplung, Kolonisierung, Zielverschiebung. Wissenschaft unter dem Regime des Exzellenzwettbewerbs zwischen Universitäten. Schweizerische Gesellschaft für Soziologie
Nature (2007). Peer review blog. Available via: http://blogs.nature.com/peer-topeer/2007/09/peer_review_and_scientific_con.html
NZZ (2004) Unruhe auf einem Spitzenplatz. Sonderbeilage zum Forschungsplatz Schweiz. Neue Zürcher Zeitung B3
Osterloh M, Frey BS (2008) Anreize im Wissenschaftssystem. Universität Zürich, Mimeo
Reich ES (2009) Plastic fantastic: how the biggest fraud in physics shook the scientific world, 1st edn. Palgrave Macmillan, New York

Rossner M, Van Epps H, Hill E (2007) Show me the data. J Gen Physiol 131(1):3–4
Schatz G (2001) How can we improve European research? Unpublished manuscript of the inaugural of the president of Switzerland's science council
Simkin MV, Roychowdhury VP (2005) Stochastic modeling of citation slips. Scientometrics 62 (3):367–384
Six A (2008) Schreibkrampf unter Forschern. Neue Zürcher Zeitung am Sonntag, 6 July 2008, p 67
Starbuck WH (2006) The production of knowledge: the challenge of social science research. Oxford University Press, Oxford
Thorpe C (2008) Capitalism, audit, and the demise of the humanistic academy. Workplace 15:103–125
Ware M, Mabe M (2012) An overview of scientific and scholarly journal publishing. The STM-report, 3rd edn. International Association of Scientific, Technical and Medical Publishers, The Hague
Wehrli C (2009) Das hohe Gut wissenschaftlicher Redlichkeit. Neue Zürcher Zeitung, 23 Sept 2009
Weingart P (2005) Impact of bibliometrics upon the science system: Inadvertent consequences? Scientometrics 62(1):117–131
Winiwarter V, Luhmann N (2009) Die Vermessung der Wissenschaft. Gaia 1(18):1
Wuchty S, Jones BF, Uzzi B (2007) The increasing dominance of teams in production of knowledge. Science 316(5827):1036–1039

Science in the Hands of Strategic Management: The Metrification of Scientific Work and Its Impact on the Evolution of Knowledge

Richard Münch

Abstract Following the global triumph of New Public Management (NPM), universities are required to improve their performance in research and teaching by strategic management. The entrepreneurial university striving for monopoly rents in terms of competitive advantages as regards the availability of money and prestige on the outside is coupled with the audit university on the inside, which tries to improve research and teaching by quality management. Both are complemented by the third-party funded university, which targets its research to recruiting third-party funds. This essay focusses on the non-intended effects counterproductive to progress in knowledge that are triggered by this transformation of universities in the context of the German science system's specific backdrop. The evaluation of the statute of a Technical University will serve as an example here concerning its quality management regarding the system of appointments and careers.

1 Strategic Management of Universities: The USA and Germany

The global movement that strives towards transferring the principles of strategic management of businesses to organizations, which genuinely do not produce any economic individual goods to be traded on the market, has long since reached the universities. It promises an increase of the universities' performance in research and teaching. Nevertheless, the fact that universities do not produce any individual goods for private use in the first instance, but instead assume a professional trusteeship for education and knowledge as collective goods, nurtures first doubts with regard to the promise of raising universities' achievements by way of strategic management. As a rule, the transfer of the principles of strategic management to

R. Münch (✉)
Otto-Friedrich University of Bamberg, Bamberg, Germany
e-mail: richard.muench@uni-bamberg.de

I.M. Welpe et al. (eds.), *Incentives and Performance*,
DOI 10.1007/978-3-319-09785-5_3

universities generates a double hybrid, which ensures that the intended effects do not arise. Instead, unintended effects appear that are counterproductive and involve worsening performance (Kieser 2010; Münch 2014), if we take the promotion of progress in knowledge and the widest and most profound education of university graduates as the uncontested yardstick. We will do this in the following sections by focusing on progress in knowledge as the primary target of science.

Strategic management creates a double hybrid as it overlaps, on the one hand, the functional requirements of professional trusteeship for a collective good. On the other hand, its own form is shaped by the culturally anchored institutions in the different national systems of higher education. We will use here the specific case of the introduction of strategic management at German universities. The statute for quality management of the Technical University Munich (TUM) related to its system of appointments and careers will serve as an example. The TUM document reveals how strategic management is translated into quality management on the inside.

Compared to the USA, above all, special features of the transfer of strategic management shine up in Germany, which were less evident in the USA up until recently, but have gained in significance in the meantime. In the USA, fund-raising, augmenting a university's assets by profitable investments and by investing funds in prestigious research has become a vital task of university managements due to the guiding role of the leading private universities and the dependence of both private and public universities on private sponsors. As a result, universities necessarily require professional management to ensure an effective administration of the available funds and, if possible, increase them further. Traditionally, however, the university management is faced by strong departments that administrate their research and teaching matters in a completely autonomous way in the USA. The appointment of new colleagues, their evaluation and promotion across the various career levels is completely in the hands of the department (Parsons and Platt 1973). From the appointment of an assistant professor right through to the last career step of a distinguished university professor, everything is in the hands of the department in question. Appointments are negotiated with the department chair and not with the university management. The university's strategic management does not have a direct impact on these affairs, and will influence them indirectly only in as far as a department recognized by the scientific community can expect the university management's financial support. However, the way in which achievements are consolidated is exclusively a matter of the department. As a result, strategic actions and quality assurance remain very close to the genuine demands of research and teaching. Strategic action implies, above all, the appointment of promising young scientists to junior faculty and of reputed colleagues to senior faculty. Quality assurance is geared toward a single person in his/her particular field and can, therefore, take the particular requirements, situation and performances into account in each case. It takes place in the decision on salaries and teaching loads and on the promotion to the next step of one's career immediately in the department.

Two changes have, however, increased the university management's influence on the departments in the USA, too, and have unbalanced their relationship: (1) the grown proportion of third-party funds in funding universities, and (2) the public power of rankings (even of privately established rankings such as the US News & World Report) along with the orientation of sponsors to rankings. Both changes imply that university managements link funding of departments to the latters' position in rankings more than before. Hence, the inevitable standardization and simplification of measuring performance in research and teaching by numbers increasingly affect the departments' strategic action and quality assurance from the outside. Consequently, the leeway for diversity—the crucial source of progress in knowledge—is restricted (Espeland and Sauder 2007, 2009; Sauder and Espeland 2009). In the USA, this development is discussed under the keyword "academic capitalism" (Berman 2012; Lorenz 2012; Münch 2011, 2014; Slaughter and Leslie 1997; Slaughter and Rhoades 2004).

The excesses of academic capitalism in the USA include especially the growing dominance of the university management over the departments. This is demonstrated especially by the exorbitantly rising staff numbers and top incomes of the top managers. Their managerial habitus has introduced the typical practices of businesses striving to position themselves on the market at the universities, too. Research and teaching are increasingly considered private goods that serve for making profit. This development is mirrored by the top incomes of top managers in universities, but also by the top incomes of "star scientists" or "global", that is globally active and visible professors (Ginsberg 2011). At the same time, an ever growing part of teaching is taken over by "adjunct faculty", i.e., low-paid teachers with temporary jobs. Up to 75 % of teaching in the BA curriculum is, in the meantime, covered by adjunct faculty with temporary appointment (Ginsberg 2011). The rising competition of so-called for-profit universities has contributed substantially to this process. It is claimed that they achieve billion-dollar profits with a business model that yields highest tuition fees at the lowest cost. This development is accompanied by high student loans that can often not be paid back and by drop-out quotas of 60 % and more (Frank 2013; Ginsberg 2011; Kempf 2014). These for-profit universities invest a mere 17.4 % of their budget into teaching, which usually consists of online courses. The highest amount of their income is allocated to management, marketing and profits. Another crucial element of the colonization of departments by the university managements' profit-making interests is the orientation of research to generate patents, whose number has risen tremendously since the Bayh Dole Act of 1980. This Act allows universities to cash patent yields generated with the help of federal funding themselves (Rudy et al. 2007; Washburn 2005). Studies show that knowledge and resources are no longer generally shared in research fields where this new entrepreneurial habitus has gained a foothold, but are shared with strategically selected partners only (Kurek et al. 2007; Shibayana et al. 2012). Ever since the universities in England have been mainly financed by tuition fees and research funds recruited in competition, a similar development can be observed there (Brown and Carasso 2013; Collini 2012, 2013; McGettigan 2013). The latest development is the increasing use of business finance methods in running the university, linking fees and loans directly to performance measures like work processes in business. In this respect, a kind of financialization of the university is emerging (McGettigan 2013). In the Netherlands, the reform act of 1997 and the transfer of real estate to the universities have also paved the way to their increasing financialization. Running the university requires loans and their hedging by derivatives, which call for an increasing number of finance managers in the administration. These managers introduce finance models into the university, implementing a strong governing management along with splitting up the university into internal markets with all its consequences of metrification of research and teaching (Engelen et al. 2014).

In Germany, this development towards an increased orientation to rankings and their both standardized and simplified measuring of performance by numbers set in with a delay of around 20 years, though against a completely different backdrop than in the USA. The task of winning private sponsors for fund-raising, of investing funds in a profitable way and of allocating them to the departments does not exist for German university managements. The necessary culture of private sponsoring on the part of affluent citizens is lacking here, i.e., sponsors who balance the state's low tax income when compared to the GDP according to their private ideas and thus determine in a private setting what public tasks will be carried out to what extent and in what way. In Germany, in contrast, a culture of public responsibility for public goods still reigns that is based on the idea of the culture state rooted in the nineteenth century, which shaped the exemplary foundation of the Berlin university in 1810 according to Wilhelm von Humboldt's organization plan.

Nevertheless, the entrepreneurial university is also making its way in Germany (Kieser 2010). In its recently published recommendations, the German Science Council puts enormous emphasis on improving the strategic capacities of the university management (Wissenschaftsrat 2013). Due to the lack of private fund-raising, the glance of a German university management, which is monitored from outside by rankings, is geared to serving the numbers used by these rankings. In this context, the acquisition of third-party funds holds first place in the indicators used not least of all due to its easy measurability. Because it is easier to collect, to standardize, to understand and to present the acquisition of third party funds to the public than other measures such as bibliometrics, altmetrics or peer review, it has become the dominating indicator of research performance in Germany. Hence, German university managements must interfere more in faculties and departments than their American counterparts, if they want to report visible achievements in acquiring third-party funds and to succeed in rankings. The Excellence Initiative of the Federal Government and the States has underlined this by the fact that the university managements were represented as applicants during the assessment of the applications. As not each type of research is suitable for acquiring third-party funds, this way of proceeding results in a one-sided focus on strongly third-party funded research and hence affects the diversity of research programs. The strategically planning university management tries to concentrate research in clusters so as to raise the opportunities for acquiring comprehensive funding for research collaborations that goes far beyond individual research. Research that cannot be integrated into a cluster will find only little legitimacy.

This interference of university managements into the matters of faculties and departments is made easier in Germany when compared to the USA by the fact that the professorial self-administration traditionally allows the senate to decide ultimately on all matters of research and teaching with the rector acting as primus inter pares. Appointment negotiations were held by the rector on behalf of the senate and in accordance with the ministry, which did not follow the university's proposal only in absolute exceptions, i.e., usually contested cases. The new laws on higher education in the individual states limit the senate's decision-making competence considerably, and with it the preceding power of the faculty councils. In appointment processes, they only have a consulting role and can make statements, but no longer possess any binding decision-making power. In contrast, the university management's decision-making competences were substantially extended both on the outside towards the ministry and to the inside towards the senate and the faculty councils. This process is linked with a strongly extended monitoring of the university managements from the outside, which is provided by the new university councils, but which is increasingly determined by rankings due to the latters' lacking proximity to research and teaching. Hence, due to its lacking possibilities of standing out by fund-raising, German university managements must interfere in the faculties and departments more than their American counterparts in order to maintain their standing in the struggle for visibility based on simplifying numbers. It also means that research is subjected to measures of producing visibility and to a kind of quality assurance that are in contrast with the genuine requirements of promoting progress in knowledge.

University managements' interests must be geared to establishing research collaborations that allow for large-scale recruitment of third-party funds to generate external visibility. The German Research Foundation (DFG) supplies the necessary funds with its coordinated programs, which unite 59 % of the funding budget (DFG 2012). The run to such programs, i.e., collaborative research centers, Ph.D. programs, research groups, excellence clusters and graduate schools is correspondingly strong, especially in the framework of the Excellence Initiative. Competition has become ever fiercer and granting quotas have dropped accordingly. The expense linked with pre-investments, the generation of critical mass in particular research fields and with application activities for third-party funds has increased enormously. The application and evaluation of research collaborations rob an increasing amount of time from the parties concerned which then lacks to carry out their own research. Along with the large-scale acquisition of third-party funds, the number of staff is growing who do research in a dependent position and with little opportunities to advance their careers. This trend goes hand in hand with a particular feature of the German university: the organization of research and teaching by chairs manned by a chair-holder and a series of assistants who do their research under his guidance, even long after their Ph.D. graduation. Scientific staff at German universities covers only 15 % professors, 17 % assistants with unterminated contracts, and 68 % assistants with terminated contracts. In the USA, we find 85 % professors—of these 24 % assistant professors—and 1 % assistants with unterminated contracts and 14 % assistants with terminated contracts (Konsortium

Bundesbericht wissenschaftlicher Nachwuchs 2013). However, we have to recognize here that the increasing financialization of the American university has put undergraduate teaching on the shoulders of a growing number of adjunct faculty with terminated contracts (Ginsberg 2011), who do not show up in these statistics.

Large-scale research in special research areas or excellence clusters keeps collaboration within the confines of the university to a large extent and neglects the fact that most often collaboration is needed across the boundaries of universities, as has been pointed out by Kaube (2013). Kaube refers to Peter Blau's (1973) classical study, which shows that research advances across university structures and beyond these in freely chosen cooperations. Strategic management, in turn, will encase it within the confines of the university. The vast amount of collaborative research along with the oligarchical structures in Germany sees to it that large part of research is not done freely and does not enter into cooperations freely beyond institutional boundaries, but is confined within strategic institutional concepts and is put under both external and internal control. We can talk here of a third-party funded university, which puts the chaos of free research under the control of external evaluation and internal hierarchy. Restricting the chaos immanent in research means that diversity will shrink as will the scope available for progress in knowledge. Diversity is important for possessing a rich pool of ideas, research programs and methods, which is nurturing any research that aims to overcome existing limits of knowledge. This is the major and still valid teaching of Paul Feyerabend's (1970/1993) plea for pluralism in science (see also Galis et al. 2012; Page 2010).

2 Tightening Control: Recent Developments in the Quality Management of Universities in the German Context

So far we have learned that the strategic management of US-American research universities has far more opportunities of fund-raising from private sources than the management of German universities and concentrates largely on increasing these funds and investing them in well performing departments and research centers. The top ranking universities have become extremely rich in this way. Rich universities can provide space for risky ideas, programs and methods and can be patient with regard to the progress made by their individual members. This is a major source of following new lines of research. The less rich universities, in contrast, forming the middle class of higher education are more strongly compelled to follow the mainstream and to exert tighter control on student and faculty performance in order to maintain their position. In comparison with the internationally leading elite of US-American top universities, German universities are middle class and, hence, have to apply similar measures of tighter control in order to secure international competitiveness. Because the German university management lacks the affluence resulting from private funds, it is particularly strongly driven to direct publicly funded research together with leading scientists at the cost of spontaneous developments within and across departments. It has to streamline research by organizing it in clusters and it has to control performance closely to guarantee success. This is why quality management is now increasingly tightening control at German universities as will be shown in this section.

In the German context, the translation of the university's strategic management into quality management subjects research to a particularly tight-meshed control system. This is proven by an assessment of the statute of the TUM's quality management with regard to the system of appointments and careers (TUM 2012). It is not a unique feature, but highlights the general trend spreading to all universities in a particularly striking manner. The statute establishes very tight control of research and teaching, from the university management down to the individual mentors whose job is to guide the appointed young scientists. TUM has, at least, dared a step forward and has gradually introduced a greater number of tenure track assistant professorships. It is intended to raise their proportion to around 30 % of the faculty. Nevertheless, the chair principle will not be abandoned. Assistant professorships simply serve as a first step on the way to a chair. Hence, this is not a real adjustment to the American department structure that does not know the chair system.

The university management's considerably wider interference in the appointment of new colleagues and quality assurance in the departments when compared to the USA is mirrored by the central position of the TUM Appointment and Tenure Board. It is charged by the university's executive board "with ensuring that the quality of tenure track appointment processes and tenure evaluations aligns with TUM's strategy across all subjects and faculties" and to see to it that "candidates selected dovetail with TUM's development plans" (TUM 2012, p. 78). The members of this central body include the managing vice president for research and innovation (chair), ten TUM professors with permanent contracts and an external member from the Max Planck Society. Apart from the vice president for research and innovation they are appointed for a 3-year term by the university's executive board and can be re-elected after that period (TUM 2012). All appointments and evaluations must be submitted to this central advisory body of the university's executive board. On the faculty level, appointments and assessments are in the hands of faculty search and evaluation committees. They not only initiate appointments, but also subject the appointed professors to thorough quality controls at all stages of their careers. Assistant professors are additionally accompanied by two mentors to whom they have to report every year. After 2 and 4 years each, a basic evaluation takes place, which decides upon the take-over to the next career level. If the evaluation turns out positive, the candidate will be taken over to the permanent position of an associate professor, usually after 6 years. It is in each case the TUM Appointment and Tenure Board, which gives recommendations on appointments and promotions to the next career step to the university's executive board that takes the ultimate decision. Hence, everything happens under the close supervision of the university's executive board that acts strategically to the outside and exercises tight control on the inside. The extent of these measures probably depends on the fact that the certainty needs of the legally trained German administration are transferred into the new system of New Public Management. Numbers now provide that kind of certainty which was guaranteed by watertight paragraphs in the past. It is estimated that the legally trained administration staff tends especially towards recovering lost certainties by numbers, target agreements and contracts.

Appointments and career systems guided by business administration are used by an ever increasing number of universities and seem to be mainly driven by the fear that scientists do not know what to do and might lead an easy life at the expense of the university budget and the tax payers. This view of research and teaching has been born from the spirit of McKinsey and is completely inappropriate for the genuine requirements of creativity in research. The fear of freedom issues from institution economics' principal-agent model, which business administration applies to ensure management effectiveness. The principal, who depends on the agent, is busy keeping the agent under control. This disagrees basically with the model of professional trusteeship based on the ethical obligation to the professional community that has so far marked the self-understanding of university professors.

Business-like quality management and its obsession with control and numbers (Balanced Score Card) has meanwhile been heavily contested in enterprises, too (Kieser 2000; Lawrence and Sharma 2002; Pandey 2005; Schneidermann 1999). It is inappropriate for supporting creativity in science (Amabile 1998). Studies reveal that creativity in science needs freedom from control, freedom to fail, because every achievement is based on learning from failures (Heinze et al. 2009; Hollingsworth 2006). When scientists have to report every year on what they have done from their first year to the most advanced levels, they can only learn strategic behavior of how to produce numbers, but cannot unfold actual creative strengths that will flourish only in an environment of genuine academic freedom. The new business-like quality management at the university breathes the spirit of a culture of mistrust, while trust should be essential for exercising trusteeship, which is at the core of scientific work. Little trust is paid to newly appointed young scientists when they have to deal with a wide range of qualifying accompanying programs and are guided by their mentors (TUM 2012). In this way, only subordinating normal scientists are reared instead of self-assured scientists who are confident to leave beaten tracks and deviate from the ruling thoughts.

There is an explanation why control is so crucial to the new appointment and career systems. If we assume the viewpoint of science for which it does not matter where new knowledge is created as long as it is considered a collective good (which is basically at risk in the case of patent-related research), neither much expense is needed in the search for the best brains (they will exist somewhere) nor for their control. It can be simply left to chance. At the same time, it guarantees the diversity of locations and paradigms. Failures at one place will be compensated for by commitment and progress at other places. It can easily be accepted that around 80 % of cited knowledge is generated by approximately 20 % of scientists. And the new control system will change nothing about this situation, just because of the limited capacity of paying attention to published research results of every scientist. It is also counterproductive to measure every scientist with the yardstick of high-performers. This can only result in broad paralysis among scientists with only average achievements. Nor is it reasonable to subject scientists dedicated to their work 60–80 h every week to increasingly tighter controls instead of trusting in the whole system of academic freedom that will create new knowledge somewhere, wherever this might be.

From the entrepreneurial, most extremely the financialized university's viewpoint, the world of science looks completely different. It cannot stay calm with regard to the place where knowledge is generated and scientists are active. To provide justification for its stakeholders, it must produce success figures year by year in as short-sighted a way as the management of stock-market companies, must supply successful appointments, more third-party funding, more publications, and more awards. It must enter a cut-throat competition on the outside, since the number of creative scientists is limited. And it must exercise control on the inside, since it cannot afford any failures. Nevertheless, the nature of science is based on many failures that are necessary to generate few achievements. If all universities act like enterprises, there will be no tolerance at all left for failures, everything must be kept under control. Carrying this consistently to conclusion, it would involve the end of curiosity and creativity in science. The only reason why this will not happen so quickly is that controls are not carried out everywhere to the same extent and that there is still underlife underneath the formal controls even in the worst of control systems as much as there is an underlife in any closed organization as Goffman (1961/1991) demonstrated in his famous study on asylums. To reassure the controllers, the desired keywords and figures are included in the organization's report to the controllers, but the liberty is reserved not to act in precisely that way. Anyway, the observers are not present when research is done. Sociology offers us the neoinstitutionalist explanation saying that many control measures are applied not because it is assured that they are effective and efficient, but because they provide legitimacy in view of globally ruling models (Meyer and Rowan 1977). We then have the opportunity to decouple actual practice from the external formal structure, but we also run the risk of seeing scientific practice colonized by the formal structure. According to Michael Power's (1997) *Audit Society*, this is a form of comforting. The control system tells the scientists what to do to gather scores and tells the controllers what to keep in mind when assessing research achievements. The problem arising here is that an imaginary world of controlled achievements is generated and we do not see what happens behind the scenes and how far creativity is restricted (Falk and Kosfeld 2006; Frey 1997, 2003, 2007; Osterloh and Frey 2008).

Ultimately, both those controlling and those controlled will suffer from an excessive control system, since its practicing robs them more and more of the time needed for research and since their thinking and acting is squeezed into confines that are deadly for the unfolding of creativity. The new control system requires professors to pursue a level of commitment as controllers that is not wholesome for their actual work. It starts with the double quality control in appointment processes on a faculty and university level and reaches right through to the annual controls of the progress made by young scientists. When will time be left for research?

When transferring the strategic management of American universities to the situation in Germany, we come across a far-reaching interference of university managements in research and teaching, which is exclusively in the hands of the departments in the USA. Summarizing the difference between the USA and Germany, we might say that the American university managements make big business in sponsoring and fund-raising and leave decisions on research programs more frequently to the departments, though NPM has empowered the management and weakened the departments in the USA, too, in recent times. In Germany, in contrast, university managements need the departments far more to be able to acquire third-party funds by large-scale research collaborations, which can then be presented to the outside as an achievement made by the university management.

Quality management in business-administrative terms involves excessive university monitoring and thus makes scientists first and foremost employees of the university working for the latter's benefit and prestige. In contrast, their loyalty to the scientific community, which is far more crucial for progress in knowledge, is pushed to the background. Nevertheless, it is still the scientific community that is the first evaluation instance to assess research, and not the university. This holds true at least as long as we are interested in promoting scientific progress. Well socialized scientists have internalized primarily striving for knowledge and, in the second instance, striving for recognition by the scientific community (Bakar 2012; Bourdieu 1975; Dunwoody et al. 2009; Frey 1994; Jacobsen and Andersen 2014; Jindal-Snape and Snape 2006; Osterloh and Frey 2008; Toren and King 1982). This ambition determines their habitus, and they are driven by the competition for progress in knowledge and recognition by the scientific community. It is underlined by citations, invitations, cooperations and awards as far as such features of achievement are not turned into standardized measure, which also have their own shortcomings. Interfering in this genuinely scientific process of quality assurance by way of business-administrative monitoring according to the principal-agent model would destroy the functionally required backdrop for the unfurling of progress in knowledge instead of supporting it. This also includes the common practice of decorating PowerPoint presentations for scientific lectures with an eye-catching logo of the university where research and teaching is done. In the hallowed halls of science not even this kind of advertising is appropriate, since it undermines the scientific community whose strength is required when it comes to fueling genuinely scientific quality assurance. We see here a growing conflict between the particular interests of the university management and the universalistic ethos of the scientific community, which is represented by the different specialist communities.

3 Why Science Needs Freedom to Fail to Advance Knowledge

After their master exams at the very latest, promising young scientists should be in a position to unfold their own ideas in their Ph.D. thesis, present them to their supervisors and acquire their Ph.D. degree. This applies even more for Habilitation or working as an assistant professor. As early as during the Ph.D. stage it should be clear to the scientists what they have to do to get ahead, since from the very beginning, orientation to the relevant research literature and the publication of their findings determine their work. This must be explored by the scientists themselves in early independence and supported by a *well-meaning* and *curious* but not *controlling* environment. This is how established professors should work with their young colleagues in order to be rewarded by very successful careers. We witness today that the younger colleagues are no longer familiar with this practice and are very frightened when they are unable to keep their young scientists on a short leash by way of Ph.D. contracts. Apart from the comprehensive gearing to the production of standardized peer reviewed journal articles, no farther reaching achievements of this practice are recognizable, which follows the principle of replacing mutual trust with contractual relationships. It is, however, crucial for a scientific career to ensure that nobody will give any orders to a young scientist of 30 years or more as regards the contents and way of his/her research and teaching. The only way of interfering can be agreement or disagreement in a scientific discourse among equals. In teaching bachelor or master courses, the crucial criterion is the students' understanding or not understanding of the subject matter in study courses, and their genuine feedback in the form of more or less successful achievements in written and oral exams. These are the genuine scientific yardsticks of good teaching, and not the satisfaction of customers in the standardized evaluation of study courses by the central quality assurance body. Of course, this academic freedom also involves dropouts, failures and withdrawal. However, orienting the entire system to avoiding such negative results will restrict science's potential to unfold, far too much to be justified.

How scientific knowledge is narrowed down to an overly homogeneous and one-sided perspective on reality by excessive quality control is demonstrated by the history of economics and its devotion to neoclassical model-building (Hodgson 2001; Hodgson and Rothman 1999). The financial crisis of 2008 has led many prominent economists to call for a return of economics from its splendid isolation and to open the doors to other disciplines like anthropology, psychology and sociology (Hodgson 2009). They see economics trapped in its model-building without contact to other disciplines. Institutional economics, behavioral economics, neuro-economics and most recently identity economics have paved the way for this recognition of the need for more diversity and collaboration with other disciplines. What is especially remarkable in our context is how much its dedication to quality assurance has driven economics into its splendid isolation. Internal and external controls have worked together in producing this splendid isolation. Striving for acknowledgement as a hard science that is good enough to be awarded the prize of the Swedish Riksbank in memory of Alfred Nobel every year since 1969 has paved the way for this development internally. In this way, scientific quality assurance has put objectification far ahead of creativity emerging from diversity. A study on the effects of the Research Assessment Exercise (RAE) and the work of the Quality Assurance Agency (QAA) in Britain comes to the following conclusion: “As suggested above, the RAE and the QAA combined with the anti-heterodox proclivities of mainstream economists produced the dominance of mainstream economics in British universities that is so visible today [. . .] Mainstream economists also used the RAE [. . .] promoting only a single paradigmatic view and eliminating dissenting voices.” (Lee 2007, p. 322; see also Lee et al. 2013). The peer reviewed journal article has become the vehicle of guaranteeing quality though strikingly low inter-rater reliabilities call the validity of this quality measure into question. The first step toward external control was counting the impact of journals by dividing the number of published articles by the number of citations. A further step was the use of such measures for decisions in hiring committees for professorships. Turning impact measures into rankings such as the ranking of economists and economics departments published by the *Handelsblatt* is the crucial step from internal to external control making all simplifying and misleading effects of ruling by numbers (Porter 1995) a new reality that guides research and publication strategies of scientists.

The reign of the impact factor and its stratification of the world of publications into A, B and C journals is largely responsible for the consolidation of economics as a hard science, but also for its running into the dead end of pure model building with little explanatory power when it comes to understanding and explaining the more complex nature of reality. Objectification has grown at the expense of diversity and creativity. This development has largely been driven by the superimposition of external on internal control. External needs for metrification called for by quality management have captured internal scientific quality assurance to lead economics into the dead end of pure neoclassical model building. In the meantime, there has been mounting critique of such negative effects of the impact factor’s reign (Campbell 2008; Falagas and Vangelis 2008; Kurnis 2003; Seglen 1997; Vauclay 2012). It narrows down science to reproducing established paradigms and methods, compels scientists from the beginning of their career to follow strategies of confirmation so that there is no chance of learning from failures, which is important for scientific progress. It rewards conservative publication strategies, and it is misleading because it attributes a majority of articles high significance though their own citation rate might be rather low, just because the high impact factor of the journal in which they have been published—resulting only from a minority of highly cited articles—serves as quality measure. The San Francisco Declaration (2012) wants to ban the impact factor from the evaluation of scientific work. It has been signed by many scientists all over the world. Leading scholars have published a widely supported open letter to mobilize the scientific community against the *Handelsblatt* ranking of management science (Kieser and Osterloh 2012). Nevertheless, the seizure of internal scientific quality assurance by external quality management will make it very hard to make this ban a reality. How much science needs freedom for diversity in order to advance knowledge, however, can still be learned from Paul Feyerabend's *Against Method* (1970/1993).

Conclusion

Added to the traditional structures prevailing at German universities, strategic management forms a hybrid that binds progress in knowledge in a novel way. The university rector as primus inter pares, who chairs appointment and retention negotiations on behalf of the senate and the faculty, becomes a president who exercises this right to a larger extent under his own guidance and has to report to the university council. At the same time, the academic bodies—senate and faculty council—have lost much of their decision-making competence. In this way, standardized and simplifying external evaluation tools provided by rankings become the crucial parameters of the university management's strategic management. In contrast, internal criteria close to research and teaching and their diversity lose in significance. An essential consequence of this implementation of strategic management at German universities is a more comprehensive interference of the university managements in the faculties and departments. This interference aims at establishing large-scale research groups and acquiring third-party funds and is pursued by target agreements and strategic appointments. This striving for collaborative research goes hand in hand with the German universities' oligarchic chair structure, which is still in stark contrast with the American department structure. Hence, large part of research takes place under the control of external evaluation and internal hierarchy. This will restrict the representation and institutionalization of diversity on the level of professorial positions. The entrepreneurial university's strategic management breaks down the chaos of research to calculable standard yardsticks. It cannot tolerate the normal chaos of science. For strategic management there is no space for failure. From its viewpoint it is unthinkable, for example, to see only 10 out of 100 research projects succeed. To control this risk, the entrepreneurial university subjects research to a seamless quality management, which sweeps away even the last crumbs of academic freedom and any leeway for creativity. Strategic management does not serve the progress in knowledge, but the increase of ostensible success figures behind which substantial losses are hidden. Scrutinized more closely, numbers of success such as the growth of third-party funding turn out to be nothing but an illusion serving to make the university management shine, which can thus justify its top salaries that are meanwhile common practice in the USA.

References

Amabile TM (1998) How to kill creativity. Harv Bus Rev (Sept–Oct):77–87
Bakar D (2012) Socialization of science – the acquisition of skills and values. Momentum 1(1), Article 7
Berman EP (2012) Creating the market university. How academic science became an economic engine. Princeton University Press, Princeton
Blau PM (1973) The organization of academic work. Wiley, New York
Bourdieu P (1975) The specificity of the scientific field and the social conditions of the progress of reason. Soc Sci Inf 14(6):19–47
Brown R, Carasso H (2013) Everything for sale? The marketization of UK higher education. Routledge, London; New York
Campbell P (2008) Escape from the impact factor. Ethics Sci Environ Polit 8(1):5–7
Collini S (2012) What are universities for? Penguin, London
Collini S (2013) Sold out. Lond Rev Books 35(20):3
DFG (Deutsche Forschungsgemeinschaft) (2012) Förderatlas 2012. Wiley, Weinheim
Dunwoody S, Brossard D, Dudo A (2009) Socialization or rewards? Predicting U.S. scientist-media interactions. J Mass Commun Q 86(2):299–314
Engelen E, Fernandez R, Hendrikse R (2014) How finance penetrates its other: a cautionary tale of the financialization of a Dutch University. Antipode 26 Mar 2014
Espeland WN, Sauder M (2007) Rankings and reactivity. How public measures recreate social worlds. Am J Sociol 113(1):1–40
Espeland WN, Sauder M (2009) Rankings and diversity. Southe Calif Rev Law Soc Just 18 (3):587–608
Falagas ME, Vangelis AG (2008) The top ten in impact factor manipulation. Arch Immunol Ther Exp 56(4):223–226
Falk A, Kosfeld M (2006) The hidden costs of control. Am Econ Rev 96(5):1611–1630
Feyerabend PK (1970/1993) Against method, 3rd edn. Verso, London
Frank T (2013) Academic fight song. Baffler 23
Frey BS (1994) How intrinsic motivation is crowded out and in. Ration Soc 6(3):334–352
Frey BS (1997) Not just for money. An economic theory of personal motivation. Edward Elgar, Cheltenham
Frey BS (2003) Publishing as prostitution. Choosing between one's own ideas and academic failure. Pub Choice 116(1–2):205–223
Frey BS (2007) Evaluierungen, Evaluierungen - Evaluitis. Perspekt Wirtsch 8(3):207–220
Galis ZS, Hoots WK, Kiley JP, Lanev MS (2012) On the value of portfolio diversity in heart, lung and blood research. Blood 120(12):2361–2364
Ginsberg B (2011) The fall of the faculty. The rise of the all-administrative university and why it matters. Oxford University Press, Oxford
Goffman E (1961/1991) Asylums. Penguin, Harmondsworth
Heinze T, Shapira P, Rogers JD, Senker JM (2009) Organizational and institutional influences on creativity in scientific research. Res Policy 38:610–623
Hodgson GM, Rothman H (1999) The editors and authors of economics journals: a case of institutional oligopoly? Econ J 109(453):165–186
Hodgson GM (2001) How economics forgot history. Routledge, London; New York
Hodgson GM (2009) The great crash of 2008 and the reform of economics. Camb J Econ 33 (6):1205–1221
Hollingsworth JR (2006) The dynamics of American science: an institutional and organizational perspective on major discoveries. In: Beckert J, Ebbinghaus B, Hassel A, Manow P (eds) Transformationen des Kapitalismus. Campus, Frankfurt; New York, pp 361–380
Jacobsen CB, Andersen LB (2014) Performance management for academic researchers. How publication command systems affect individual behavior. Rev Pub Pers Administ 34(2):84–107

Jindal-Snape D, Snape JB (2006) Motivation of scientists in a government research institute: scientists' perceptions and the role of management. Manag Decis 44(10):1325–1343
Kaube J (2013) Soziologiekolumne. Universität, Prestige, Organisation. Merkur 67(4):342–348
Kempf W (2014) Akademischer Kapitalismus. Süddeutsche Zeitung 70(3):11
Kieser A (2000) Die Balanced Scorecard als Managementmode. Kostenrechnungspraxis 44(2):123–124
Kieser A (2010) Unternehmen Wissenschaft? Leviathan 38(3):347–367
Kieser A, Osterloh M (2012) Warum wir aus dem Handelsblatt BWL-Ranking ausgestiegen sind – ein offener Brief an das Handelsblatt, unterzeichnet von Professoren der Betriebswirtschaftslehre. http://www.handelsblattranking.wordpress.com/2012/08/29/handelsblatt-ranking/. Accessed 30 May 2014
Konsortium Bundesbericht wissenschaftlicher Nachwuchs (ed) (2013) Bundesbericht wissenschaftlicher Nachwuchs. W. Bertelsmann Verlag, Bielefeld
Kurek K, Geurts PATM, Roosendaal HE (2007) The research entrepreneur: strategic positioning of the researcher in his societal environment. Sci Public Policy 34(7):501–513
Kurnis AP (2003) Understanding the limitations of the journal impact factor. J Bone Joint Surg 85 (12):2449–2454
Lawrence S, Sharma U (2002) Commodification of education and academic labour – using the balanced score card in a university setting. Crit Perspect Account 13(5):661–677
Lee FS (2007) The research assessment exercise, the state and the dominance of mainstream economics in British universities. Camb J Econ 31(2):309–325
Lee FS, Pham X, Gu G (2013) The UK research assessment exercise and the narrowing of UK economics. Camb J Econ 37(4):693–717
Lorenz C (2012) If you're so smart, why are you under surveillance? Universities, neoliberalism, and new public management. Crit Inq 38(3):599–629
McGettigan A (2013) The great university gamble: money, markets and the future of higher education. Pluto, London
Meyer JW, Rowan B (1977) Institutionalized organizations: formal structure as myth and ceremony. Am J Sociol 83:55–77
Münch R (2011) Akademischer Kapitalismus. Zur Politischen Ökonomie der Hochschulreform. Suhrkamp, Berlin
Münch R (2014) Academic capitalism. Universities in the global struggle for excellence. Routledge, New York; London
Osterloh M, Frey BS (2008) Anreize im Wissenschaftssystem. CREMA-Arbeitspapier, Zürich
Page SE (2010) Diversity and complexity. Princeton University Press, Princeton
Pandey IM (2005) Balanced scorecard: myth and reality. Vikalpa J Decis Mak 30(1):51–66
Parsons T, Platt GM (1973) The American university. Harvard University Press, Cambridge
Porter TM (1995) Trust in numbers: the pursuit of objectivity in science and public life. Princeton University Press, Princeton
Power M (1997) The audit society. Rituals of verification. Oxford University Press, Oxford
Rudy AP, Coppin D, Konefal J, Shaw BT, Eyck TT, Harris C, Bush L (2007) Universities in the age of corporate science. The UC Berkeley – Novartis controversy. Temple University Press, Philadelphia, PA
San Francisco Declaration of Research Assessment (2012) Putting science into the assessment of research. http://www.am.ascb.org/dora/files/sfdeclarationfinal.pdf. Accessed 30 May 2014
Sauder M, Espeland WN (2009) The discipline of rankings: tight coupling and organizational change. Am Sociol Rev 74(1):63–82
Schneidermann AM (1999) Why balanced scorecards fail. J Strat Perform Manag 2(56):11
Seglen PO (1997) Why the impact factor of journals should not be used for evaluating research. Br Med J 314(7079):498–502
Shibayana S, Walsh JP, Baba Y (2012) Academic entrepreneurship and exchange of scientific resources. Material transfer in life and material sciences in Japanese universities. Am Sociol Rev 77(5):804–830

Slaughter S, Leslie LL (1997) Academic capitalism: politics, policies and the entrepreneurial university. The Johns Hopkins University Press, Baltimore; London

Slaughter S, Rhoades G (2004) Academic capitalism and the new economy. Markets, state and higher education. The Johns Hopkins University press, Baltimore; London

Toren N, King J (1982) Scientists' orientation toward their work: the relative effect of socialization versus situation. Int J Comp Sociol 23(1–2):34–46

TUM (Technical University Munich) (2012) TUM Berufungs- und Karrieresystem. Statut zum Qualitätsmanagement, Munich

Vauclay JK (2012) Impact factor: outdated artefact or stepping-stone to journal certification. Scientometrics 92(2):211–238

Washburn J (2005) University Inc.: the corporate corruption of higher education. Basic Books, New York

Wissenschaftsrat (2013) Perspektiven des deutschen Wissenschaftssystems. Cologne

A Theory Exploring How Expert Leaders Influence Performance in Knowledge-Intensive Organizations

Amanda H. Goodall and Agnes Bäker

Abstract Leadership has been deemed, by some earlier scholars, to be less necessary in organizations that are knowledge-intensive. It has been assumed that because experts and professionals are driven largely by intrinsic motivation, extrinsic management and leadership factors are less important. We believe this assumption is wrong. Leaders have been shown in recent studies to have a considerable influence on organizational performance in universities, research institutes, hospitals and in high-skill sports settings. What matters, we argue, is the kind of leader. Experts and professionals need to be led by other experts and professionals, those who have a deep understanding of and high ability in the core-business of their organization. Our contribution will summarize the literature on the relationship between expert leaders and organizational performance, and then we will present a theory of expert leadership in a new model that outlines the possible transfer processes through which expert leaders generate better organizational performance.

1 Introduction

In 1996 Google was created by two Stanford University Ph.D. students, Larry Page and Sergey Brin, as part of their research. Although it has become a billion dollar empire, Google continues to promote its scientific credibility. Almost all members of the Google board hold at least one computer science or engineering undergraduate degree, Master's degree or Ph.D. There are two university presidents on the board—Stanford's John Hennessy and the former Princeton President Shirley Tilghman, both of whom are eminent scholars; and proud mention is made by the

The work presented in this chapter draws on earlier work by Goodall (2012) and Goodall (2014), among other non-technical articles.

A.H. Goodall (✉)
Cass Business School, City University London, London, United Kingdom
e-mail: amanda.goodall.1@city.ac.uk

A. Bäker
Lehrstuhl für Personal und Organisation, Eberhard Karls Universität Tübingen, Tübingen, Germany
e-mail: agnes.baeker@uni-tuebingen.de

I.M. Welpe et al. (eds.), *Incentives and Performance*,
DOI 10.1007/978-3-319-09785-5_4

founders of their links with the National Academy of Engineering and other scientific bodies.

Putting scientists on corporate boards may seem counterintuitive to some; similarly, the idea of leadership may appear anomalous to knowledge-workers (Jung 2001). This is partially because of the assumption that knowledge workers are somehow feline (e.g., Davies and Garrett 2010), and as cats cannot apparently be herded, leadership of knowledge workers is, therefore, not possible.[1] Because experts and professionals are more likely to be driven by intrinsic motivation (Amabile 1993), it has sometimes been assumed that management and leadership practices are less important. This is not correct. Organizations of very different kinds perform better when good management practices are in place (Bloom and Van Reenen 2007; Bloom et al. 2010a; McCormack et al. 2014). Similarly, there is evidence that knowledge-intensive organizations (hereafter KIOs) perform more effectively under the right kind of leadership (e.g., Goodall 2009a, b). Although management practices are important to performance outcomes, the evidence suggests that organizational leaders should be more than managers; they should be individuals who have a deep knowledge of the core-business[2] activity of their organizations, what we term ‘expert leaders’. For example, the core-business of a research university is scholarship. Universities were found to have improved in their performance if they were led, a number of years earlier, by presidents (vice chancellors, rectors) with strong research records (Goodall 2006, 2009a, b). The same finding exists at the academic department level. In a recent study assessing the success of US economics departments over time, we found that departments improved the most over a 15-year period if they were led by Chairs whose own research was highly cited (Goodall et al. 2014). A relationship between the core-business knowledge held by a leader and organizational success has been found in a number of settings: basketball (Goodall et al. 2011), hospitals (Goodall 2011), and Formula 1 Championships (Goodall and Pogrebna forthcoming). The suggestion from these studies is that leader characteristics should align with the organizations core-business activity.

The issue of expert leaders vs. leaders with a predominantly managerial background is germane because over the last few decades’, major firms have moved away from hiring CEOs with technical expertise, towards selecting leaders who are generalists (Frydman 2007; Bertrand 2009). An extreme example of a sector that has gone over to manager-CEOs and away from technical experts is that of healthcare. In the past, qualified doctors ran hospitals. In the United States (US) today only 4 % of hospitals are led by medically trained doctors (Gunderman and Kanter 2009). Similarly, in the United Kingdom (UK), most CEOs are now professional managers (Goodall 2011). The evidence and supporting arguments

[1] It is notable that despite popular belief cats are a social species and they adhere to hierarchies.

[2] The core business is defined here as the primary or underlying activity; it is that which is considered to be the most important or central endeavor in an organization, and it generates the most attention and income.

presented here suggest that the pendulum may have swung too far away from core-business functions towards management functions in the selection of leaders in many sectors.

In this chapter we will focus on leadership in KIOs. In Sect. 2 we summarize the evidence in support of expert leaders and draw from research in other fields. Section 3 elaborates on the concept of expert knowledge. In Sect. 4 we explore why organizations headed by experts appear to perform better. That is, what are the transmission channels through which expert leaders may generate better performance? We try to explain these processes through a theory of expert leadership (Goodall 2012; Goodall and Pogrebna forthcoming) presented here in a new conceptual framework. Finally, in Sect. 5 we discuss the potential weaknesses associated with expert leaders and Sect. 6 concludes.

2 Expert Leaders and Organizational Performance

To estimate leaders' effects on organizational performance in an exact way within real-world settings is known to be problematic (Antonakis et al. 2010; Blettner et al. 2012). Nevertheless, there is a growing research literature that claims to have captured leaders' influence on performance. Much of the work attempts to separate CEO effects from industry or firm effects to calculate the explanatory power of leaders and their characteristics (e.g., Waldman and Yammarino 1999; Mumford, Scott, Gaddis, and Strange 2002; Bertrand and Schoar 2003; Jones and Olken 2005; Bennedsen et al. 2007; Mackey 2008).

In the context of KIOs, recent studies have shown that leaders can have a positive effect on organizational performance (e.g., Goodall 2009a, b, 2014). In these studies the key leader characteristic that is observed, and is associated with a change in performance, is described as 'expert knowledge'. Expert knowledge has two components: industry experience and expert ability in the core business activity. Expert knowledge is not viewed as a proxy for management ability or leadership experience, both of which are always necessary as suggested above. However, the earlier studies outlined above do not measure these factors. A key suggestion is that expert knowledge about the core business should be viewed as a first-order requirement when hiring panels appoint organizational leaders (Goodall and Pogrebna forthcoming). Once this is established, other important factors can be scrutinized as a secondary process; for example, the more subjective attributes like style of leadership (transactional/transformational), personality (charisma/traits), or the nature of their relationships (leader–member exchange). These secondary factors are likely to be disparate. This expert leader proposition may be especially important in KIOs, where work tasks are less structured and workers often need greater autonomy to be creative (Amabile and Gryskiewicz 1989).

Employee creativity is known to be enhanced when supervisors themselves have creative expertise. Mumford et al. (2002) summarize this research, drawn predominantly from the psychology literature. They report that to lead creative individuals requires both "technical and creative problem-solving skills" since, as the authors suggest, "they provide a basis for structuring an inherently ill-defined task and because they provide the credibility needed to exercise influence" (2002, p. 712). The earliest of these studies, Andrews and Farris (1967), examined scientists' productivity. The ability of the supervisor as a scientist was the best predictor of a researcher's creative performance, when compared with other factors such motivation, maintaining group relationships, and the amount of autonomy granted to staff. These results were replicated in a similar study by Barnowe (1975).

A positive association between leaders' expert knowledge and firm performance has also been found in other KIOs. For example, in a study of US hospital CEOs, Goodall (2011) found that hospitals led by medically trained doctors, as opposed to professional managers, were more likely to be ranked higher in performance. In the sports setting of US basketball, coaches who were former All-Stars players with long playing careers were associated with the greatest winning success (Goodall et al. 2011). In the competitive industry of Formula 1 Championships, the performance of constructor teams (e.g., Ferrari, Mercedes, McLaren) were examined across the whole history of the industry. Team principals who were former racing drivers made the best leaders (Goodall and Pogrebna forthcoming). In these studies between 8 and 15 % of organizational performance is explained by the expert knowledge effect.

The research on expert leaders links to the literature on CEO origin, that attempts to identify a link between firm performance, among other outcomes, and whether a CEO has been hired from outside a firm or sector, or promoted from within (see Kesner and Dalton 1994; Shen and Cannella 2002; Wiersema 1995; Zajac 1990; Karaevli 2007; Zhang and Rajagopalan 2004, 2010) and therefore possesses expert knowledge. The evidence, that is ambiguous, in general reveals that insider and outsider CEOs bring different perspectives that may prove beneficial under different conditions; these are dependent upon, for example, pre or post-succession, firm performance during periods of environmental munificence or turbulence, the level of strategic change that is introduced, and so on (Harris and Helfat 1997; Karaevli 2007; Zhang and Rajagopalan 2010).

What can be observed is that outsider CEO hires have risen since the early 1990s (Lucier et al. 2003). A study on CEO succession in the world's largest 2,500 public companies revealed that in 2011 22 % of new CEOs came from outside their organization, compared to 14 % in 2007 (Booz & Co 2011). This study also showed that insider CEOs tend to stay for longer and leave their companies with higher shareholder returns, supporting our proposition that expert knowledge contributes positively to leadership performance.

3 Expert Knowledge in the Core-Business

What is meant by expert knowledge? There are over 4,000 academic articles that mention the term ‘expert knowledge’ in the disciplines of computer science, ecology and environmental science, engineering, medicine, psychology and sociology. In the social sciences the term tends to be linked to experts; those who have acquired a focused knowledge and expertise such that it affects how they perceive the world (see Ericsson et al. 2006). Experts differ from non-experts in a number of ways: knowledge is represented and bundled differently, they tend to think more holistically about problems (Bradley et al. 2006), and experts are more likely than novices to use abstract concepts to solve problems (Sembugamoorthy and Chandresekaren 1986).

Expert knowledge is acquired through a combination of technical education, domain-specific knowledge, practice and experience (Chase and Simon 1973; de Groot 1978); it combines explicit and tacit knowledge (Nonaka and Takeuchi 1995) and it might also be thought of as a deep understanding that aids intuitive decision-making, akin to wisdom (Tichy and Bennis 2007). We suggest that when a leader has expert knowledge in the core-business it influences decision-making through a process of ‘expertise-based intuition’ (Salas et al. 2010), an idea that combines the work on intuition and decision-making (e.g., Tversky and Kahneman 1981; Burke and Miller 1999; Lowenstein 2000; Gigerenzer 2007) with the literature on expertise (Ericsson et al. 2006). Salas et al. (2010) argue that it is intuition informed by expertise that leads to effective intuitive decision making. Performance is attained through mechanisms of expert decision-making derived from domain-knowledge, experience and practice (Ericsson et al. 1993; Salas et al. 2010).

As suggested above, our interpretation of expert knowledge is that it is tied to the core business of the organization through two mechanisms: the first is expert ability in the core business activity. This can be explained by an example from earlier studies of university leadership and organizational performance (Goodall 2006, 2009a, b). In some UK universities it has been possible for non-research focused academics to go into senior leadership positions (e.g., vice chancellors); these heads may include academics who dropped out of research completely at an early stage in their career, or those who maintain minimal research output. The evidence shows that university presidents who follow this kind of career trajectory are associated with reduced organizational performance, compared with presidents who were instead among the best researchers (Goodall 2006, 2009a, b). Thus, better researchers go on to make better university presidents.

The second factor we suggest that contributes to expert knowledge is industry experience, which accounts for the amount of time a head or supervisor has worked in an industry. This idea links directly with the research on CEO origin, outlined above, and to a new study showing that supervisor’s competence is the single strongest predictor of workers’ well-being (Artz et al. 2014). The authors examine data from three different sources. In a cross-section of 1,600 British workers, satisfaction levels are shown to be higher among individuals whose supervisor could if necessary step in competently to do their job; and in pooled cross-sections totaling 27,000 individuals, workers' job satisfaction is found to be highly correlated with the perceived competence of supervisors. Finally, in a cross-section of 6,000 young U.S. workers, the job satisfaction of employees is found to be positively associated with whether the supervisor worked his or her way up within the company, or started the company (Artz et al. 2014). Supervisor competence and industry experience are shown to be associated with workers' job satisfaction; this is important because happier workers also make more productive workers (Edmans 2012; Oswald et al. forthcoming).

Arguably, industry experience is valuable, but not in isolation. For example, it might be claimed that managers in hospitals have a great deal of knowledge about healthcare administration, finance and health policy because they have worked in the sector for many years. However, professional managers do not have expert knowledge in the core-business of hospitals which is the practice of medicine; only clinically-trained medics have this.

4 How do Expert Leaders Positively Influence Organizations?

Mumford et al. (2002) review the psychology literature to partially answer the question: why does a leader's technical expertise matter to the performance of their creative subordinates? They report, first, that the evaluation of creative people and their ideas can only be done by individuals who share their competencies; in short, it takes one to know one (or competently assess one). Second, leaders who share the same creative and technical perspective and motivation as their followers can communicate more clearly; finally, in relation to performance, they can better articulate the needs and goals of the organization. We build on these ideas to try and address the question: how do expert leaders influence organizational performance? In our conceptual framework in Fig. 1, we suggest there are two main channels through which expert leaders, in contrast to manager leaders, have a positive impact on performance. First, experts diverge in their *decisions and actions* from professional managers; and second, their expertise serves as a *signal* to insiders and outsiders—always holding constant management and leadership experience.

Next we summarize our model of expert leadership.

4.1 Expert Leaders

Expert knowledge, we argue, should be viewed as a first-order requirement, after controlling for management skills and leadership experience which are factors

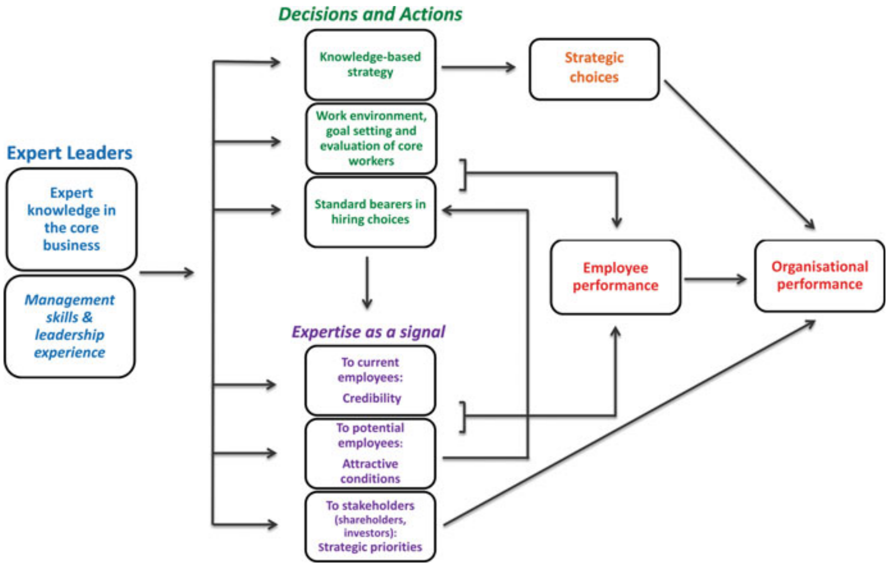


Fig. 1 Model of expert leadership: conceptual framework

required by all senior executives. Expert knowledge of the core-business is not a substitute for management skills and leadership experience. All leaders must be competent in these areas. Consequently, not all experts will make good managers and leaders.

Proposition 1 Expert knowledge is not a proxy for management skills and leadership experience, both are necessary prerequisites.

4.2 Decisions and Actions

Leaders are involved in multiple decisions and actions. To explain performance differences between expert leaders and manager leaders, we focus on those decisions and actions where the two types of leaders might reasonably differ. These are: (1) making knowledge-based strategic decisions, (2) creating a suitable work environment and providing adequate goals, and informed evaluation to core workers, and (3) hiring behavior.

4.2.1 Knowledge-Based Strategy

A knowledge-based strategy is that which combines a leader's expert knowledge of the core business with the strategic direction of their organization. We propose this happens in two ways: first, experts make better strategic choices compared with manager leaders because their preferences align with what is best for the organization; second, expert leaders are intrinsically motivated by the core-business activity, which we believe makes it more likely that experts will adopt the long view.

To become a successful expert, whether in architecture or engineering, an individual will normally focus intensely on their subject and sector for a number of years, thereby amassing a deep knowledge base. This is often referred to in the literature as domain knowledge; that which has been acquired through education, training, and experience within a particular context. Aligning a leader's own career preferences and priorities with the requirements of the core business will, we argue, shape decision-making and organizational strategy. An example can be found in Barker and Mueller (2002), who show that research and development (R&D) spending is significantly increased in firms where CEOs have advanced science-related degrees (Barker and Mueller 2002; see also Narayanan 2001).

This proposition connects with Hambrick and Mason's (1984) Upper Echelons (UE) theory (see Carpenter et al. 2004 for a review). UE theory argues that top managers make strategic choices that are reflections of their own values and cognitions. Members of the top management team will be influenced in their decision-making by individual and group demographic factors (such as age, education, functional track, top management team heterogeneity). UE theory focuses on the top management team (TMT) demographics, whereas we concentrate on CEO characteristics or generally any leaders characteristics.

Because of the extensive time and effort necessary to develop expertise to a high level, experts are more often self-motivated—driven by intrinsically motivated curiosity—rather than by purely extrinsic factors, such as money (Amabile 1993, 1996). Intrinsic motivation is defined as 'the drive to do something for the sheer enjoyment, interest, and personal challenge of the task itself (rather than solely for some external goal)' (Hennessey and Amabile 2010, p. 581). We suggest that leaders who are intrinsically motivated by the core-business may be more likely to adopt strategic choices that follow a long view with regards income and profit generation. In contrast, it is the adoption of a short view—or short-termism (Laverty 1996; Palley 1997; Marginson and McAulay 2008)—that is often linked by scholars (Nesbitt 2009; Dallas 2011) and commentators (e.g., Matthew Bishop in The Economist, Nov 13, 2009) to the financial crisis of 2008. CEO short-termism, demonstrated, for example, in the length of CEO decision horizon, is linked to a preference for investments with faster paybacks, which may be to the detriment of long-term value creation (Antia et al. 2010). Instead, we suggest that a long term strategy as adopted by expert leaders may increase organizational performance.

Proposition 2 Expert leaders implement more profitable organizational strategies than manager leaders.

4.2.2 Work Environment, Goal Setting and Evaluation of Core Workers

Expert leaders might be described as being the first among equals because they originated from the core workers. Having been 'one of them', expert leaders understand the culture and value system of core workers, and also their incentives and motivations. We argue that expert leaders are, therefore, more likely to create the right conditions for core workers, compared with leaders who are non-experts (e.g., those who ceased working in the core-business activity early in their careers, or are professional managers). As work environments are known to be important to employees' creativity and to their creative performance (Shalley 1991, 1995; Oldham and Cummings 1996), we argue that expert leaders will increase the performance of KIOs by creating the right work environment for the core workers. When the work environment complements the creative requirements of the job, individuals report higher job satisfaction and lower intentions to quit (Shalley et al. 2000), increasing organizational performance. The best core workers are expensive because they are in demand. Thus, if key employees are to be held onto, KIOs must offer competitive incentives and a fertile work environment.

How can leaders create the right work environment for knowledge workers? Suggestions from the literature on the role of supervisors in promoting creativity include: support and encouragement, effective communication, appraisal, and mentoring, and ensuring that appropriate human resource practices are in place (Drazin et al. 1999; Mumford 2000; Mumford et al. 2002; Oldham and Cummings 1996; Shalley et al. 2000; Tierney et al. 1999; Tierney 2008). Following Mumford et al. (2002) but applying their reasoning not only to creative industries, we argue that expert leaders are better placed to evaluate the ideas of their core workers and offer constructive feedback, because they have the same technical expertise as those being appraised (Mumford et al. 2000; Basadur et al. 2000). An important aspect with respect to the specific context of KIOs, is that these kinds of organizations require risk-taking insofar as new ideas often stem from unknown phenomena. Failure must therefore be tolerated (Watkins and Marsick 1993); expert leaders who originated from the core workers may be more likely to accept and tolerate failure and to live with the ambiguity of the context (Alvesson 1993). A key reason why managerial processes have been so widely introduced in various settings may be because managers do not understand how to assess, monitor or feedback to knowledge workers. If managers do not share expert knowledge with core workers then arguably trust will also be absent. A lack of trust may lead to the introduction of overly cumbersome management systems, and inappropriate assessment may create a counterproductive climate, leaving employees feeling unfairly treated and demotivated.

Tying in with evaluation and feedback is the setting of goals, as a prerequisite for appraising performance. To increase motivation, goals should be set in such a way that they are neither too low, nor too high (Locke et al. 1981). However, it may be hard for non-experts to establish the right balance, because they do not understand the complexity of key tasks and projects. Whereas experts, who understand the competencies and abilities of their employees, are more likely to set goals that are both attainable and challenging (Mumford et al. 2002). Indeed, if a firm wants to be among the best in its field, then, we argue, the board should hire a leader who is already one of the best in that same field. As it is the leader's responsibility to establish the quality threshold in an organization, if an outstanding expert is hired as leader, the bar is automatically raised. Also, importantly, it may be easier for a leader to be an effective quality enforcer if she or he has first met the standard that is to be imposed (i.e. the standard bearer should first bear the standard).

Evaluating the performance of core worker, setting appropriate goals and giving constructive feedback are important factors for a productive work environment. Autonomy is also important for knowledge workers (Bucher and Stelling 1969; Robertson and Hammersley 2000; Robertson and Swan 2003). Work environments that are found to be managed by supervisors who are supportive and not overly controlling foster creativity (Oldham and Cummings 1996). Expert leaders understand the conditions that are required because, as suggested above, they have direct knowledge of the field and understand the culture and value system of core workers, their incentives and motivations. Thus, they will likely trust their employees with greater autonomy. In contrast, to compensate for their lack of core-business knowledge, a non-expert or a professional manager may be more likely to use managerial processes that they (as managers) have learned through training, and also from their own experience of being supervised by other managers.

Proposition 3 Expert leaders create a more appropriate work environment for core workers than manager leaders.

4.2.3 Standard Bearers in Hiring Choices

Most CEOs and HR directors would likely agree that hiring the best people is central to the success of any organization. Individuals who have excelled in their field of expertise (top scientists, surgeons, etc.) might be expected to hire others who are also outstanding in their field. If higher quality core workers are employed, this is likely to lead to improved organizational performance. Homophily in hiring and promotion happens when recruiters seek to 'reproduce themselves in their own image' (Kanter 1977). That people select others who are like themselves is a form of assortative matching (Becker 1973). We suggest that an outstanding expert may be more likely to recognize other similar talent, and be willing to hire someone who is more able than they are. For example, in a school setting, the undergraduate backgrounds of principals' (school heads) are found to correlate with the academic undergraduate backgrounds of the teachers a head hires. Principals who attended more selective universities are more likely to hire teachers who have stronger academic backgrounds, which is shown to produce better student outcomes (Baker and Cooper 2005).

It has been suggested that people find it hard to hire others who are better than themselves.[3] Sometimes, negative self-feelings can be traced directly to, and are antecedents of, processes of social comparison (Festinger 1954). Job satisfaction and happiness have been shown to be related to how the self compares with similar others (Stiles and Kaplan 2004; Clark and Oswald 1996; Luttmer 2005).

Proposition 4 Expert leaders hire better employees than manager leaders.

4.3 Expertise as a Signal

We have argued that expert leaders make different decisions and take different actions compared with manager leaders. But they may also signal different messages about themselves and their organizations to their own workers and to outsiders.

4.3.1 Signals Credibility

Expert leaders may appear credible and command more respect because of their proven track record in the core-business activity. The idea that credibility legitimizes leaders' authority is well documented in the literature (e.g., Bass 1985; Bennis and Nanus 1985; Kouzes and Posner 2003). This approach focuses on the social interactions between leaders and their followers. We suggest that expert leaders are viewed as credible because they have 'walked-the-walk' to a high standard. It also signals that a head understands the culture and value system, incentives and priorities of those being led. A manager leader might have equal levels of executive power, but expert leaders are likely to have both power and influence particularly among the core workers. Although credibility can be acquired because of expert knowledge, arguably, in the long run, it must be maintained through good performance. This is why in the conceptual framework in Fig. 1 we show that the decisions and actions of a leader flow towards credibility. As workers are able to observe the decisions and actions of their leader as well as the consequences of those decisions and actions, the importance of expertise for creating credibility may be reduced over time.

[3] This is captured in a statement attributed to André Weil, a French mathematician from the mid-twentieth century, in his 'Weil's Law of university hiring': 'First rate people hire other first rate people. Second rate people hire third rate people. Third rate people hire fifth rate people.' Likewise, the American writer Leo Rosten is cited as having said, 'First-rate people hire first-rate people; second-rate people hire third-rate people.' In interviews with university presidents (Goodall 2009a), a number of heads commented on the need to put the most outstanding scholars on hiring panels to ensure that the best academics are hired.

Proposition 5 Because expert leaders are more credible than manager leaders, they are more willingly followed by core workers.

4.3.2 Signals Work Conditions

The credibility of a CEO or president may send an important signal to the firm's current employees; it also signals information to *potential* employees who may be at an informational disadvantage with respect to organizational characteristics like, for example, the work environment. If the firm is headed by an expert leader this might suggest to potential core employees that an appropriate work environment exists (for reasons explained earlier). Signaling that an optimal work environment exists in an organization may expand the applicant pool of potential outstanding core workers and thereby increase performance.

Proposition 6 Expert leaders attract better potential employees than manager leaders.

4.3.3 Signals Strategic Priorities

Finally, hiring an outstanding expert leader may also signal credibility to a wider audience. For example, an organization's board may choose to hire a noted expert or specialist to send out a signal about strategic priorities to employees, and also to external stakeholders such as shareholders, customers, suppliers, the media, and donors.

Proposition 7 Expert leaders appear in a more positive light for external stakeholders than manager leaders.

5 Discussion: The Potential Vulnerabilities of Expert Leaders

We have so far presented a positive view of expert leaders and explained how organizational performance may be enhanced (see Fig. 1). However, it is also important to consider the potential shortcomings associated with the individual characteristics of specialists. We have identified four possible drawbacks associated with expert leaders in knowledge intensive settings which are outlined below.[4]

[4] These are the most common critiques raised by those attending presentations of this work and also in media interviews.

(1) Experts may have an overly narrow perspective. Being intrinsically motivated, to the extent that experts are, will likely require shutting out the world for some years lest it distracts from the goal in hand. This intense focus may have a detrimental effect on other areas of personal development. It might also stifle original thinking or lead to 'groupthink' (Janis 1971). In the case of a hospital, a critic might suggest that an expert surgeon may be less able to make judgments about other fields, like acute care or preventative medicine, and, therefore, a generalist or lay person may be better placed to weigh up competing arguments. This, we believe is incorrect. Arguably, a general manager may also have a contracted perspective if, for example, he or she has come from a finance background or marketing. Another often heard suggestion is that managers and administrators make it possible for core workers to have more time to focus on their own priorities, for example, patient care. This is an ideal scenario; however, the opposite situation may happen if administrators design managerial systems without consideration about how these processes will affect core workers. Universities and hospitals are places where these types of complaints are common place. Finally, a discrepancy may also arise in performance management and assessment, especially in the provision of complex services, such as healthcare or education. The priorities of an expert will likely differ from those of a manager.

(2) There is a risk that professional rivalries may occur between an expert leader and other experts. This is often perceived to be a potential problem in universities, because of fears that disciplinary favoritism might distort organizational priorities (e.g., the London School of Economics never hire's internal faculty members to become Director for this reason). We reject this criticism because it is unclear why prospective rivalries would be any more likely in KIOs than in other kinds. In universities among other settings, there has been a preference (bearing on the extreme) to hire external candidates into leadership roles. It is unclear how this pattern emerged, however it is likely to have been introduced by head-hunters who cannot charge clients for internal candidates. Hiring only externals may have its benefits, insofar as it brings in new blood and possibly new and better practices; but it is also likely to change the culture of an organization, and, importantly, it fails to recognize the potential of internal candidates and capitalize on their loyalty to the institution.

(3) Industries often decline in prominence because of outdated norms and operations, which may require an injection of new blood and new systems from outside (Spender 1989). In the UK public sector, and more generally around the world, a New Public Management (Hood 1991) led to a revolution in the delivery of public services. Professional managers may have assumed greater powers in organizations because of the recognition that modern management practices contribute towards successful performance (Bloom and Van Reenen 2007). Experts and professionals may have been less inclined to adopt innovative managerial applications. This could in part be explained by the generic style of management training offered. Experts and professionals tend to be reluctant leaders and managers. But if an appropriate form of training were

made available (e.g., bite-sized portions, the right incentives, use of common language with reduced jargon), take-up may be enhanced. We have promoted the idea that being a successful manager alone is not a sufficient condition for effective leadership; but, correspondingly, we have also argued that experts must also be excellent managers if they are to take on leadership positions.

(4) Finally, the success that experts accrue in one field, may imbue them with the belief that success is as likely in all fields they later engage with, thus inducing hubris. There is a tendency for successful people (both experts and non-experts) to attribute their ascent to their own brilliant talent—as predicted by social psychology's attribution theory (Heider 1958). For example, a successful surgeon praised for his or her technical skills, more adored because of saving lives, might over-attribute to their own talent. Success, arguably, comes also from luck, networking and providence. It has been suggested that leaders need empathy if they are to be effective (Maccoby 2000; Kelletta et al. 2006). Intrinsic motivation combined with self-motivation may weaken the specialist's ability to place themselves in others' shoes.

Conclusion

Those hiring senior leaders in knowledge intensive settings, from hospitals to technology firms, may be inclined to appoint CEOs who are talented managers but have little or no expertise in the core business of the organization. The arguments laid out in this chapter suggest that this is a mistake. KIOs should, according to the evidence, look for leaders who have expert knowledge (those who understand the core-business, and are competent in the organization's core activity). In this chapter we present a model of expert leadership that builds on previous conceptual literatures and evidence. The framework in Fig. 1 suggests that organizations will perform more effectively when led by expert leaders, as compared to capable general managers.

Given our promotion of expert leadership, this paper is implicitly critical of the empirically-documented rise of the professional manager and generalist CEO. Whilst acknowledging possible vulnerabilities, we argue that expert leaders can be expected to improve organizational performance through both their actions and decisions, and also the signals they convey. Specifically, we propose several mechanisms through which expert leaders might improve firm performance: by implementing a knowledge-based strategy; by creating an attractive environment for core workers, and appropriate goal setting and evaluation; by hiring the best workers; and finally, by signaling credibility to current and potential employees and other important stakeholders.

We argue that these mechanisms might be especially relevant in knowledge-intensive settings, where: (1) The relevance of organizational strategy for staying competitive might be especially strong (*knowledge-based strategy*); (2) Workers effort cannot be easily measured and therefore

(continued)

requires appropriate or technical appraisal (*evaluation*); (3) Failure needs to be tolerated by managers in experimental settings such as KIOs (*work environment*); (4) Knowledge-workers can be among the most expensive to retain, therefore requiring the right incentive structures and work environment (*work environment*); and (5) Performance is more closely tied to employees' ability (*hiring)*.

The theoretical framework in Fig. 1 raises propositions in the form of transfer mechanisms to be tested in new research.[5] Furthermore, the theory of expert leadership outlined here draws from studies of organizations that are either knowledge-intensive (e.g., research groups, academic departments, universities, hospitals) or high-skill (e.g., basketball teams and Formula 1 Championships). These settings have been examined partly because measuring productivity is relatively uncomplicated; for example, it is possible to apportion research outputs and citations to individuals, groups and institutions. Similarly, hospitals are publicly ranked according to quality measures, and so on. However, the question of how the expert leader proposition will fare in other kinds of settings—the boundary conditions of our theory—will need to be established empirically in future research.

References

Alvesson M (1993) Organizations as rhetoric: knowledge-intensive firms and the struggle with ambiguity. J Manag Stud 30(6):997–1015

Amabile TM (1993) Motivational synergy: toward new conceptualizations of intrinsic and extrinsic motivation in the workplace. Hum Resour Manag 3(3):185–201

Amabile TM (1996) Attributions of creativity: what are the consequences? Creativity Res J 8 (4):423–426

Amabile TM, Gryskiewicz ND (1989) The creative environment scales: work environment inventory. Creativity Res J 2:231–253

Andrews FM, Farris GF (1967) Supervisory practices and innovation in scientific teams. Pers Psychol 20:497–515

Antia M, Pantzalis C, Park JC (2010) CEO decision horizon and firm performance: an empirical investigation. J Corp Financ 16(3):288–301

Antonakis J, Bendahan S, Jacquart P, Lalive R (2010) On making causal claims: a review and recommendations. Leadership Q 21(6):1086–1120

Artz B, Goodall AH, Oswald AJ (2014) Boss competence and worker well-being, working paper

Baker BD, Cooper BS (2005) Do principals with stronger academic backgrounds hire better teachers? Policy implications for high poverty schools. Educ Admin Q 41(3):449–479

Barker VL, Mueller GC (2002) CEO characteristics and firm R&D spending. Manag Sci 48 (6):782–801

Barnowe JT (1975) Leadership and performance outcomes in research organizations. Organ Behav Hum Perform 14:264–280

[5] Bäker and Goodall began a new study of leadership and scientific productivity in March 2014.

Basadur M, Runco MA, Vega LA (2000) Understanding how creative thinking skills, attitudes, and behaviors work together: a causal process model. J Creative Behav 34:77–100
Bass BM (1985) Leadership and performance beyond expectation. Free Press, New York
Becker GS (1973) A theory of marriage: part I. J Polit Econ 81(4):813–846
Bennedsen M, Nielsen KM, Pérez-González F, Wolfenzon D (2007) Inside the family firm: the role of families in succession decisions and performance. Q J Econ 122(2):647–691
Bennis WG, Nanus B (1985) Leaders. Harper & Row, New York
Bertrand M (2009) CEOs. Annu Rev Econ 1:1.1–1.29
Bertrand M, Schoar A (2003) Managing with style: the effect of managers on firm policies. Q J Econ 118(4):1169–1208
Bishop M (2009) Now for the long term: it's time for businesses to think about the future again. The Economist Nov 13th
Blettner DP, Chaddad FR, Bettis RA (2012) The CEO performance effect: statistical issues and a complex fit perspective. Strateg Manag J 33(8):986–999
Bloom N, Van Reenen J (2007) Measuring and explaining management practices across firms and countries. Q J Econ 122(4):1351–1408
Bloom N, Genakos C, Martin R, Sadun R (2010a) Modern management: good for the environment or just hot air? Econ J 120(544):551–572
Bradley JH, Paul R, Seeman E (2006) Analyzing the structure of expert knowledge. Inf Manag 43 (1):77–91
Bucher R, Stelling J (1969) Characteristics of professional organizations. J Health Soc Behav 10:3–15
Burke LA, Miller MK (1999) Taking the mystery out of intuitive decision making. Acad Manag Exec 13:91–99
Carpenter MA, Geletkanycz MA, Sanders WG (2004) Upper echelons research revisited: antecedents, elements, and consequences of top management team composition. J Manag 30 (6):749–778
Chase WG, Simon HA (1973) Perception in chess. Cogn Psychol 4:55–81
Clark AE, Oswald AJ (1996) Satisfaction and comparison income. J Public Econ 61:359–381
Booz & Company (2011) CEO succession report: 12th annual global CEO succession study. http://www.booz.com/media/uploads/BoozCo_CEO-Succession-Study-2011_Extended-Study-Report.pdf. Accessed 11 Apr 2014
Dallas L (2011) Short-termism, the financial crisis and corporate governance. San Diego Legal Studies Paper No 11-052. Available at SSRN: http://ssrn.com/abstract=1794190
Davies GJ, Garrett G (2010) Herding cats: being advice to aspiring academic and research leaders. Triarchy Press, Devon
de Groot A (1978) Thought and choice in chess, 2nd edn. Mouton De Gruyter, The Hague
Drazin R, Glynn MA, Kazanjian RK (1999) Multilevel theorizing about creativity in organizations: a sense making perspective. Acad Manag Rev 24:286–329
Edmans A (2012) The link between employee satisfaction and firm value, with implications for corporate social responsibility. Acad Manag Perspect 6(4):1–19
Ericsson KA, Krampe RT, Tesch-Romer C (1993) The role of deliberate practice in the acquisition of expert performance. Psychol Rev 100:363–406
Ericsson KA, Charness N, Hoffman RR, Feltovich PJ (eds) (2006) The Cambridge handbook of expertise and expert performance. Cambridge University Press, Cambridge
Festinger L (1954) A theory of social comparison processes. Hum Relat 7:117–140
Frydman C (2007) Rising through the ranks: the evolution of the market for corporate executives, 1936-2003. J Econ Hist 66(2):516–517
Gigerenzer G (2007) Gut feelings: the intelligence of the unconscious. Penguin, New York
Goodall AH (2006) Should top universities be led by top researchers, and are they? A citations analysis. J Doc 62:388–411
Goodall AH (2009a) Highly cited leaders and the performance of research universities. Res Policy 38(7):1079–1092

Goodall AH (2009b) Socrates in the boardroom: why research universities should be led by top scholars. Princeton University Press, Princeton; Oxford

Goodall AH (2011) Physician-leaders and hospital performance: is there an association? Soc Sci Med 73(4):535–539

Goodall AH (2012) A theory of expert leadership. IZA Working paper, IZA Institute for the Study of Labor, Bonn, Germany

Goodall, AH (2014) Leadership: experts versus generalists. Forthcoming in: Shaw K (section ed.) IZA World of labor: evidenced based policy making. Bloomsbury Press, New York

Goodall AH, Pogrebna G (Forthcoming) Expert leaders in a fast moving environment. IZA working paper no. 6715, IZA Institute for the Study of Labor, Bonn, Germany

Goodall AH, Kahn LM, Oswald AJ (2011) Why do leaders matter? A study of expert knowledge in a superstar setting. J Econ Behav Organ 77:265–284

Goodall AH, McDowell JM, Singell LD (2014) Leadership and the research productivity of university departments. IZA Working paper, IZA Institute for the Study of Labor, Bonn, Germany (Under review)

Gunderman R, Kanter SL (2009) Educating physicians to lead hospitals. Acad Med 84:1348–1351

Hambrick DC, Mason P (1984) Upper echelons: the organization as a reflection of its top managers. Acad Manag Rev 9:193–206

Harris D, Helfat C (1997) Specificity of CEO human capital and compensation. Strateg Manag J 18 (11):895–920

Heider F (1958) The psychology of interpersonal relations. Wiley, New York

Hennessey BA, Amabile TM (2010) Creativity. Annu Rev Psychol 61:569–598

Hood C (1991) A public management for all seasons. Public Admin 69(Spring):3–19

Janis IL (1971) Groupthink. Psychol Today 5(6):43–46, 74–76

Jones BF, Olken BA (2005) Do leaders matter? National leadership and growth since World War II. Q J Econ 120(3):835–864

Jung DI (2001) Transformational and transactional leadership and their effects on creativity in groups. Creativity Res J 13:185–197

Kanter R (1977) Men and women of the corporation. Basic Books, New York

Karaevli A (2007) Performance consequences of new CEO "outsiderness": moderating effects of pre- and post-succession contexts. Strateg Manag J 28:681–706

Kelletta JB, Humphrey RH, Sleeth RG (2006) Empathy and the emergence of task and relations leaders. Leadership Q 17(2):146–162

Kesner IF, Dalton DR (1994) Top management turnover and CEO succession: an investigation of the effects of turnover on performance. J Manag Stud 31:701–713

Kouzes JM, Posner BZ (2003) Credibility: how leaders gain and lose it, why people demand it. Jossey-Bass, San Francisco

Laverty KJ (1996) Economic "short-termism": the debate, the unresolved issues, and the implications for management practice and research. Acad Manag Rev 21(3):825–860

Locke EA, Shaw KN, Saari LM, Latham GP (1981) Goal setting and task performance: 1969–1980. Psychol Bull 90(1):125–152

Lowenstein R (2000) When genius failed: the rise and fall of long-term capital management. Random House, New York

Lucier C, Schuyt R, Handa J (2003) CEO Succession 2003: the 'perils' of good governance. Booz Allen Hamilton, McLean, VA

Luttmer EFP (2005) Neighbors as negatives: relative earnings and well-being. Q J Econ 120 (3):963–1002

Maccoby M (2000) Narcissistic leaders. Harv Bus Rev 78(1):68–72

Mackey A (2008) The effect of CEOs on firm performance. Strateg Manag J 29(12):1357–1367

Marginson D, McAulay L (2008) Exploring the debate on short-termism: a theoretical and empirical analysis. Strateg Manag J 29(3):273–292

McCormack J, Propper C, Smith S (2014) Herding cats? Management and university performance. Econ J 124(578): F534–F564

Mumford MD (2000) Managing creative people: strategy and tactics for innovation. Hum Resour Manag R 10:313–351
Mumford MD, Marks MA, Connelly MS, Zaccaro SJ, Reiter-Palmon R (2000) Development of leadership skills: experience, timing, and growth. Leadership Q 11:87–114
Mumford MD, Scott GM, Gaddis B, Strange JM (2002) Leading creative people: orchestrating expertise and relationships. Leadership Q 13(6):705–750
Narayanan VK (2001) Managing technology and innovation for competitive advantage. Prentice Hall, Englewood Cliffs, NJ
Nesbitt J (2009) The role of short-termism in financial market crises. Aust Account Rev 19 (4):314–318
Nonaka I, Takeuchi H (1995) The knowledge-creating company. Oxford University Press, New York
Oldham GR, Cummings A (1996) Employee creativity: personal and contextual factors at work. Acad Manag J 39:607–634
Oswald AJ, Proto E, Sgroi D (forthcoming) Happiness and productivity. J Labor Econ (forthcoming)
Palley TI (1997) Managerial turnover and the theory of short-termism. J Econ Behav Organ 32 (4):547–557
Robertson M, Hammersley GOM (2000) Knowledge management practices within a knowledge-intensive firm: the significance of the people management dimension. J Eur Ind Train 24(2/3/4):241–253
Robertson M, Swan J (2003) 'Control–what control?'Culture and ambiguity within a knowledge intensive firm. J Manag Stud 40(4):831–858
Salas E, Rosen MA, DiazGranados D (2010) Expertise-based intuition and decision making in organizations. J Manag 36:941–973
Sembugamoorthy V, Chandresekaren B (1986) Functional representation of devices and compilation of diagnostic problem-solving systems. In: Kolodner JL, Reisbeck CK (eds) Experience, memory, and reasoning. Erlbaum, Hillsdale, pp 47–71
Shalley CE (1991) Effects of productivity goals, creativity goals, and personal discretion on individual creativity. J Appl Psychol 76:179–185
Shalley CE (1995) Effects of coaction, expected evaluation, and goal setting on creativity and productivity. Acad Manag J 38:483–503
Shalley CE, Gilson LL, Blum TC (2000) Matching creativity requirements and the work environment: effects on satisfaction and intensions to leave. Acad Manag J 43(2):215–223
Shen W, Cannella AA (2002) Revisiting the performance consequences of CEO succession: the impact of successor type, post-succession senior executive turnover, and departing CEO tenure. Acad Manag J 45:717–733
Spender JC (1989) Industry recipes: an enquiry into the nature and sources of managerial judgement. Basil Blackwell, London
Stiles BL, Kaplan HB (2004) Adverse social comparison processes and negative self-feelings: a test of alternative models. Soc Behav Pers 32(1):31–44
Tichy NM, Bennis WG (2007) Making judgement calls. Harv Bus Rev (October):94–102
Tierney P (2008) Leadership and employee creativity. In: Zhou J, Shalley CE (eds) Handbook of organizational creativity. Lawrence Erlbaum, New York, pp 95–124
Tierney P, Farmer SM, Graen GB (1999) An examination of leadership and employee creativity: the relevance of traits and relationships. Pers Psychol 52:591–620
Tversky A, Kahneman D (1981) Judgment under uncertainty: heuristics and biases. Science 185:1124–1131
Waldman DA, Yammarino FJ (1999) CEO charismatic leadership: levels-of-management and levels-of-analysis effects. Acad Manag Rev 24(2):266–285
Watkins KE, Marsick VJ (1993) Sculpting the learning organization: lessons in the art and science of systemic change. Jossey-Bass, San Francisco
Wiersema MF (1995) Executive succession as an antecedent to corporate restructuring. Hum Resour Manag 34:185–202

Zajac EJ (1990) CEO selection, succession, compensation, and firm performance: a theoretical integration and empirical evidence. Strateg Manag J 11:217–230
Zhang Y, Rajagopalan N (2004) When the known devil is better than an unknown God: an empirical study of the antecedents and consequences of relay CEO successions. Acad Manag J 47:483–500
Zhang Y, Rajagopalan N (2010) Once an outsider, always an outsider? CEO origin, strategic change, and firm performance. Strateg Manag J 31(3):334–346

Incentives and Power: An Organizational Perspective

Otto Hüther and Georg Krücken

Abstract Over the last years new incentives for professors were introduced into the German university system in order to strengthen the external motivation and the productivity of professors. At the same time a critical reflection has begun, in which the effects on internal motivation and deficiencies concerning the measurement and the overall effects on the science and higher education system were discussed. In addition to these criticisms, we will argue from an organizational perspective. From that perspective, incentives are a central aspect of power in organizations. This allows us to put incentives in a broader perspective, in which also other forms of power in higher education organizations come into light. Such forms are the power over resources, careers, and, ultimately, membership. The article argues that due to the specificities of the German university system and its organizations, the nexus between power and incentives is rather weak as compared to other systems. However, such a structure is not per se problematic. It generates a specific set of advantages and disadvantages with regard to the missions of universities in a knowledge society and some critical side-effects of a strong nexus between power and incentives can thus be avoided.

1 Introduction

In the past decades, higher education in Europe has undergone significant reforms (Paradeise et al. 2009; de Boer et al. 2007b; Krücken et al. 2013). Many of the changes made pertain to the governance of higher education, the expansion of organizational management capacities, the courses offered and modes of financing. In addition to targeting organizational and structural aspects of higher education, the reforms are also laid out to affect the motivational structures of academic staff. In this regard, strategies involving new incentive structures linked to performance have been introduced. Attempts to set incentives in the form of remuneration and

O. Hüther • G. Krücken (✉)
International Centre for Higher Education Research Kassel (INCHER), University of Kassel, Kassel, Germany
e-mail: kruecken@incher.uni-kassel.de

I.M. Welpe et al. (eds.), *Incentives and Performance*,
DOI 10.1007/978-3-319-09785-5_5

resource allocation amongst professors are particularly pertinent. While traditionally in science studies, recognition from academic peers was considered the most relevant incentive for researchers as part of an intrinsic motivational framework (Hagstrom 1965; Luhmann 1992; Merton 1973), current reforms aim to introduce more extrinsic motivational structures. These reforms are therefore at least implicitly accompanied by the suspicion that academics have been deliberately underperforming and aim to mobilize productivity reserves in universities.[1]

The reforms are somewhat controversial. It has been postulated that an increased emphasis on external incentives is likely to undermine academics' intrinsic motivation. Such effects are widely discussed in the crowding out theory of motivation (e.g., Frey and Oberholzer-Gee 1997; Osterloh and Frey 2000). Other researchers point out that measuring academic performance is generally highly problematic with possibly unforeseeable effects on the academic system as a whole (e.g., Maasen and Weingart 2006; MacRoberts and MacRoberts 1996; Martin and Whitley 2010; Osterloh and Frey 2009; Espeland and Sauder 2007). The academic system could be affected if academics increasingly set out to tailor their projects around the new performance criteria (e.g., Osterloh 2012). According to critics, this could lead to a mainstream research culture (Münch 2006; Lee 2007) built around determining the smallest possible unit of content fit for publication (Butler 2003).

In the literature, the effects of monetary and non-monetary incentives are mainly considered in regard to effects at the level of the individual researcher, research groups and/or to the academic system as a whole (e.g., Stern 2004; Jansen et al. 2007; Heinze et al. 2009; Bruneau and Savage 2002; Partha and David 1994; Sutton 1984; Merton 1973; Hagstrom 1965). The effects at the internal organizational level of formal scientific organizations are hardly analyzed. Among other reasons, this is surprising because a great deal of monetary incentives and resource allocation takes place within the organization 'university'. Effects on the organization are therefore to be expected. If the analysis of incentives focuses on the internal organizational level, then the incentives are part of the power structure of the organization. In this case, the organization positively sanctions desired behavior. The capacity to enact positive sanctions comprises only one aspect of an organization's power structure; the potential to enact negative sanctions makes up the other side of the coin. The aim of this paper is to place the effects of newly enacted incentive structure reforms in relation to the power structures in German universities. Our organizational perspective is of particularly relevance because most of the recent reforms with regard to the scientific system in Germany are indeed organizational reforms of universities (Hüther 2010). Research on organizational change indicates that the success of such reforms strongly depends on the power structure within the organizations (e.g., Royston and Hinings 1996; Hannan and Freeman 1984; Cyert and March 1963). Therefore, the capacity of changing the German scientific system and the internal power structure of

[1] The discussion shares many parallels with that currently underway in regard to the public sector as a whole (Frey et al. 2013).

universities are heavily intertwined. Furthermore, our perspective relates the discussion of incentives in the social system of science to a central issue in organizational research: the issue of power in organizations.

Generally, it can be said that in comparison to universities in other countries, German universities traditionally exercised hardly any power over their operational core, the academic staff (Paradeise et al. 2009; Hüther and Krücken 2013). Consequently, academics could hardly be obligated to adhere to organizational goals. Before giving evidence of the increasing importance of organizational goals, one should stress that for a long time the traditional power structure was seen as an advantage of the German system as compared to other national system, in particular at the end of the nineteenth and the beginning of the twentieth century, when the success of German academia had been recognized world-wide (Ben-David 1971; Flexner 1930). One could argue that the high autonomy of professors and their major orientation towards the scientific community instead of their formal university setting better suits creative research and, ultimately, scientific progress as compared to planning via organizational goals (Partha and David 1994; Krücken 2008). However, for at least three reasons a stronger role for the university as an organization vis-à-vis its professorial members can be witnessed since the mid 1990s.

First, with the expansion of higher education the teaching function of universities is becoming of ever-increasing importance. Currently, a bit more than 50 % of the relevant age cohort study at German higher education institutions (BMBF 2014). In this, a global trend towards the massification of higher education is expressed (Trow 2010; Schofer and Meyer 2005). Organizational goals and structures are of paramount importance in the realm of teaching, while the research function of universities is based on research networks and scientific communities, and to a far lesser extend it is bound to the university as an organization.

Second, one can see that the goals of universities have been multiplied over time (Schimank 2001). While the traditional missions of universities consist in teaching and research, in addition to these two missions universities are currently expected to promote the direct transfer of knowledge and technology (Krücken 2003), to integrate women, people from lower social classes and migrants (Shils 1991) and to offer continuing education (Wilkesmann 2010).

Third, as Brunsson and Sahlin-Andersson (2000) claim, there has been a trend since the 1980s towards constructing public sector organizations as complete organizations with identity, hierarchy and rationality. The organizational form of the arena, in contrast, is losing ground. The traditional German university is prototypical for this organizational form, in which „members perform their tasks relatively free from control by the local leadership. Instead, they (. . .) are controlled by, external parties"(Brunsson and Sahlin-Andersson 2000, p. 734). Arenas can be regarded as highly functional, for example, in the field of knowledge production. Nevertheless, they are losing legitimacy in their broader socio-political environments. Therefore, the organizational form of what Brunsson and Sahlin-Andersson (2000) call 'complete organization' is gaining importance among universities (Krücken and Meier 2006; de Boer et al. 2007a), health care organizations (Preston

1992; Reay and Hinings 2005; Bode 2010) and public organizations as such (Pollitt and Bouckaert 2004).

The three reasons have led to question the adequacy of the German university model as one cannot simply assume that university professors are intrinsically motivated to pursue teaching in an era of mass education as well as additional organizational goals, which both are in open conflict with the research function of universities. While the functionality of a relatively weak organizational power structure within universities has proven its fruitfulness for research as individual researchers are internally motivated to carry out this task, it is doubtful whether such a power structure is adequate for the pursuit of mass education and additional organizational goals. In addition, the legitimacy of the arena model, which underlies the traditional German university, has been questioned in a variety of organizational sectors.

One possible reaction to the processes described above is to establish stronger linkages between university professors and the organization and, thereby, changing the power structures within universities. We assume that exactly this happened by the introduction of performance-related pay and restricted resource allocation. The rationale is based on the assumption that such new incentive structures enable the organization to create the motivation to further their multiple and in part new organizational goals. The reforms can also be interpreted as one aspect of the transformation of universities into ‘complete organizations’. In addition to the classical steering bodies of the higher education system—the state on the one hand and academic self-organization on the other—the organization and its leadership are now assuming additional steering functions (Krücken and Meier 2006). However, significant limitations are also clearly visible, which we will discuss in this paper.

To illustrate the changes to the power structures in Germany, the article focuses on three central formal power structures in organizations: the promotion or hindrance of careers within the organization (personnel power), the possibility of exclusion from the organization (organizational power) and the provision of resources to and remuneration of individuals (resource power).[2] In the following, these power structures will be analyzed not so much in terms of their coercive nature, but in terms of their potential to create motivation for desired behavior. It is, however, necessary to define what we mean by organizational power within the framework of this article before we concentrate on the various power structures.

[2] For more details on these three power structures in German higher education institutions see Hüther and Krücken (2011, 2013).

2 Power in Organizations

Power is one of the fundamental and central concepts in organizational research. There is, however, no scientific consensus on how power should be defined or which methods should be used to examine power phenomena in organizations (Diefenbach et al. 2009). Whereas classic texts on organizational research focus more strongly on formal power (Blau and Scott 1969; Mayntz 1968; Luhmann 1964; Etzioni 1964), within the last two decades the focus has shifted towards informal power, self-disciplining and legitimation of power (Clegg et al. 2006, pp. 290–319; Diefenbach and Sillince 2011; Kärreman and Alvesson 2004; Brown and Lewis 2011). There is no doubt that these recent developments in the analysis of power in organizations are highly relevant. Nonetheless, in the following we will use a more traditional understanding of power with a strong focus on formal power. We do this because our interest is the motivational effect of power in universities. Since there is hardly any research available in this area we use formal power structures as a starting point.

Our definition of power is as follows: in a social relationship, power exists when it is mutually assumed that one actor has control or influence over something the other actor desires. The base and the degree of power are therefore determined by the desires of the subordinate and the importance of the desire (Emerson 1962). As Scott noted: "The power of superordinates is based on their ability and willingness to sanction others—to provide or withhold rewards and penalties (. . .) what constitutes a reward or a penalty is ultimately determined by the goals and values of the subordinates in the relation" (Scott 1981, p. 276). In organizations such as universities we can find many power relations based on many desires. Nevertheless, important parts of the power relations in organizations are typically attached to formal positions and the ability to sanction subordinates. Usually a superordinate decides over who can remain within an organization and therefore has organizational power. A superordinate also decides over careers or can at least exert considerable influence through his or her assessments of the subordinates' performance. Furthermore, in most cases a superordinate decides over resource allocation and performance-related pay. In the following we will use the example of German universities to examine whether the university as an organization really has the sources of power to decide about who stays in the organization and may pursue a career, and the allocation of resources and performance-related pay.

2.1 Personnel Power

In numerous organizations the actions of members are among other things influenced by the fact that their superordinates can affect their careers within the organization. Members are therefore motivated by the prospect of having a career (Luhmann 1980). German universities cannot apply this motivational instrument that Luhmann (1980) called *Personalmacht* (personnel power) to most of the academics working within their structures.

First of all, it is important to consider the career tracks of German academics. It is striking that after a professorship[3] has been attained, no further career steps within the German academic system are intended. There is, however, a distinction between professorships (C2 to C4 or W2 to W3[4]). The C4 or W3 professorships come with a higher salary and more resources. But the organizations cannot exploit these differences for motivational purposes because moving from a lower into a higher professorship within one university is usually impossible. This is due to the German *Hausberufungsverbot*. The *Hausberufungsverbot* is a traditional part of the German academic system and means that members of the university issuing the call usually may not or should not be considered for the open position. On the one hand, this ban is backed by law, although exceptions are possible. On the other hand it is enforced by informal norms provided by the shared convictions of the professors who view it as a legitimate means to prevent patronage. The combination of formal and informal rules means that in practice to accept a call or to move into a higher graded professorship means to change university. In other words, the market for professorships in Germany is an external labor market (Musselin 2005).

This has consequences. Although the desire to have a career within the academic system is an important source of motivation for academics, decisions about careers are not made within their particular work organization. In the German case, this means that the organization is not able to motivate its academics to pursue those of its goals that do not coincide with those of the academic system. Career prospects can hardly be utilized to motivate professors to excel at teaching or fully engage in academic self-administration. The German *Hausberufungsverbot* leads to career prospects and ambitions being channeled into motivation for research and reputation building, which are the overarching criteria for a successful career in the academic system. In addition, research is far more visible and easier to evaluate for other universities that decide about careers, in contrast to teaching and participation in academic self-administration.

The situation for the vast majority of positions below the professorship is slightly different. Here, promotion within the organization from graduate student to a postdoctoral position is possible and common. An internal labor market, characterized by patronage, can be clearly seen. It starts with professors recruiting their graduate students from within their student body. Recruiting postdoctoral staff is also characterized by personal contacts, and staff associated with the professor's own chair are often preferred (Enders and Bornmann 2001; Enders 2008). It should be noted that personnel power does exist at the level of the professorship. By offering career prospects, professors can motivate their staff to commit to a wide range of

[3] For reasons of simplicity we will not consider the Junior Professors because they hardly figure in the German system. Only 4 % of professors are Junior Professors (Statistisches Bundesamt 2013).

[4] In contrast to other systems, this distinction is usually not directly visible to outsiders. Normally one can not find this information on the professors' websites or their business cards.

behaviors, including behaviors that further the academic career of the professor more than those of their staff. The explanation is that the dependency of the staff on the professor relates to their prospects for further qualification and therefore their chances of remaining within the academic system at all. The professor's support is mandatory for both a doctorate and a habilitation. In Germany, the doctoral and habilitation phase traditionally has little structure (Röbken 2007) and the still prevalent "master-apprentice model" (Bosbach 2009) not only influences age, the type of support etc. but also the power structures and sources of motivation within the university. For the staff, there is personal loyalty towards the professor and not the organization. Put bluntly, the professor has something to offer and can therefore be a central force for external motivation.

Overall, it can be said that German universities as organizations cannot motivate their operational core of academics by means of career incentives. They can neither hinder nor help careers. As we have shown in our introduction there were historically good reasons for the rather weak power structure within German academia, in particular with regard to the research function of universities. However, in many other university systems around the world we find different power structures.[5] There are either clearly structured internal career paths within a university (e.g., USA, UK) or there is at least no strict house ban on internal calls—or if there is, it is not as rigidly enforced by informal norms as in Germany (e.g., France, Italy). Not only professors are affected by these career paths, but also the vast majority of the staff with academic duties in higher education (Hüther and Krücken 2011, 2013). Internal career paths towards a professorship are therefore possible in a number of other national university systems. If such an internal career is possible, those who pursue it will, to a much greater extent, be subject to an organizational logic. The organization can thus utilize career prospects and ambitions as incentives but also as negative sanctions to motivate desired behavior. This will make it possible to include behaviors not centered on the academic system (in particular, publications), but around the multiple other goals of the organization (such as the development of further education programs, the provision of additional services for students or regional economic cooperation).

The effective prevention of careers within universities in Germany is also a unique feature in comparison to other professional organizations. Sociological profession theory suggests that in contrast to other occupational groups, professionals are more likely to pursue careers between different organizations (Scott 1966). However, this does not mean that the organizations in question do not try to break this logic, at least partially. Internal careers of medical doctors in hospitals for example, are quite common. The same applies to large law firms with their distinctive internal career patterns (Heinz et al. 2005). The complete renunciation of this organizational power instrument within the German university system is thus neither typical in an international comparison of higher education systems nor in

[5] See, for example, Musselin (2010), Enders (2000b).

comparison with other professional organizations, business firms or public administration.

2.2 Organizational Power

In addition to personnel power, there is a second source of power in most organizations based on the power to exclude members: organizational power. According to Luhmann (1980), organizational power is characterized by the power of organizations to exclude members if they fail to meet the minimum standards of the organization. Minimum standards include the acceptance of the formal rules within an organization or refraining from actions which are counterproductive to the organization's goals. Organizational power is therefore a means to shore up at least the basic motivation of the members to pursue the organization's goals. But here as well, German universities can hardly use this source of power to motivate their academics.

Let us take a look at the professors first. It becomes obvious that organizational power as a foundation for motivation cannot be utilized. There is no procedure for "how a professor could be fired even if he or she is lazy, incompetent or no longer needed" (Dilger 2007, p. 109).

The reason for this is that the vast majority of professors in Germany have life tenure with a special civil servant status. Furthermore, freedom in research and teaching is guaranteed by the German constitution in Article 5, paragraph 3. This freedom is closely connected with the German tradition of the independence of professors (Baker and Lenhardt 2008). Due to the traditionally high status of professors in the German system and the highly detrimental effects of the Third Reich on the individual autonomy and the freedom in research and teaching German professors are protected by the constitution, also vis-à-vis the organization in which they are embedded. This leads to a strong, secure position of professors toward their organization which makes it extremely difficult to dismiss a professor. A credible threat of exclusion from the organization is therefore nearly impossible. The organization's leadership would have to take recourse to risky legal action in order to assert its organizational power which, considering the high costs in terms of time and personnel and the uncertain outcome, only happens very rarely. Summing up: Organizational power is not relevant to professors. The organization has no credible sanctions at its disposal and therefore cannot generate motivational effects.

Organizational power is also ineffective towards the great majority of other academic personnel. As with personnel power, organizational power over academic staff rests with the professors. Professors, not deans or university management "hire and potentially fire" (Dilger 2007, p. 103) junior academic staff. It is therefore professors who decide about the inclusion or exclusion of junior members of the organization. The fact that professors and not institutes or faculties are the gate keepers for academic careers up to the level of the professorships is a result of the traditional German chair structure with its emphasis on professorial independence.

We also find here, as before with personnel power, that power and the resulting motivational possibilities are concentrated in the hand of professors.

Overall, it can be said that organizational power as a means of motivating behavior plays no part in the German system at the organizational level. Although we can assume that academics do not want to be excluded from the organization, this decision is not made by the organization: neither for the professors who usually have life tenure, nor for the young academics with their fixed-term contracts. This lack of organizational power at German universities is unique in international comparison. Firstly, the special employment status as a civil servant and the resulting general lack of grounds for dismissal of professors does not exist in many countries. If it did, as in Sweden (Askling 2001), it has since been abolished. In the Netherlands and Great Britain professors can be excluded from the organization (de Weert 2000; Fulton 2000). In the American system, a strengthening of organizational power can be observed. Latest figures show that the tenure-track system is declining and short-term contracts, which do not have to be renewed, are on the rise at higher education institutions.[6] Whereas personnel power through the career incentives offered by the tenure track was prevalent in the 1970s in universities in the USA, today the use of organizational power by the use of short-term contracts has become easier. Similar developments can also be observed at British universities (Kreckel 2008; Fulton 2000).

It has been shown above that neither personnel nor organizational power plays a part in supporting the motivation of professors in Germany. In relation to the academic staff below the level of professor, the chair holder has access to both sources of power. Since the professors are not firmly bound to the multiple organizational goals it is at least questionable whether professors motivate their staff towards the organization's goals. This is a severe problem for the university because it cannot orientate its members towards its multiple goals (e.g., teaching, academic self-administration, knowledge transfer), nor does it have the power to enforce such an orientation.

2.3 *Power over Resources*

Since German universities have hardly any career incentives or sanctions, the question arises of whether there are alternatives to these typical sources of power in organizations. A central alternative is the power over resources. In this case, power is built up by the allocation of resources in order to give incentives both to

[6] The number of tenured academics in the USA declined from 65 % in 1980/1981 to 49 % in 2007/2008 [cf. National Center for Education Statistics (NCES)]. According to Chait (2002, p. 19) the number of part-time professorships nearly doubled from 22 % in 1970 to 41 % in 1995. According to Donoghue (2008) this trend is particularly dramatic in subjects like humanities from which no immediate economic utility can be expected, or at higher education institutions that are orientated toward profit and/or training.

single members or to organizational units (Hüther and Krücken 2013). In the following we will concentrate on the individual level of academics.

At the individual level, the recent changes to professors' remuneration and fixed-term funding are of particular relevance in Germany. In principle, both of these reform measures strengthen the organizational potential for exercising power. Following our previous line of argumentation, the innovation is that granting or withdrawing funding and bonuses is delegated to the university as an organization and are no longer the object of negotiations between the professor and the relevant state ministry, as was the case in the traditional system. In principle, as Musselin (2013) stated, these reforms change the relationship between academics and their university. This becomes especially apparent with performance-related pay that can be granted for exceptional achievements in the fields of research, teaching and/or academic self-administration. Incentives are therefore possible for behavior that is central to the organization but that is not necessarily of equally high importance to the academic system.

However, problems are recognizable with both the remuneration and the fixed-term funding. First, it should be noted that in the German case both sources of power can potentially only be exercised over certain groups of members. The new dynamic of the remuneration and funding structures is initially only applicable to professors; other academic staff is not directly affected. A direct motivational effect can therefore only pertain to the professors. Indirectly, the new incentives could reach other academic members via the professors if they use their organizational and personnel power according to the incentive systems.

Second, there are limits to the efficacy of the application of variable remuneration and funding structures that are primarily related to differences between subjects and disciplines. Performance-related pay will not be an attractive incentive to professors who have sources of income from outside the university. This can quite often be the case in medical, law or engineering departments. Precisely the same differences apply to the fixed-term funding of chair resources. Classic liberal arts and humanities subjects, in which research is chiefly individual research, are more independent in this respect than more strongly networked sciences that, like most of the natural sciences, require significant human and material resources in order to be able to conduct research at all (Jansen et al. 2007). Therefore, we can assume that performance-related pay and resources will only be sought after by some professors for whom they can then work as motivational incentives. However, motivational effects will not or only hardly be possible among professors who do not seek the incentives or do not think they are important.

In addition, negative sanctions, such as reducing funding and other resources, are limited by law. Professorial resources cannot be reduced at will because the constitutional right to freedom of research and teaching guarantees minimal resources for a professor (Seidler 2004; BVerfG 2010, p. 114). Negative sanctions can therefore only be applied in a limited way. Not only the resources but also the remuneration of professors is guaranteed by the constitution. In 2012 the constitutional court ruled that the regulations governing performance-related pay of professors were unconstitutional because the basic salary without the incentives was too low (BVerfG 2012). The federal states that have already drafted new laws have incorporated a higher basic salary but a lower performance-related bonus (HRK 2012). It is to be expected that the motivational effects of performance-related pay, which has been questioned in the literature cited in the introduction, will further diminish in the future.

A further point is important: incentives—as opposed to sanctions—are expensive. If the rewards are really meant to be motivating, they have to be paid out. Incentives cost money and demand flexible financial resources. Both areas, performance-related resources and performance-related pay, face problems. When performance-related pay was introduced, the overall amount of money for the salary of professors was not increased. No extra funding for incentives was provided, which means that possible performance-related bonuses are very small. It can be assumed that there is not enough money for incentivizing professors. If this is the case, the possible motivational effect diminishes because despite good performance no or only a very small amount of performance-related bonuses are available. However, as we assume a particular strong intrinsic motivation among professors, such incentivizing strategies can negatively affect this motivational base as research on the crowding out phenomenon has shown consistently (e.g., Frey and Oberholzer-Gee 1997; Osterloh and Frey 2000).

Incentives based on resources face a similar problem. First, it is necessary to mention that the German university system is comparatively underfinanced (OECD 2012, p. 244). Again this leads to the question of whether incentives can be a credible strategy for motivation if finances are tight. Second, it is doubtful that within the German system the universities have sufficient freely disposable funds for incentives. In recent years, so called global funds have been introduced to increase flexibility in the allocation of funds. However, in practice the new flexibility has little effect because, for example, existing personnel plans at universities still dictate how most of the funds are used (Schubert and Schmoch 2010). Overall, it can be seen that the attempts to introduce motivational incentives by means of positive sanctions via resources in Germany are limited by a lack of funding. Due to legal restrictions on negative sanctions, funds can rarely be generated by negatively sanctioning low performing professors and departments.

In accordance with the literature on motivation in academia cited in the introduction, we can say that the reforms so far have been unilaterally directed at bolstering the organizational level by means of allocating financial resources so that economic capital becomes the dominant steering medium. It could be objected that in academia and higher education money is the wrong steering medium because it is not capable of affecting the behavior of academics or academic self-organization (Minssen and Wilkesmann 2003; Luhmann 1992). Even if this is not true, the effect of incentives in the academic system is different than, for example, in business firms. Since in the academic system reputation is the more important currency than money, we can assume that money will only have a steering influence if reputation is simultaneously increased. This would also explain, for example, why financial incentives for better teaching have been more or less ignored by professors in Germany (Wilkesmann and Schmid 2012).

3 Summary and Discussion

Overall, it can be said that German universities traditionally had only very limited means to motivate academics to achieve the multiple goals of the organization. Universities could exercise neither career incentives nor threats of exclusion. In international comparison, the strong emphasis on academic freedom marks a significant weakness of the organizational level with respect to the individual academic. Despite the many reforms in Germany, there is still a considerable weakness of both personnel and organizational power. The newly introduced incentive systems have, however, strengthened the organizational level because decisions about the level of remuneration and resources are made within the universities. The organization can better motivate academics to pursue its multiple goals. Incentives therefore change the power structures within universities to the benefit of the organization and its leadership. From this perspective the new incentives can be viewed as one aspect of how German universities are becoming actors in their own right.

But the article has also shown that there are severe problems associated with incentives based on remuneration and resources. In agreement with the literature on motivation cited in the introduction, it is doubtful whether professors actually pursue the incentives offered or deem them relevant to themselves. If the incentives are not pursued or are not important to the academics, the power they could give to the organization and its leadership is limited. Consequently, the motivational effect also has to be viewed as limited. In addition, incentive systems are cost intensive and for an underfunded system in which the option to impose negative sanctions is severely restricted by the constitution and public sector employment legislation, as in Germany, they can hardly be considered an appropriate means of motivation.

We would like to address two pertinent issues resulting from our analysis. First, what comparative advantages and disadvantages do result from the peculiar power structure within German universities? Second, what are sensible options in order to strengthen the organizational level vis-à-vis its individual academic members?

Let us begin with the first question. The weakness of the organizational level with respect to the individual academic might be considered as strength with regard to the pursuit of the universities' research function as academic freedom is a precondition of scientific creativity. The strong individual orientation at trans-organizational and trans-national research networks and scientific communities goes hand in hand with a weak organizational level. However, with regard to other goals of universities (e.g., mass education, continuing education, knowledge and technology transfer) the current power structure might be more problematic. The internal motivation to pursue such goals in many cases is rather low. As such motivation is not cherished by the wider scientific community and the weak organizational power structure can hardly motivate either, universities might systematically fall short of such goals and, ultimately, lose legitimacy. Here, a more nuanced discussion on the multiple goals and related means of universities is needed. Furthermore, one should add that the effects of attempts at strengthening the organizational level of universities become increasingly visible in a variety of European countries (Paradeise et al. 2009). Such effects to a large extent include unintended effects (e.g., Capano 2008; Mignot 2003; Enders et al. 2013). To give just one example: In Great Britain, the strengthening of the organizational level vis-à-vis its individual academic members between 1992 and 2007 has led to a decrease of their organizational commitment, and the percentage of universities professors who contemplate leaving academia for good is the highest in international comparisons (Jacob and Teichler 2011, p. 49, 142). Taking such an unintended effect into account might advice caution on shifting the power balance between the organization and its academic members.

With regard to the second question we would briefly like to discuss two concrete options to strengthen the organizational level vis-à-vis its individual academic members: the shift from a chair to a department system, and the introduction of tenure track positions. Both options are vividly discussed, and partially also implemented in Germany.

One might strengthen the organizational level by shifting to a department system as it has been done in a variety of national systems before (Neave and Rhoades 1987; Enders 2000a). In this case, the dean, not the university leadership will be strengthened. For the German system, we expect strong resistance among professors who will lose some of their power vis-à-vis the dean and little efforts among the deans to exercise their power. Though the formal power structure of the dean has increased in most higher education laws of the states in Germany (Hüther 2008), empirical studies show that this increase is hardly reflected in changing practices among deans (Scholkmann 2011). Furthermore, the new power structure is also limited by legal constraints as the Constitutional Court recognizes the individual level of the professor as the most important one when it comes to the defense of academic freedom, while the organization and organizational units like departments are rather seen as a possible threat to such freedom (Baker and Lenhardt 2008; BVerfG 2010). Another way of strengthening the organizational level and its power is the introduction of tenure track positions. This allows for organizational careers up to the level of the full professor. In the terminology we employed in the article 'personnel power' can be exercised and the role of the organization and its leadership becomes more important as compared to the traditional system. Multiple organizational goals can be connected more easily to the career trajectories of individual within their university setting. However, also here one must expect a variety of unintended effects, which result from introducing an entirely new career system within a university system that for a long time had been based on the premise that academic careers do not take place within the work organization, but outside.

At the end of our article we would like to point out some relevant future research perspectives. Combining our perspective from power and organizational theory more closely with the perspective of motivational theory on the topic could yield a very interesting perspective for future research that goes beyond this article. The focus would be on both the relationship between the individual and the organization and also on the effects of incentive strategies on the academic system as a whole. In general, manifold unintended effects might result from different reform efforts aiming at shifting the power balance between the organization and the individual that are worthwhile to explore (Krücken 2014). An additional perspective would be to widen the notion of power we chose. As we focused mainly on formal aspects of exercising power, one should also try to incorporate informal aspects of power. University organizations in particular are characterized by manifold mechanisms for exercising power that are not tied to formal rules but are very important in creating desired behavior. The same is true for the aspects of self-disciplining and legitimation of power emphasized by the more recent power concepts which we mentioned in the beginning of Part 2.

The article also highlights that an international comparative perspective on power structures would be desirable for research on universities. Whereas changes in the overall governance structures have been well studied, hardly any material on the power structures on the organizational level is available. This would be particularly important to better assess the preconditions and effects of the new incentive systems. In the German case, for example, we cannot find any amplifying effects between the three observed power structures. This is due to the extreme weakness of both organizational and personnel power. However, the article also showed that organizational and/or personnel power is stronger in other higher education systems. The question of the type of interaction between the power structures and its influence on the effects of the new incentive systems is, in our opinion, a central question for future research.

References

Askling B (2001) Higher education and academic staff in a period of policy and system change. High Educ 41(1–2):157–181

Baker DP, Lenhardt G (2008) The institutional crisis of the German research university. High Educ Pol 21(1):49–64

Ben-David J (1971) The scientist's role in society. A comparative study. Prentice-Hall, Englewood Cliffs

Blau PM, Scott RW (1969) Formal organizations. A comparative approach. Routledge; Kegan Paul, London

BMBF (2014) Datenportal. Studienanfänger/-innen im 1. Hochschulsemester und Studienanfängerquoten nach Geschlecht und Land des Erwerbs der Hochschulzugangsberechtigung. http://www.datenportal.bmbf.de/portal/Tabelle-2.5.73.xls. Accessed 18 May 2014

Bode I (2010) Der Zweck heil(ig)t die Mittel? Ökonomisierung und Organisationsdynamik im Krankenhaussektor. In: Endreß M, Matys T (eds) Die Ökonomie der Organisation - die Organisation der Ökonomie. VS Verlag, Wiesbaden, pp 63–92

Bosbach E (2009) Von Bologna nach Boston? Perspektiven und Reformansätze in der Doktorandenausbildung anhand eines Vergleichs zwischen Deutschland und den USA. Akademische Verlags-Anstalt, Leipzig

Brown AD, Lewis MA (2011) Identities, discipline and routines. Organ Stud 32(7):871–895

Bruneau W, Savage DC (2002) Counting out the scholars: the case against performance indicators in higher education. James Lorimer & Company, Toronto

Brunsson N, Sahlin-Andersson K (2000) Constructing organizations: the example of public sector reform. Organ Stud 21(4):721–746

Bundesamt S (2013) Bildung und Kultur: Personal an Hochschulen 2012. Statistisches Bundesamt, Wiesbaden

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

BVerfG (2010) 1 BvR 748/06 vom 20.7.2010

BVerfG (2012) 2 BvL 4/10 vom 14.02.2012

Capano G (2008) Looking for serendipity: the problematical reform of government within Italy's universities. High Educ 55(4):481–504

Chait R (2002) The questions of tenure. Harvard University Press, Cambridge, MA

Clegg SR, Courpasson D, Phillips N (2006) Power and organizations. SAGE Publications, Los Angeles, CA

Cyert RM, March JG (1963) A behavioral theory of the firm. Prentice-Hall, Englewood Cliffs, NJ

de Boer H, Enders J, Leisyte L (2007a) Public sector reform in Dutch higher education: the organizational transformation of the university. Public Adm 85(1):27–46

de Boer H, Enders J, Schimank U (2007b) On the way towards new public management? The governance of university systems in England, the Netherlands, Austria, and Germany. In: Jansen D (ed) New forms of governance in research organizations, Disciplinary approaches, interfaces and integration. Springer, Dordrecht, pp 137–152

de Weert E (2000) Academic staff in the Netherlands. In: Enders J (ed) Employment and working conditions of academic staff in Europe. German Trade Union for Education & Science, Frankfurt/Main, pp 168–180

Diefenbach T, Sillince JAA (2011) Formal and informal hierarchy in different types of organization. Organ Stud 32(11):1515–1537

Diefenbach T, By RT, Klarner P (2009) A multi-dimensional analysis of managers' power. Functional, socio-political, interpretive-discursive, and socio-cultural approaches. Manag Rev 20(4):413–431

Dilger A (2007) German universities as state-sponsored co-operatives. Manag Rev 18(2):102–116

Donoghue F (2008) The last professors: the corporate university and the fate of the humanities. Fordham University Press, New York

Emerson RM (1962) Power-dependence relations. Am Sociol Rev 27(1):31–41

Enders J (2000a) Academic staff in the European Union. In: Enders J (ed) Employment and working conditions of academic staff in Europe. German Trade Union for Education & Science, Frankfurt/Main, pp 29–53

Enders J (ed) (2000b) Employment and working conditions of academic staff in Europe. GEW, Frankfurt/Main

Enders J (2008) Professor werden ist sehr schwer, professor sein dann gar nicht mehr? Ein Betrag zur Personalstrukturreform an den Hochschulen. In: Matthies H, Simon D (eds) Wissenschaft unter Beobachtung. Effekte und Defekte von Evaluationen. VS Verlag, Wiesbaden, pp 83–99

Enders J, Bornmann L (2001) Karriere mit Doktortitel? Ausbildung, Berufsverlauf und Berufserfolg von Promovierten. Campus-Verlag, Frankfurt/Main u.a

Enders J, Boer H, Weyer E (2013) Regulatory autonomy and performance: the reform of higher education re-visited. High Educ 65(1):5–23

Espeland WN, Sauder M (2007) Rankings and reactivity: how public measures recreate social worlds. Am J Sociol 113(1):1–40

Etzioni A (1964) Modern organizations. Prentice-Hall, Englewood Cliffs, NJ

Flexner A (1930) Universities - American, English, German. Oxford University Press, Oxford

Frey BS, Oberholzer-Gee F (1997) The cost of price incentives: an empirical analysis of motivation crowding-out. Am Econ Rev 87:746–755

Frey BS, Homberg F, Osterloh M (2013) Organizational control systems and pay-for-performance in the public service. Organ Stud 34(7):949–972

Fulton O (2000) Academic staff in the United Kingdom. In: Enders J (ed) Employment and working conditions of academic staff in Europe. German Trade Union for Education & Science, Frankfurt/Main, pp 224–234
Hagstrom WO (1965) The scientific community. Basic Books, New York
Hannan MT, Freeman J (1984) Structural inertia and organizational change. Am Sociol Rev 49 (2):149–164
Heinz JP, Nelsen RL, Sandefur RL, Laumann EO (2005) Urban lawyers: the new social structure of the bar. University of Chicago Press, Chicago
Heinze T, Shapira PRJ, Senker J (2009) Organizational and institutional influences on creativity in scientific research. Res Policy 38:610–623
HRK (2012) HRK Senate criticises the draft amendments to the "W" professorial salary scale. http://www.hrk.de/press/press-releases/press-release/meldung/hrk-senate-criticises-the-draft-amendments-to-the-w-professorial-salary-scale-2340/
Hüther O (2008) Starke Dekane - Schwache Professoren? Vergleich und Analyse der deutschen Landeshochschulgesetze. Hochschulmanagement 3(1):23–27
Hüther O (2010) Von der Kollegialität zur Hierarchie? Eine Analyse des New Managerialism in den Landeshochschulgesetzen. VS Verlag, Wiesbaden
Hüther O, Krücken G (2011) Wissenschaftliche Karriere und Beschäftigungsbedingungen – organisationssoziologische Überlegungen zu den Grenzen neuer Steuerungsmodelle an deutschen Hochschulen. Soziale Welt 62(3):305–325
Hüther O, Krücken G (2013) Hierarchy and power: a conceptual analysis with particular reference to new public management reforms in German universities. Eur J High Educ 3(4):307–323
Jacob AK, Teichler U (2011) Der Wandel des Hochschullehrerberufs im internationalen Vergleich. Ergebnisse einer Befragung in den Jahren 2007/08. BMBF, Bonn
Jansen D, Wald A, Franke K, Schmoch U, Schubert T (2007) Drittmittel als Performanzindikator der wissenschaftlichen Forschung. Zum Einfluss von Rahmenbedingungen auf Forschungsleistungen. Kölner Zeitschrift für Soziologie und Sozialpsychologie 59(1):125–149
Kärreman D, Alvesson M (2004) Cages in tandem: management control, social identity, and identification in a knowledge-intensive firm. Organization 11(1):149–175
Kreckel R (2008) Großbritannien. In: Kreckel R (ed) Zwischen promotion und professur. Akademische Verlags-Anstalt, Leipzig, pp 124–165
Krücken G (2003) Learning the 'new, new thing': on the role of path dependency in university structures. High Educ 46(3):315–339
Krücken G (2008) Lässt sich Wissenschaft managen? Wissenschaftsrecht 41(4):345–358
Krücken G (2014) Higher education reforms and unintended consequences: a research agenda. Studies of Higher Education.
Krücken G, Meier F (2006) Turning the university into an organizational actor. In: Drori GS, Meyer JW, Hwang H (eds) Globalization and organization. World society and organizational change. Oxford University Press, Oxford, pp 241–257
Krücken G, Blümel A, Kloke K (2013) The managerial turn in higher education? On the interplay of organizational and occupational change in German academica. Minerva 51(5):417–442
Lee FS (2007) The research assessment exercise, the state and the dominance of mainstream economics in British universities. Camb J Econ 31(2):309–325
Luhmann N (1964) Funktionen und Folgen formaler Organisation. Duncker & Humblot, Berlin
Luhmann N (1980) Trust and power. Two works by Niklas Luhmann. Wiley, Chichester
Luhmann N (1992) Die Wissenschaft der Gesellschaft. Frankfurt/M, Suhrkamp
Maasen S, Weingart P (2006) Unternehmerische Universität und neue Wissenskultur. die hochschule 15(1):19–45
MacRoberts MH, MacRoberts BR (1996) Problems of citation analysis. Scientometrics 36 (3):435–444
Martin B, Whitley R (2010) The UK research exercise: a case of regulatory capture? In: Whitley R, Gläser J, Engwall L (eds) Reconfiguring knowledge production. Changing authority relationships in the sciences and their consequences for intellectual innovation. Oxford University Press, Oxford, pp 51–80
Mayntz R (ed) (1968) Bürokratische organisation. Kiepenheuer & Witsch, Köln

Merton RK (1973) The sociology of science. Theoretical and empirical investigations. University of Chicago Press, Chicago, IL
Mignot G (2003) Who are the actors in the government of French universities? The paradoxal victory of deliberative leadership. High Educ 45(1):71–89
Minssen H, Wilkesmann U (2003) Lassen Hochschulen sich steuern? Soziale Welt 54(2):123–141
Münch R (2006) Die Entsorgung wissenschaftlicher Kreativität Latente Effekte der kennziffern-gesteuerten Mittelverteilung im akademischen Feld. die hochschule 15(2):98–122
Musselin C (2005) European academic labor markets in transition. High Educ 49(1):135–154
Musselin C (2010) The market for academics. Routledge, New York
Musselin C (2013) Redefinition of the relationships between academics and their university. High Educ 65(1):25–37
Neave G, Rhoades G (1987) The academic estate in western Europe. In: Clark B (ed) The academic profession. National, disciplinary and institutional settings. University of California Press, Berkeley, CA, pp 211–270
OECD (2012) Education at a glance 2012. OECD Publishing, Paris
Osterloh M (2012) New public management'versus ,Gelehrtenrepublik'- rankings als instrument der qualitätsbeurteilung in der wissenschaft? In: Wilkesmann U, Schmid C (eds) Hochschule als organisation. Organisationssoziologie. VS Verlag, Wiesbaden, pp 209–221
Osterloh M, Frey BS (2000) Motivation, knowledge transfer, and organizational forms. Organ Sci 11(5):538–550
Osterloh M, Frey BS (2009) Research governance in academia: are there alternatives to academic rankings? CESifo working paper. http://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID1460691_code968850.pdf?abstractid=1460691&mirid=1. Accessed 15 Apr 2014
Paradeise C, Bleiklie I, Ferlie E, Reale E (eds) (2009) University governance: western European comparative perspectives. Springer, Dordrecht
Partha D, David PA (1994) Toward a new economics of science. Res Policy 23(5):487–521
Pollitt C, Bouckaert G (2004) Public management reform: a comparative analysis. Oxford University Press, Oxford
Preston AM (1992) The birth of clinical accounting: a study of the emergence and transformations of discourses on costs and practices of accounting in U.S. hospitals. Acc Organ Soc 17(1):63–100
Reay T, Hinings CR (2005) The recomposition of an organizational field: health care in Alberta. Organ Stud 26(3):351–384
Röbken H (2007) Postgraduate studies in Germany - how much structure is not enough? South Afr J High Educ 21(8):1054–1066
Royston G, Hinings CR (1996) Understanding radical organizational change: bringing together the old and the new institutionalism. Acad Manag Rev 21(4):1022–1054
Schimank U (2001) Festgefahrene Gemischtwarenläden - Die deutschen Hochschulen als erfolgreich scheiternde Organisationen. In: Stölting E, Schimank U (eds) Die Krise der Universitäten. Westdeutscher Verlag, Wiesbaden, pp 223–242
Schofer E, Meyer JW (2005) The worldwide expansion of higher education in the twentieth century. Am Sociol Rev 70(6):898–920
Scholkmann A (2011) Deans in German universities: goal acceptance and task characteristics. Tert Educ Manag 17(4):337–353
Schubert T, Schmoch U (2010) Finanzierung der Hochschulforschung. In: Simon D, Knie A, Hornbostel S (eds) Handbuch Wissenschaftspolitik. VS Verlag, Wiesbaden, pp 244–261
Scott RW (1966) Professionals in bureaucracies - areas of conflict. In: Vollmer HM, Mills DL (eds) Professionalization. Prentice-Hall, Englewood Cliffs, NJ, pp 265–275
Scott RW (1981) Organizations: rational, natural, and open systems. Prentice-Hall, Englewood Cliffs, NJ
Seidler HH (2004) Hochschulfinanzierung, evaluation und mittelvergabe. In: Hartmer M, Detmer H (eds) Hochschulrecht. Ein Handbuch für die Praxis. Verlagsgruppe Hüthig-Jehle-Rehm, Heidelberg, pp 478–510

Shils E (1991) The idea of a university and the right to academic freedom. In: Altbach PG (ed) International higher education. An encyclopedia, vol 1. New York; London, pp. 1–22
Stern S (2004) Do scientists pay to be scientists? Manag Sci 50(6):835–853
Sutton JR (1984) Organizational autonomy and professional norms in science: a case study of the Lawrence Livermore Laboratory. Soc Stud Sci 14(2):197–224
Trow M (2010) Twentieth-century higher education: elite to mass to universal. JHU, Baltimore, MD
Wilkesmann U (2010) Die vier Dilemmata der wissenschaftlichen Weiterbildung. Zeitschrift für Soziologie der Erziehung und Sozialisation 30(1):28–42
Wilkesmann U, Schmid CJ (2012) The impacts of new governance on teaching at german universities. Findings from a national survey. High Educ 63(1):33–52

Performance Management and Incentive Systems in Research Organizations: Effects, Limits and Opportunities

Stefanie Ringelhan, Jutta Wollersheim, and Isabell M. Welpe

Abstract This chapter gives an overview on the effects, limits and opportunities of the different performance management and incentive systems employed in research organizations. Specifically, we discuss the existing performance management and incentive systems in research organizations as well as their advantages and disadvantages, and provide theoretically—as well as empirically—grounded recommendations for performance management and incentive systems in research organizations. In particular, we provide two recommendations. First, we advocate focusing on input governance as well as on trust in the intrinsic motivation of scholars, and independence as ways to ensure quality. Second, with regard to incentive systems, we recommend emphasizing informal-interpersonal acknowledgment.

1 Introduction

The investigation of performance management and incentive systems in research organizations[1] is important for three reasons. First, a profound understanding of appropriate performance management and incentive systems serves as a basis for successfully mastering challenges arising from the increasing competition concerning excellence in research and teaching within and between nations. Second, governments and the public are increasingly interested in evaluating the performance of scholars and research organizations. Consequently, performance management and incentive systems have changed. In Germany, the funding system builds increasingly on output instead of input control (Schmoch 2009) in order to measure whether funding by the state (based on tax payments) is used efficiently. It is unclear whether this transformation is based on a profound understanding of

[1] The term "research organization" predominantly refers to higher education institutions such as universities in the following.

S. Ringelhan (✉) • J. Wollersheim • I.M. Welpe
Chair for Strategy and Organization, Technische Universität München, Munich, Germany
e-mail: s.ringelhan@tum.de

I.M. Welpe et al. (eds.), *Incentives and Performance*,
DOI 10.1007/978-3-319-09785-5_6

academic performance. Third, current performance management and incentive systems have repeatedly been called into question in both the private and public sectors (Jacquart and Armstrong 2013; Weibel et al. 2010). Scholars point to distinctive characteristics of knowledge-oriented organizations such as research organizations (Binswanger 2011; Keller 2012; Minssen and Wilkesmann 2003; Osterloh 2010; Ringelhan et al. 2013; Schmoch et al. 2010) highlighting their unstandardized and complex tasks, many of which require creativity. Additionally, they often aim at education and knowledge generation. No functioning market exists for such tasks (Binswanger 2011). Furthermore, Weibel et al. (2010) demonstrate that pay-for-performance often produces hidden costs, for example, a decrease in intrinsic motivation. Thus, we propose that a positive relationship between pay-for-performance and academic performance is questionable.

What should performance management look like in order to be appropriate for research settings? Thus far, no systematic comparison of different performance management and incentive systems exists. In addition, alternative concepts and recommendations for performance management and incentive systems tailored for research settings are scarce (e.g., alternatives to pay-for-performance). Therefore, we aim to address this research gap by giving a literature-based systematic overview of (1) input and output performance management systems, and (2) existing monetary and non-monetary incentive systems. Note that incentive systems are part of performance management systems. Based on this overview, we derive recommendations for an appropriate performance management and incentive system for research organizations.

Our book chapter is structured as follows: In the next section, we present existing performance management and incentive systems in research organizations. Thereafter, we discuss our own empirical results on performance management and incentive systems, supplemented by results reported in the literature. Finally, we consolidate the existing conceptual and empirical knowledge by giving practical recommendations and by pointing to future research avenues.

2 Performance Management and Incentive Systems in Research Organizations

2.1 Performance Management Systems in Research Organizations

Performance management systems can ideally be categorized as output, process and input control. These three control forms are outlined in the following, one after the other, starting with output control. Note that quantitative and qualitative evaluations methods are not necessarily equivalent to output and input control. For example, informed peer-review may be described as a qualitative evaluation of scholars, which is comprised of quantitative *and* qualitative information.

Performance management systems in the form of output control have been introduced to higher education institutions with the rise of New Public Management (Wilkesmann and Würmseer 2009), in response to the call to become more competitive, efficient, effective, and accountable to stakeholders (Geuna and Martin 2003; Lange 2008; Melo et al. 2010; Wissenschaftsrat 2014). According to the German Wissenschaftsrat (2011), the new management system has been introduced in Germany for two reasons: first, to allocate resources in an indicator-based way, and second, to reward excellent performance. Generally, there are two main ways in which behavior is governed by output governance mechanisms: (1) indirectly, for example, by more transparency through rankings (Wissenschaftsrat 2011) and (2) directly, for example, by performance-based remuneration.

Indirect output governance mechanisms are characterized by a shift to entrepreneurial models; this has been fostered by New Public Management (Vilalta 2001). This leads to two indirect output control mechanisms: first, (quasi) competition (Binswanger 2011), for example, by relatively more third-party funding in higher education (Wilkesmann and Würmseer 2009); and second, more transparency, for example, by rankings of scholars or organizations (Wissenschaftsrat 2011).

Direct output controls consist of, for example, the distribution of funds according to outputs, performance-based payment and target agreements (Forschungsgemeinschaft 2012; Jaeger 2006a; Jaeger et al. 2005; Jansen et al. 2007; Melo et al. 2010; Sousa et al. 2010; Wilkesmann and Schmid 2011). Direct output control also involves more autonomy for higher education institutions while increasing external governance by stakeholders (Hicks 2012; Lange 2008), as well as introducing own (global) budget responsibility (Wilkesmann and Würmseer 2009). Furthermore, interim evaluations of junior professors as well as research and teaching evaluations also represent direct output control methods (Wilkesmann and Würmseer 2009). Note that interim evaluations of junior professors, as well as research and teaching evaluations, are not necessarily direct output controls, but they can be.

In output control, measurable performance indicators were established to govern and distribute funds (Forschungsgemeinschaft 2012; Jaeger 2006a; Jaeger et al. 2005; Jansen et al. 2007; Melo et al. 2010; Sousa et al. 2010; Wilkesmann and Schmid 2011). Examples measuring teaching include numbers of exams or graduates. Examples for assessing research performance include third-party funding, number of doctoral students, or number of publications and citations (Wissenschaftsrat 2011).

Which indicators can and should be used, for example, in formula-based budgeting, is highly debated (Aguinis et al. 2012; Baum 2011; Fiedler et al. 2008; Gendron 2008; Goodall 2009; Jansen et al. 2007; Kieser 2010, 2012; Schimank 2005; Wilkesmann and Würmseer 2009; Wissenschaftsrat 2011). Researchers have questioned whether performance indicators can measure performance in an exact and objective way, distinguishing good from bad performance (Wissenschaftsrat 2011). According to critics, performance in academia is manifold and complex; therefore, qualitative evaluation should be emphasized (Aguinis et al. 2012; Goodall 2009; Harmon 2006; Jansen et al. 2007; Keller 2012; Melo et al. 2010; Osterloh 2012; Schmoch et al. 2010).

While in some European countries, a wide and regular evaluation of higher education based on outputs was launched, for example, in the UK with the Research

Assessment Exercise in 1986 (Hicks 2012), such evaluation in Germany started much later (Jansen et al. 2007). The German Wissenschaftsrat (2013) concludes that the use of research ratings (which may not be only quantitative) supports the governance of research organizations by strategic steering and the provision of a performance proof. Furthermore, it considers comparing performance evaluations an important tool for governing research organizations. Research and teaching evaluations are applied as a quality management tool (Wilkesmann and Würmseer 2009).

To avoid undesired adaptation of scholars, the German Wissenschaftsrat (2013) proposed use of a wide variety of governance instruments, intending that output control is not the sole basis of evaluation. Previous research has discussed alternative ways of evaluating the quality of research performance and identified (informed) peer-review (Moed 2007; Osterloh 2010; Wissenschaftsrat 2011). Informed peer-review refers to a qualitative evaluation which is informed by quantitative and qualitative indicators. For interim evaluations of junior professors' research performance (Wilkesmann and Würmseer 2009), (informed) peer-review may be used. For example, the Netherlands has an evaluation system implemented based on peer review. The system is not used for distribution decisions but rather to develop strategies (Geuna and Martin 2003; Hicks 2012). In addition to (informed) peer review, post-publication review is also supported as an alternative and qualitative method of research evaluation (Kriegeskorte 2012).

Though New Public Management considers mainly output control, this is not the only form of management control. According to management control theory (Ouchi 1977, 1979), in certain situations, process and input control have to be applied (Frey et al. 2013; Frey and Osterloh 2011; Osterloh 2010; Ouchi 1979). In contrast to output control, process control can be applied when measurability of the outcome might be low. However, process control requires precise knowledge of the cause-and-effect relationship. Different forms of process control are, for instance, peer control as applied in accreditation of research organizations, and regulations when their abeyance is controlled for (Kaulisch and Huisman 2007; Wissenschaftsrat 2011). In the accreditation process, performance is evaluated, for instance, by the number of foreign professors, and creative performance is standardized.

According to management control theory, input control should be used (Frey et al. 2013; Frey and Osterloh 2011; Osterloh 2010; Ouchi 1979) when both the measurability of the outcome and the cause and effect relationship are low (difficult to determine and assess). In other words, input control should be applied when output and process control are not possible, as well as when tasks are complex and ambiguous, as in academia (Keller 2012).

Input control is exerted in academia, for example, in the United States. In US research universities, a formal periodic performance control on the basis of defined criteria (output control) is not widely used, according to Kieser (2010). Rather universities, such as Harvard, use the performance management method of relying on the basic principle of a thorough selection of qualified scholars (by a rigorous use of academic standards) who are then backed-up and can work autonomously (Letter to the New York Times, August 13th 1945, cited in Kieser (2010); see also www.

fas.harvard.edu/~research/greybook/principles.html). By this means, Harvard trusts in the intrinsic motivation of scholars, thus exerting input control. Furthermore, to trust in intrinsic motivation, independence should be granted to scholars as, for example, guaranteed by article 5 of the Basic Law for the Federal Republic of Germany. Applicants to research organizations ("input") can be evaluated against diverse selection criteria such as the ability for critical thinking, as well as an urge for exploration and experimentation, and professional standards (Osterloh and Frey 2011). Input control is also supported as the method of choice by some scholars (Osterloh and Frey 2011; Ouchi 1979).

The previous detailed explanation of management controls in research organizations is summarized in a tabular overview; Table 1 displays performance management systems discussed in the literature.

Table 1 Overview of performance management systems applied in research organizations divided by management control methods: output control, input control and process control

Management control method	Governance mechanism	Performance management tool	Evaluation method
Output control	Indirect	• Constitution of (quasi) competition	• Performance indicator
		• More transparency (e.g., through rankings)	
	Direct	• Distribution of funds	
		• Performance-based payment	
		• Target agreements	
		• Autonomy and budget responsibility	
		• Central evaluation (e.g., of teaching)	
Input control	Indirect	• Selection of qualified scientists	• Qualitative evaluation and selection criteria for applicants, such as professional standards (e.g., critical thinking, as well as an urge for exploration and experimentation)
		• Trust in the intrinsic motivation of employees	
		• Independence as, for example, guaranteed by article 5 of the Basic Law of the Federal Republic of Germany	
Process control	Direct	• Accreditation of research organizations	• Peer control, for example, in the accreditation process (evaluation of performance, for example, by the number of foreign professors)
		• Regulations that are obeyed	

2.2 Incentive Systems in Research Organizations

Incentives are traditionally set to enforce a certain desired behavior according to the reinforcement theory, operant conditioning and the law of effect[2] (Luthans and Stajkovic 1999; Stajkovic and Luthans 2003; Wissenschaftsrat 2011). Incentives in the higher education system can be given at an institutional level [e.g., performance-related budgeting (Wilkesmann and Schmid 2012)], at an individual level (Wissenschaftsrat 2011), or at the faculty level, and in a monetary or in a non-monetary form (Colyvas et al. 2002; Lam 2011; Staatsministerium 2005). In the following paragraphs, we first address monetary incentives reported in the literature, followed by non-monetary incentives. We also address whether these incentives are given at an institutional (or faculty) level or at an individual level.

Monetary incentives at the institutional level might be, for example, an award of centers of excellence to universities (Hicks 2012; Kaulisch and Huisman 2007), or target agreements at the faculty level which are linked to monetary incentives. With the award of a center of excellence, performance-based funding follows. Target agreements were introduced in Germany to higher education institutions at the end of the 1990s (Jaeger 2006b). Target agreements are reported as a way by which goals can be linked to financial incentives at a faculty level (Herrmann 2001), for example, at the Technische Universität München (Germany). At most German universities (87 %), target agreements are linked to budgeting (Jaeger 2006b). At the individual level, there are several monetary incentives, which are discussed in the following. In German higher education institutions, a new incentive system, the so-called "W-Besoldung", was introduced in 2004 in compliance with the *Bundestag*'s adoption of an amendment to the Framework Act in 1998 (Geuna and Martin 2003; Wissenschaftsrat 2011). Prior to 2004, professors in research organizations were paid according to the so-called "C-Besoldung". In the "C-Besoldung" system, adjustments of remuneration required appointment as professor to another research organization (Süß 2007). Otherwise, remuneration differences only occurred by an automatic age progression (seniority wage system) (Wilkesmann and Schmid 2012). In 2004, variable remuneration schemes were introduced as part of the "W-Besoldung" in order to enhance motivation and thus performance of professors (Süß 2007; Wissenschaftsrat 2011). Specifically, the "W-Besoldung" system incorporates a considerably lower base salary—down to two-thirds of the relative previous level (Wissenschaftsrat 2011)—but allows variable performance remuneration—up to one-third of the salary (Wilkesmann and Schmid 2012). Variable performance remuneration can be granted in the following three cases (Bundesgesetzblatt 2002): (1) when renegotiations with a professor are held to retain the professor at the research organization if the professor received a job offer from another research organization (Schäffer 2005); (2) in the case of special performance in research, teaching, art, continuing education and teaching of young academics; or (3) when carrying out a function or fulfilling a

---

[2] The law of effect states that a positive or pleasant outcome of a behavior leads to an increasing frequency of displaying this specific behavior.

special task in the research organization's self-administration or governance. Corresponding guidelines for the allocation of the variable remuneration have been defined by the respective Land and the respective organization, which possess autonomy in this area. The total amount of variable remuneration must be allocated in such a way that the average remuneration costs of the W2, W3, C2 and C4 professors equal the average remuneration costs for these professors in the year 2001 (Bundesgesetzblatt 2002). Thus, in the variable remuneration system of higher education institutions, it is not possible to increase or decrease general labor costs. Specific criteria for a high research performance (e.g., third party funding, patents) may therefore vary across research organizations according to their individual strategic goals. At the Institute for Research Information and Quality Assurance (Germany), for example, publication activities of scholars are considered in determining performance-based payment and discussed in job and appraisal interviews (Wissenschaftsrat 2014).

Hicks (2012) states that in many countries (e.g., United Kingdom, Germany, Australia) traditional command-and-control systems are replaced with quasi-market incentives in research. While in the old system, government agencies used to distribute funds directly to each individual scholar independent of relative performance, a broader shift to performance-based funding occurred. Research organizations provide such external (e.g., monetary) incentives in order to align the behavior of academics with the goals of the research organization (following the general idea of the principal-agent theory) (Minssen and Wilkesmann 2003; Wissenschaftsrat 2014). This movement also includes competition for funding, which displays the quasi-market side of the reform (Hicks 2012). Countries vary in the use of performance-based research funding. For example, in Norway and Italy, the share of performance-based funding out of total funding is 2 %, while it is 6 % in Australia, 10 % in New Zealand, and 25 % in the United Kingdom (Hicks 2012).

With regard to remuneration for special performance, for example, Kieser (2010) elaborates on the implementation of the performance bonus system for professors at the University of Konstanz. At the University of Konstanz, professors who meet expectations with regard to their performance (or fall short of expectations) receive their basic salary (level 0). For professors who exceed the performance usually expected of professors, a performance bonus system is implemented that ranges from level 1 *(performance in research, teaching, young academics advancement and/or continuing education that exceeds the performance that would usually be expected of professors)* to level 5 *(exceptional, internationally recognized and fundamental contributions in research and/or teaching, young academics advancement and continuing education of exceptional researchers of international and interdisciplinary reputation)*, with an additional level (above level 5) for exceptional researchers.[3] To stay within the limits of allocation, a certain percentage of allocation across levels is fixed, for example, only 2 % are

[3] It is not further explained in Kieser (2010) how exceptional performance at the University of Konstanz is defined and how it is evaluated.

allowed to be absolute top researchers (above level 5). For top researchers, no specific bonus is specified at the University of Konstanz.

Another example is the European Business School (Germany), which provides bonuses from gross profits attained through advanced education of the European Business School, to professors in accordance to their performance (Schäffer 2005). In addition to these variable remuneration possibilities, scholars can also attain bonuses in the form of prize money when winning an award (Krempkow 2015; Wilkesmann and Schmid 2012).

Additionally, there are two more monetary incentives present at research organizations. One is the allowance of additional tangible means, for example, third-party funds providing specific tangible assets (Wilkesmann and Würmseer 2009). Another monetary incentive is the allowance of additional human resources, for example, for the recruitment of additional employees (Jaeger et al. 2005; Wilkesmann and Würmseer 2009). For the allocation of tangible means and human resources, target agreements are often applied in German universities (Jaeger et al. 2005; Wilkesmann and Schmid 2011).

Apart from these monetary rewards, there are also non-monetary incentives for academic performance. At the institutional level, accreditations and regulations are present (Frey and Osterloh 2011; Kaulisch and Huisman 2007; Osterloh 2010; Ouchi 1979; Wissenschaftsrat 2011).

At the individual level, there are also non-monetary incentives which are discussed next. The scientific community can reward performance in several ways. One possibility is a formal-symbolic acknowledgement of exceptional performance in the form of, for example, an award, prize, or certificate (Schäffer 2005; Wilkesmann and Schmid 2012; Wilkesmann and Würmseer 2009), or also in the form of rankings (Hicks 2012). Furthermore, the scientific community can reward performance via informal-interpersonal acknowledgment (e.g., praise by colleagues) (Deemer et al. 2010; Minssen and Wilkesmann 2003). Such formal or informal acknowledgments are important, because they contribute to an increase in academic reputation (Minssen and Wilkesmann 2003). This is an important point to consider, as reputation and acknowledgments are major motivators for researchers, next to intrinsic motivation (Lam 2011). Hence, major motives of scholars arise from themselves (intrinsic motivation) or are externally provided, for instance, by their peers (acknowledgment). Therefore, governance of research organizations has only restricted (in)direct performance management possibilities via providing external incentives (Minssen and Wilkesmann 2003) and providing work conditions that, for example, foster autonomy which, among other things, is beneficial for intrinsic motivation. The vital importance of reputation within the scientific community is also supported by Hicks (2012), who states that the competition for prestige generated by performance-based funding schemes creates a powerful incentive within research organizations. Next to these formal-symbolic or informal-interpersonal acknowledgments, material (non-monetary) acknowledgments can be distributed in the form of, for example, advanced education, a dedicated research laboratory, or dedicated teaching rooms (Jaeger et al. 2005; Wilkesmann and Würmseer 2009). The European Business School, for example, offers attendance at advanced education (seminars) in combination with their awards (Schäffer 2005), which can support career advancement. Another non-monetary incentive in research organizations is an increase in autonomy/freedom of scope, for example, with regard to personal planning, or time provided in the form of a research-free semester or exemption of teaching duties (Wissenschaftsrat 2011). The European Business School, for example, provides a teaching load reduction of 2 hours per week per semester for the three researchers who perform best (Schäffer 2005). Furthermore, in the case of new contracts, a research-free-semester (sabbatical) can take place at the European Business School when the ex-ante defined research goals are met. Other non-monetary incentives that are used in research organizations include (1) increased responsibility, for example, for one's own budget (Wilkesmann and Würmseer 2009), for additional employees, or (external) lecturers; (2) additional resources, for example, more scientific employees, who support one's own research or teaching (Jaeger et al. 2005); as well as (3) target agreements contracted between the research organization and an individual scholar (Herrmann 2001; Hippler 2013; Jaeger 2006a, b; Jaeger et al. 2005; Schäffer 2005; Wilkesmann and Schmid 2011; Wilkesmann and Würmseer 2009; Wissenschaftsrat 2011). By now, most German higher education institutions have established target agreements with professors and/or faculties (Jaeger et al. 2005).

The monetary and non-monetary incentive systems in research organizations explained above are presented in a detailed, tabular form in Table 2.

Table 2 Overview of monetary and non-monetary incentive systems applied in research organizations

Incentive type	Governance level	Specific incentive
Monetary incentives	Institutional	• Awarding centers of excellence to universities inclusive of performance-based funding
	Faculty	• Target agreements linked to financial incentives
	Individual	• "W-Besoldung" (remuneration system in Germany) • When renegotiations with a professor are held • In case of special performance • When carrying out a function in research organization's self-administration/governance
		• Quasi-market incentives and competition
		• Prize money
		• Allowance of additional tangible resources
		• Allowance of additional human resources
		• Target agreements linked to financial incentives
Non-monetary incentives	Institutional	• Accreditation
		• Regulations
	Individual	• Formal-symbolic acknowledgment
		• Informal-interpersonal acknowledgment
		• Material acknowledgment
		• Increase in autonomy
		• Increased responsibility
		• Additional resources
		• Target agreements

3 Empirical Results on Performance Management and Incentive Systems in Research Organizations

In this chapter, we report primarily our own findings and complement these with findings from the literature. We start with a report on studies about performance management tools applied in research organizations. We also address contextual and psychological factors influencing the effects of such performance management tools. This is followed by a discussion of how research and teaching performance is measured and how it should be measured. In the following paragraphs, motivational effects of monetary and non-monetary incentives are contrasted.

Economists generally draw different conclusions regarding the effectiveness of pay-for-performance (Jacquart and Armstrong 2013). There are studies in the literature that report positive effects of performance-based financial rewards for easy and creative tasks (Bradler et al. 2012). However, pay-for-performance is also criticized in the literature. For example, Jacquart and Armstrong (2013) report that higher pay actually fails to foster performance of executives and instead undermines intrinsic work motivation, inhibits learning, leads to ignoring stakeholders, leads to neglecting long-term effects of their own decisions, and encourages unethical behavior. Moreover, a goal displacement effect at the individual level and a lock-in effect at the institutional level may be the result of pay-for-performance (Frey and Osterloh 2011). Hence, the mentioned authors make the following four propositions: reduce compensation, eliminate incentive plans, strengthen democracy and employee ownership, and use more validated hiring methods for top executives. Similar suggestions are given by DeNisi and Smith (2014), who propose that bundles of human resource practices aligned to the organization's strategic goals can create a climate for performance. These aspects might also be important to consider in a higher education setting.

The importance of managing employees and focusing on their concerns is emphasized by the results of our cross-sectional survey of young scholars (doctoral students, postdoctoral students, junior and assistant professors) in the fields of business or economics. Our study supports the crucial role of intrinsic and extrinsic work motivation, as well as that of job satisfaction in academia. Our findings reveal that intrinsic work motivation, extrinsic work motivation and job satisfaction have a direct influence on research performance (Ringelhan et al. 2013). According to our study, intrinsic work motivation increased job satisfaction, whereas extrinsic work motivation did not increase job satisfaction. This finding highlights the particular importance of intrinsic work motivation for performance management in academia (Ringelhan et al. 2013; Wollersheim et al. 2014a). Interestingly, the contextual factor of whether a university was involved in the excellence initiative (e.g., graduate schools and/or clusters of excellence) moderated the effect of extrinsic work motivation on academic performance (Wollersheim et al. 2014a). Specifically, Wollersheim et al. (2014a) observed that highly extrinsically motivated scholars who were employed by faculties that were involved in the excellence initiative performed significantly worse than highly extrinsically motivated scholars who were employed by faculties not involved in the excellence initiative. Another individual factor that influenced academic success was gender (Wollersheim et al. 2014a), with women outperforming male scholars.[4]

To analyze how scientific performance *is* measured and how it *should be* measured, we compared results from our survey among university professors on German-speaking business and economics faculties with results from our survey of young scholars at German-speaking business or economics faculties (Wollersheim et al. 2014b). Professors indicated that research performance *currently is* measured primarily by quantitative criteria at their faculties, such as by the number of publications (especially highly ranked journal publications), acquired/spent third-party funding, as well as the reception of a job offer from another research organization, awards, or the number of (post)doctoral students. These results indicate that, currently, qualitative criteria are used less than quantitative criteria to evaluate research performance at German-speaking business and economics faculties. Regarding the question of how research performance *should be* measured, our survey among young scholars at German-speaking business and economics faculties revealed that qualitative research performance criteria were considered to be significantly more important than qualitative research performance criteria. Specifically, young scholars rated scientific knowledge gain (the qualitative research performance criterion rated highest in our survey) significantly higher than the number of publications in highly ranked journals (the quantitative research performance criterion rated highest in our survey). These results have been supported by our student survey. A similar result pattern is present for teaching performance criteria. Here, qualitative teaching performance criteria have also been rated higher than quantitative teaching performance criteria. In particular, teaching independent working, learning, and problem solving (the qualitative teaching performance criterion rated highest in our survey) was rated significantly higher than scale and scope of teaching (the quantitative teaching performance criterion rated highest in our survey) (Wollersheim et al. 2014b).

Furthermore, our survey of young scholars provides insights into the effects of monetary and non-monetary incentives (Wollersheim et al. 2014b). The results of our study indicate that acknowledgment by the supervising professor motivated more to show central working behaviors in research (e.g., to pursue creative research ideas) than did autonomy with regard to the research topic or monetary incentives for good research performance. Furthermore, the young scholars participating in our study evaluated informal-interpersonal acknowledgment (e.g., praise by a supervising professor) as more motivating for their own research performance than other non-monetary incentives such as formal-symbolic acknowledgment in the form of awards or certificates (Wollersheim et al. 2014b). This result was supported by our survey of professors, indicating the crucial importance of non-monetary incentives, such as acknowledgment, to increase the performance of employees in knowledge-intensive settings.

[4] However, this effect was not robust when using research performance rather than academic performance as the criterion.

4 Recommendations for Theory and Practice and Conclusions

Different performance management and incentive systems have their advantages and disadvantages. Output control may increase accountability (Geuna and Martin 2003; Lange 2008; Melo et al. 2010; Wissenschaftsrat 2014). However, the measurement of qualitative performance in particular in the multidimensionality of academic performance is difficult. On the other hand, an input control also has its drawbacks, for example, group-thinking and cronyism among peers (Osterloh 2010). Nonetheless, due to the measurability problem as well as the difficulty in distinguishing between cause and effect, input control appears to be the more suitable management technique compared to output control in academia (Osterloh 2010). Input control should accordingly be implemented via a profound selection of employees, socialization of employees (internalized norms and standards), and via the application of responsibility and trust as central concepts. Another benefit of input control is the supposed sustainment of the 'taste for science' (intrinsic motivation), while New Public Management output control is proposed to substitute a 'taste for publication' (extrinsic motivation) for the taste of science (Osterloh 2010). Hence, unintended effects may take place as a result of New Public Management output control, which also imposes quasi-markets and may hinder meaningful advancements in science (Binswanger 2011; Hicks 2012; Lange 2008; Wilkesmann and Würmseer 2009). Furthermore, as a result of such an attentional shift to quantifiable research output, teaching can fall behind (Wissenschaftsrat 2011). A major change set in place by New Public Management output control is the introduction of monetary incentives in the form of performance-based remuneration. Given that performance-based remuneration aims for excellence, an objective and transparent performance-based remuneration may increase equity perceptions (Hicks 2012). A strong incentive effect of performance-based remuneration may accrue from the competition for reputation and status which performance-based remuneration creates (Hicks 2012). Geuna and Martin (2003, p. 277) examined the advantages and disadvantages of performance-based monetary incentives and conclude, "while initial benefits may outweigh the costs, over time such a system seems to produce diminishing returns." In line with this conclusion, Mukherjee (2013) proposes that a research organization cannot be governed solely from above. Hence, we suggest placing more emphasis on input control, intrinsic motivation of employed scholars, trust, autonomy and praise.

Based on our literature review and our empirical results, we derived the following recommendations for the design of an adequate performance management and incentive system in research organizations, as well as for future research in this field.

4.1 Recommendations for Practice

(1) In the realm of input control, conduct a thorough and careful selection of scholars. Ideally, intellectually gifted scholars with high professional norms should be selected (Kieser 2010; Osterloh and Frey 2011; Ouchi 1977, 1979).
(2) Account for intrinsic work motivation, which can increase job satisfaction and can in turn increase research performance (Ringelhan et al. 2013).
(3) Consider the multidimensionality of academic performance (Aguinis et al. 2012; Goodall 2009; Harmon 2006; Jansen et al. 2007; Keller 2012; Melo et al. 2010; Osterloh 2012; Schmoch et al. 2010) rather than focusing, for example, solely on research performance.
(4) Focus more on quality than on quantity with regard to the evaluation of academic performance (Wollersheim et al. 2014b).
(5) Focus more on acknowledgment rather than monetary incentives; in particular, informal-interpersonal acknowledgment appears to be motivating (Wollersheim et al. 2014b).

The most central recommendations for an adequate performance management and incentive system in research organizations are summarized in Fig. 1. These were derived based on our own empirical studies and on existing literature. While input control already addresses recruitment aspects such as a thorough selection of scholars, management of performance by output control and incentive systems addresses post-recruitment aspects, which should be refined in order to be applied in a more adequate manner in academia.

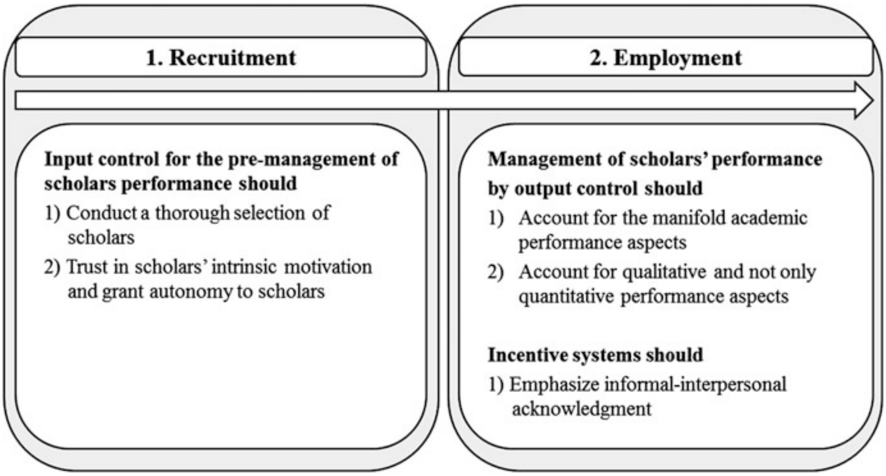


Fig. 1 Recommendations for an adequate performance management and incentive system in research organizations

4.2 Recommendations for Future Research

(1) Further develop existing performance indicators. In doing so, as many academic performance dimensions as possible should be considered and the perspectives of various research stakeholders must be embraced.
(2) Contemplate incentives for research organizations for input rather than output governance.
(3) Evaluate organizations that attempt quality control of research organizations, for instance, accreditation organizations, which set standards and greatly influence research organizations.
(4) Empirically analyze in depth the effects of different non-monetary incentives on performance in research and teaching.

In conclusion, performance management and incentive systems in research organizations face severe problems. Drastic changes in governance have been set in place with New Public Management. However, a proactive search for alternative or additional evaluating and governing mechanisms arose. This led to the investigation of solutions for performance management and incentive systems in research organizations. It also lays the foundation for highly interesting and relevant studies in the field of science economy.

References

Aguinis H, Suárez-González I, Lannelongue G, Joo H (2012) Scholarly impact revisited. Acad Manag Perspect 26(2):105–132

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 (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

Bradler C, Neckermann S, Warnke AS (2012). Anreize und Kreativität. Paper presented at Innovation, Leistungsmessung und Anreizsysteme in Wissenschaft und Wirtschaft – Governance wissensintensiver Organisationen, Munich, Germany

Bundesgesetzblatt (2002) Gesetz zur Reform der Professorenbesoldung (Professorenbesoldungsreformgesetz – ProfBesReformG). Bonn

Colyvas J, Crow M, Gelijns A, Mazzoleni R, Nelson RR, Rosenberg N, Sampat BN (2002) How do university inventions get into practice? Manag Sci 48(1):61–72

Deemer ED, Martens MP, Buboltz WC (2010) Toward a tripartite model of research motivation: development and initial validation of the research motivation scale. J Career Assess 18 (3):292–309

DeNisi A, Smith CE (2014) Performance appraisal, performance management, and firm-level performance: a review, a proposed model, and new directions for future research. Acad Manag Ann 8(1):127–179

Fiedler M, Welpe IM, Lindlbauer K, Sattler K (2008) Denn wer da hat, dem wird gegeben: Publikationsproduktivität des BWL-Hochschullehrernachwuchses und deren wissenschaftlicher Betreuer. Z Betriebswirt 78(5):477–508

Forschungsgemeinschaft D (2012) Förderatlas 2012: Kennzahlen zur öffentlich finanzierten Forschung in Deutschland. Wiley, Weinheim
Frey BS, Osterloh M (2011) Ranking games. CREMA Working Paper 2011-16
Frey BS, Homberg F, Osterloh M (2013) Organizational control systems and pay-for-performance in the public service. Org Stud 34(7):949–972
Gendron Y (2008) Constituting the academic performer: the spectre of superficiality and stagnation in academia. Eur Account Rev 17:97–127
Geuna A, Martin BR (2003) University research evaluation and funding: an international comparison. Minerva 41(4):277–304
Goodall AH (2009) Highly cited leaders and the performance of research universities. Res Policy 38(7):1079–1092
Harmon MM (2006) Business research and Chinese patriotic poetry: how competition for status distorts the priority between research and teaching in U.S. business schools. Acad Manag Learn Educ 5(2):234–243
Herrmann WA (2001) Akademisches Controlling und hochschulinterne Zielvereinbarungen - Erfahrungsbericht. München, Gütersloh, 15 Jan 2001
Hicks D (2012) Performance-based university research funding systems. Res Policy 41 (2):251–261
Hippler H (2013) Mehr Handlungsfreiheit für die Hochschulen - Fragen an den Präsidenten der HRK. Forschung Lehre 10:804–806
Jacquart P, Armstrong JS (2013) The ombudsman: are top executives paid enough? An evidence-based review. Interfaces 43(6):580–589
Jaeger M (2006a) Leistungsbezogene Budgetierung an deutschen Universitäten. Wissenschaftsmanagement 3:32–38
Jaeger M (2006b) Steuerung an Hochschulen durch interne Zielvereinbarungen - Aktueller Stand der Entwicklungen. die hoschschule 2:55–66
Jaeger M, Leszczensky M, Orr D, Schwarzenberger A (2005) Formelgebundene Mittelvergabe und Zielvereinbarungen als Instrumente der Budgetierung an deutschen Hochschulen: Ergebnisse einer bundesweiten Befragung (HIS Kurzinformation A13/2005). HIS, Hannover
Jansen D, Wald A, Franke K, Schmoch U, Schubert T (2007) Drittmittel als Performanzindikator der wissenschaftlichen Forschung: Zum Einfluss von Rahmenbedingungen auf Forschungsleistungen. Kölner Zeitschrift für Soziologie und Sozialpsychologie 59:125–149
Kaulisch M, Huisman J (2007) Higher education in Germany: country report. Center for Higher Education Policy Studies (CHEPS), The Netherlands
Keller RT (2012) Predicting the performance and innovativeness of scientists and engineers. J Appl Psychol 97(1):225–233
Kieser A (2010) Unternehmen Wissenschaft? Leviathan 38(3):347–367
Kieser A (2012) JOURQUAL - der Gebrauch, nicht der Missbrauch, ist das Problem. Die Betriebswirtschaft 1(72):93–110
Krempkow R (2015) Research and teaching awards as elements of incentive systems in science. In: Welpe IM, Wollersheim J, Ringelhan S, Osterloh M (eds) Incentives and performance - governance of research organizations. Springer, Cham
Kriegeskorte N (2012) Open evaluation: a vision for entirely transparent post-publication peer review and rating for science. Front Comput Neurosci 6:1–18
Lam A (2011) What motivates academic scientists to engage in research commercialization: 'Gold', 'ribbon' or 'puzzle'? Res Policy 40:1354–1368
Lange S (2008) New Public Management und die Governance der Universitäten. Zeitschrift für Public Policy, Recht und Management 1:235–248
Luthans F, Stajkovic AD (1999) Reinforce for performance: the need to go beyond pay and even rewards. Acad Manag Exec 13(2):49–57

Melo AI, Sarrico CS, Radnor Z (2010) The influence of performance management systems on key actors in universities. Public Manag Rev 12(2):233–254
Minssen H, Wilkesmann U (2003) Lassen Hochschulen sich steuern? Soziale Welt 54(2):123–144
Moed HF (2007) The future of research evaluation rests with an intelligent combination of advanced metrics and transparent peer review. Sci Public Policy 34(8):575–583
Mukherjee J (2013) Zwischen Kapitänsdeck und Maschinenraum - Persönliche Einblicke eines Universitätspräsidenten. Forschung Lehre 10:812–813
Osterloh M (2010) Governance by numbers. Does it really work in research? Analyse und Kritik 2:267–283
Osterloh M (2012) 'New Public Management' versus 'Gelehrtenrepublik'. Rankings als Instrument der Qualitätsbeurteilung in der Wissenschaft? In: Wilkesmann U, Schmid CJ (eds) Hochschule als Organisation. VS - Verlag für Sozialwissenschaften, Münster, pp 209–221
Osterloh M, Frey BS (2011) Input control and random choice improving the selection process for journal articles, 4 September 2011. University of Zurich, Zurich
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
Ringelhan S, Wollersheim J, Welpe IM, Fiedler M, Spörrle M (2013) Work motivation and job satisfaction as antecedents of research performance: investigation of different mediation models. Journal of Business Economics (ZfB), Special Issue 3/2013:7–38
Schäffer U (2005) Incentivierung von Forschungsleistungen an der ebs. Oestrich-Winkel
Schimank U (2005) 'New Public Management' and the academic profession: reflections on the German situation. Minerva 43(4):361–376
Schmoch U (2009) Geeignete Ansätze zur Messung wissenschaftlicher Leistung. Beiträge zur Hochschulforschung 31(1):26–41
Schmoch U, Schubert T, Jansen D, Heidler R, von Görtz R (2010) How to use indicators to measure scientific performance: a balanced approach. Res Eval 19(1):2–18
Sousa CA, de Nijs WF, Hendriks PH (2010) Secrets of the beehive: performance management in university research organizations. Hum Relat 63(9):1439–1460
Staatsministerium B (2005) Zielvereinbarung. Irsee, pp 1–26
Stajkovic AD, Luthans F (2003) Behavioral management and task performance in organizations: conceptual background, meta-analysis, and test of alternative models. Pers Psychol 56 (1):155–194
Süß S (2007) Grenzen einer leistungsorientierten Anreizgestaltung für Hochschullehrer. Hochschulmanagement 3:68–72
Vilalta J (2001) University policy and coordination systems between governments and universities: the experience of the Catalan university system. Tert Educ Manag 7(1):9–22
Weibel A, Rost K, Osterloh M (2010) Pay for performance in the public sector—benefits and (hidden) costs. J Public Adm Res Theory 20(2):387–412
Wilkesmann U, Schmid CJ (2011) Lehren lohnt sich (nicht)? Ergebnisse einer deutschlandweiten Erhebung zu den Auswirkungen leistungsorientierter Steuerung auf die universitäre Lehrtätigkeit. Soziale Welt 62(3):251–278
Wilkesmann U, Schmid CJ (2012) The impacts of new governance on teaching at German universities. Findings from a national survey. Higher Educ 63(1):33–52
Wilkesmann U, Würmseer G (2009) Lässt sich Lehre an Hochschulen steuern? Auswirkungen von Governance-Strukturen auf die Hochschullehre. die hochschule 2:33–46
Wissenschaftsrat (2011) Empfehlungen zur Bewertung und Steuerung von Forschungsleistung. Halle
Wissenschaftsrat (2013) Empfehlungen zur Zukunft des Forschungsratings. Mainz
Wissenschaftsrat (2014) Institutionelle Perspektiven der empirischen Wissenschafts- und Hochschulforschung in Deutschland. Mainz

Wollersheim J, Lenz A, Welpe IM, Spörrle M (2014a) Me, myself, and my university: a multilevel analysis of individual and institutional determinants of academic performance. J Bus Econ
Wollersheim J, Welpe IM, Ringelhan S (2014b) Sein und Sollen - Bewertung von Forschungsleistungen in den Wirtschaftswissenschaften. Forschung Lehre 4:276–278

Part II
Performance Measurement

Interrater Reliability of the Peer Review Process in Management Journals

Alexander T. Nicolai, Stanislaw Schmal, and Charlotte L. Schuster

Abstract Peer review is an established method of assessing the quality and contribution of academic performance in most scientific disciplines. Up to now, little is known about interrater agreement among reviewers in management journals. This paper aims to provide an overview of agreement among the judgments of reviewers in management studies. The results of our literature review indicate a low level of agreement among reviewers in management journals. However, low consensus is not specific to management studies but widely present in other sciences as well. We discuss the consequences and implications of low judgment agreement for management research.

1 Introduction

In order to make a scientific contribution, research work has to be shared within the scientific community and come under discussion and scrutiny (Beyer et al. 1995). In management studies, as in any other academic discipline, scholarly journals serve as a platform for scientists to communicate their work to each other. Peer review is the predominant process through which manuscripts are evaluated prior to publication. It is a "key quality control mechanism" (Campanario 1998, p. 181; Rowland 2002) and has a "gatekeeping role" (Beyer et al. 1995, p. 1219). Publication in academic journals is regarded as an indicator of the "quality of role-performance in a social system" (Zuckerman and Merton 1971, p. 66). Review processes thus play a significant role in an author's career development, salary, and recognition within the scientific community (Frey 2003; Hunt and Blair 1987; Ketchen and Ireland 2010; Miller 2006). Since the rankings of departments and universities are also frequently based on publications in peer-refereed journals (Frey 2003, p. 211), the decisions of reviewers have a significant impact on academic systems in general.

A.T. Nicolai (✉) • S. Schmal • C.L. Schuster
Carl von Ossietzky Universität, Oldenburg, Germany
e-mail: alexander.nicolai@uni-oldenburg.de; stanislaw.schmal@uni-oldenburg.de; charlotte.schuster@uni-oldenburg.de

I.M. Welpe et al. (eds.), *Incentives and Performance*,
DOI 10.1007/978-3-319-09785-5_7

Given the importance of peer review, it is not surprising that this method "arouses very diverse emotions, beliefs, and ambitions. It angers, it reassures, it intimidates, it tramples egos, and it puffs them up" (Starbuck 2003, p. 348). At the same time, peer review is a heavily discussed topic, also in management studies (Miller 2006; Nicolai et al. 2011; Starbuck 2003). The most widely discussed aspects of peer review include validity, generalizability, accuracy, and bias (Campanario 1998; Marsh et al. 2008). In particular, the biases that affect peer review receive a lot of attention. Scholars from various disciplines, including the sociology of science, have named up to 25 different biases that can affect the fairness of peer review.[1] Campanario (1998), for instance, discusses evidence for bias towards positive and statistically significant results, as well as for bias against replication (Hubbard et al. (1998) show that bias against replication also applies to the area of Strategic Management). Gans and Shepherd (1994) discuss bias against fundamentally new ideas and refer to later Nobel Laureates whose manuscripts were often rejected in the first instance. Other scholars argue that some authors are favored over others who produce the same or even better quality which results from biases of reputation (Beyer et al. 1995; Merton 1968; Miller 2006). Other authors discuss a possible gender bias (Bornmann 2008).

Biased judgments and, accordingly, a lack of fairness are certainly among the most discussed issues in peer review. Still, the proponents of peer review argue that, although this method is imperfect, it is "more effective than any other known instrument for self-regulation in promoting the critical selection that is crucial to the evolution of scientific knowledge" (Bornmann 2011, p. 202).

One of the "most basic", "damning", and "broadly supported" criticism of peer review is "its failure to achieve acceptable levels of agreement among independent assessors", which makes peer review unreliable (Marsh et al. 2008, p. 161f). According to Mutz et al. (2012, p. 1), differing reviewer judgments of manuscripts are "[o]ne of the most important weaknesses of the peer review process." The reliability of peer review is typically studied by measuring "interrater reliability" (Bornmann 2008, 2011). Cicchetti (1991, p. 120) defines interrater or interreferee reliability as "the extent to which two or more independent reviews of the same scientific document agree."

This article aims to discuss dissensus among reviewer judgments regarding the acceptance, revision or rejection of a manuscript. We specifically provide an updated overview of studies on the degree of consensus among reviewers in the peer review process of papers for publication in management journals. We compare the empirical results of management studies with those of studies from other disciplines. Finally, consequences of high dissensus that is observed among

[1] See Bornmann (2008, p. 26) and Cicchetti (1991, p. 129) for a list of literature on peer review research discussing different biases. See also Campanario (1998) who discusses fraud, favoritism, self-interest, the connections among authors, reviewers, and editors, as well as the suggestibility of particularistic criteria in the context of double-blind reviewing.

reviewers and its implications for management studies and the academic system are discussed.

2 Interrater Reliability in Management Studies

There is a controversial debate on the most appropriate statistical measure for interrater reliability (see for an overview Conger and Ward 1984; or Whitehurst 1984). Cicchetti (1991, p.120) points out that an appropriate statistical method should account for the number of referees, the matching procedure, and the degree of reviewer agreement that can be attributed to chance alone. Considering the points mentioned by Cicchetti (1991), interclass correlation (ICC) and Cohen's kappa (k) are argued to be the most appropriate measures of interrater reliability (Bartko 1976; Cicchetti 1980; Spitzer and Fleiss 1974; Tinsley and Weiss 1975). Interclass correlation provides reliability estimates of assessments made by two or more referees of the same manuscript. Full agreement between reviewers is indicated by the value 1.0 (Shrout and Fleiss 1979; Whitehurst 1984). Cohen's kappa is a statistical method for identifying the degree of agreement between two or more raters that is above the agreement that could be expected by chance (Fleiss and Cohen 1973). It ranges from −1 to 1. Negative values indicate poorer agreement than would be expected by chance, 0 indicates chance agreement, and positive values are interpreted as chance-corrected percentage agreement (Landis and Koch 1977; Weller 2001).

Interrater reliability in management journals is a seldom analyzed issue. Our systematic research of the relevant literature identified five studies that analyze the level of agreement among reviewers in management studies. Table 1 presents the results of five studies including three management journals: *Academy of Management Journal* (*AMJ*), *Administration Science Quarterly* (*ASQ*), and a German journal, *Zeitschrift für Führung und Organisation* (*ZFO*). The table shows the methods each study applied, the results it obtained, and the qualitative interpretation of the authors.

The studies presented here cover the period between 1976 and 2005. Cummings et al. (1985) initiated the debate on disagreement among referees in the management discipline. They analyzed the statements of reviewers on manuscripts submitted to the *AMJ* between 1976 and 1978. In 34 of 81 cases the authors found that the reviewers' recommendations were inconsistent. This corresponds to a disagreement rate of almost 42 %, indicating that there was no common ground among the reviewers' evaluations.

The next study, which was published 10 years later by Beyer et al. (1995), picks up the issue of interrater agreement in management sciences. Using a data sample of 400 manuscripts submitted to the *AMJ* from 1984 to 1987, Beyer et al. analyzed the effects of author, manuscript and review process characteristics on the publication decisions in the first and final review. They calculated an indicator for reviewer disagreement as the standard deviation of the five-point scaled submission

Table 1 Interrater agreement in management studies

Journal/Authors	Agreement	Sample size	Categories	Author's interpretation
Academy of Management Journal (AMJ) (Cummings et al. 1985)	42 % disagreement[a]	81		None
Academy of Management Journal (AMJ) (Beyer et al. 1995)	$SD = 0.69$	400	5	None
Administrative Science Quarterly (ASQ) (Starbuck 2003)	$\rho = 0.12$ Pearson product–moment correlation	~500	3	"Little agreement among reviewers" (p. 348)
Academy of Management Journal (AMJ) (Miller 2006)	37 % disagreement	68	5	"Dissensus is present at AMJ but certainly not to the extent that it could be" (p. 429)
Zeitschrift Führung + Organisation (ZFO) (Nicolai et al. 2011)	ρ(full sample) $= 0.19$ ρ(rejected) $= 0.02$ ρ(accepted) $= -0.25$[b]	142	5	"Weakly positive relationship between the evaluations of academics and practitioners" (p. 60)

ρ = Pearson product–moment correlation; *SD*=standard deviation
[a]In 34 of 81 cases the authors found that the reviewers' recommendations were inconsistent
[b]"Full" indicates a sample consisting of rejected and accepted manuscripts; "Rejected" indicates a sample consisting only of rejected manuscripts; "accepted" indicates a sample consisting only of accepted manuscripts

ratings. Even though this was not their main focus, some of their results provide interesting insights into interrater agreement in the *AMJ*. For new submissions the authors report a disagreement of 0.69 (averaged intra-paper *SD*). One may argue that this is not an extremely high value for a five-point Likert scale (Miller 2006, p. 429). However, a closer look at the results indicates an almost equivalent averaged dispersion from the mean value by 0.62, which may lead to a difference of more than one level on the scale.

Starbuck (2003) evaluated the review process from the editorial point of view. Drawing on his own experience as an editor for the *ASQ* in the late 1960s, he calculated the correlation among three-point scaled peer review recommendations of around 500 manuscripts. It resulted in a significant but almost negligible coefficient of 0.12. This value indicates a very low level of agreement among reviewers.

Further results on dissensus are discussed by Miller (2006). He mostly agrees that there is low agreement among reviewers in sociological and psychological research as a whole, but critically questions whether this is the case in the *AMJ*. Using a randomly drawn sample of 68 cases from Rynes editorship, he calculated the disagreement rate and the standard deviation of recommendations. He comes up with a disagreement rate of 37 %, which is similar to the result reported by

Cummings et al. (1985). Likewise his results on within-paper standard deviation[2] correspond to Beyer et al. (1995). As mentioned earlier, the level of both results indicates the existence of considerable dissensus among *AMJ* reviewers.

Nicolai et al. (2011) examined disagreement among the reviewers of a German management journal. This study is a special case in that the authors analyze a so-called bridge journal, the *ZFO*, which uses double-blind reviews conducted by one academic and one practitioner. The study's sample consists of 142 manuscripts submitted to the *ZFO* between 1995 and 2005. All examined recommendations are based on a five-point Likert scale. Correlation analysis reveals a significant ($p < 0.05$) but relatively low correlation (0.19), which implies an almost negligible relationship among the reviewers. Further analyses indicate a substantial difference in agreement between accepted and rejected manuscripts. The correlation coefficient of reviews for rejected papers was insignificant different from zero. In contrast, the coefficient for accepted papers is significant ($p < 0.10$) but negative (−0.25), which suggests a weak inverse relationship between the recommendations of academics and of practitioners.

Another study, which analyzes authors' opinions instead of directly comparison of review judgments, is not presented in Table 1. Bedeian (2003) analyzed the experience of 173 authors whose manuscripts were accepted for publication in the *AMJ* or the *AMR* between 1999 and 2001. In particular, Bedeian asked the leading authors how satisfied they were with the consistency of the different referees' comments (Bedeian 2003, p. 334). More than half of the respondents—a total of 107 (62 %)—were satisfied with the uniformity of the reviewers' statements. Another 34 (20 %) were neither satisfied nor dissatisfied, and 31 (18 %) were unsatisfied with the degree of agreement among the reviewers' recommendations. Bedeian (2003, p. 333) concluded that the "authors expressed satisfaction with the consistency of the editor and referee comments among one another." However, this result should be interpreted with caution. As Miller (2006, p. 429) points out, the nature of Bedeian's sample is biased towards successful authors. Indeed, Bedeian (2003, p. 335) himself states that as "several authors noted, they would have responded to various survey items differently had their manuscripts been rejected."

The results of all studies presented here that directly analyzed interrater agreement mostly indicate low consensus among reviewers in management studies. Moreover, the works indicate that dissensus in management studies seems to be independent of editor, journal, or period covered by the data. However, the methods these studies use (disagreement rate, correlations etc.) hardly make it possible to qualify the level of agreement or make comparisons between different studies and different measures. Furthermore, frequency statistics and correlation measures do not consider chance agreement among raters and present an additive bias (Whitehurst 1984). For this reason, in our literature review we also gathered the qualitative statements that the authors of these works have made (see column 5 in Table 1). All in all, these statements demonstrate that the degree of interrater

[2] The author or Miller (2006, p. 429) do not report numerical results.

agreement was smaller than expected. Starbuck (2003, p. 349) concludes that "knowing what one reviewer had said about a manuscript would tell me almost nothing about what a second reviewer had said or would say."

Some authors explain dissensus among reviewers with the "low paradigm development" of an academic discipline (Beyer et al. 1995; Miller 2006; Starbuck 2003). "Low paradigm development" refers to a lack of common assumptions, terminology, theory, and methodology (e.g., Pfeffer 1993; Zammuto 1984). These authors follow Lodahl and Gordon (1972), who used Kuhn's concept (1962) of the scientific paradigm to classify the major sciences as "low-paradigm" or "high-paradigm" disciplines. On the basis of the state of technological development and the level of consensus within each field of research, the social sciences are classified as "low-paradigm" disciplines, whereas the natural sciences, such as physics or chemistry, are classified as "high-paradigm" disciplines (Lodahl and Gordon 1972).

According to this scheme, management studies can be classified as a "low–paradigm" field (Beyer et al. 1995, p. 1230; Pfeffer 1993). Beyer et al. (1995) argue that in the social sciences there are no universal scientific criteria on the basis of which generally accepted decisions could be made. Whitley (1984, p. 780) describes management studies as a "fragmented type of scientific field" with a high degree of technical and strategic task uncertainty. In contrast to mathematicians or economists, management researchers do not coordinate or control their work in keeping with a "global view" of their discipline (Whitley 1984, p. 799f). The level of agreement in management studies can be described as "multitude, vague, and ever changing. Researchers do not behave as if agreements exist about some beliefs and perceptions being correct" (Starbuck 2003, p. 348). Thus, "low-paradigm" development, as well as the immaturity of an academic discipline, can be a reason for the inconsistent (usage of) judgment criteria in the review process and dissensus among reviewers (Beyer et al. 1995).

3 Interrater Reliability in Other Sciences

If the "low-paradigm" character of management studies is the reason for dissensus, agreement should be higher in "high-paradigm" disciplines. On the basis of the classification metric of Lodahl and Gordon (1972) of "low-" and "high-paradigm" disciplines, we summarize in Table 2 the results on interrater reliability among reviewers in two disciplines classified as "high-paradigm": chemistry and physics.

We identified studies on referee agreement in journal submission and grant application processes. Zuckerman and Merton (1971) examined the referee system as a whole and analyzed the patterns of decision-making among editors and reviewers in *The Physical Review* (*PR*). Examining 172 manuscripts evaluated by two referees between 1948 and 1956, the authors calculated a rate on full

Table 2 Interrater overview in "high-paradigm" disciplines

Journal/Authors	Discipline	Agreement	Sample size	Categories	Author's interpretation
The Physical Review (PR) (Zuckerman and Merton 1971)	Physics	3[a]–67 %[b] dis-agreement rate	172	2–4	"Agreement was very high" (p. 67)
Physical Review Letters (Lazarus 1982)	Physics	85–90 % dis-agreement rate			None
Angewandte Chemie (Daniel 1993)	Chemistry	K = 0.2	856	4	"Reviewer agree-ment must be described as rather unsatisfying" (p. 23)
Angewandte Chemie (Bornmann and Daniel 2008)	Chemistry	K = 0.1–0.21	1,899	4	"Low level of agreement among referees' recom-mendations" (p. 7174)
Grants within Committee on Science and Public Policy (Cole et al. 1981)	Chemical dynamics	53 % (share of total variance)	50	12	"Substantial reviewer variance" (p. 884)
Grants within Committee on Science and Public Policy (Cole et al. 1981)	Solid-state physics	47 % (share of total variance)[c]	50	12	

K = Cohen's kappa
[a]Refers to full disagreement that one referee recommended acceptance and the other rejection
[b]Refers to minor differences between referees
[c]Percentage of total variance in reviewers' ratings accounted for by differences among reviewers of individual proposals (Cole et al. 1981, p. 884)

disagreement[3] of about 3 %. This finding indicates a very high level of agreement. However, other sets of results from this study draw a slightly different picture. As the authors state, two-thirds of these recommendation judgments reveal "minor differences in the character of proposed revisions" (Zuckerman and Merton 1971, p. 67, footnote 3). Unfortunately, no further information is given on the recommendation range or its variance. Thus, the evaluation of those results is hardly possible.

Lazarus (1982) examined the issue of interrater agreement in physics. He claims that the agreement rate among reviewers on the question of accepting or rejecting a

[3] Full disagreement implies that one referee recommended acceptance and the other rejection.

manuscript in *Physical Review Letters* is about 10–15 %. This corresponds to a disagreement rate of about 85–90 %, which is more than twice as high as the disagreement rate in management journals. Unfortunately, Lazarus (1982) does not provide further details on this analysis.

Interrater reliability in chemistry was examined in Daniel (1993) and in Bornmann and Daniel (2008). Both studies used the referees' recommendations of the *Angewandte Chemie*, which is a journal published by the German Chemical Society (Gesellschaft Deutscher Chemiker, GDCh, Frankfurt am Main). As a measure of agreement, both studies used Cohen's kappa (Fleiss and Cohen 1973). In the earlier study, Daniel (1993) examined a sample of 856 manuscripts covering the mid-1980s. His calculations revealed a Cohen's kappa of 0.20, implying that agreement among the reviewers' judgments of these manuscripts is 20 % higher than would be expected by chance. Following the guidelines of Landis and Koch (1977) on how such measurements should be interpreted, this corresponds to a low level of agreement among reviewers.

In the subsequent study, Bornmann and Daniel (2008) used a larger sample of 1,899 manuscripts refereed in the year 2000. They applied a more advanced statistical method, weighted Cohen's kappa, which takes into account the different level of agreement between two or more referees. Depending on the weighting parameter, their kappa coefficients range between 0.10 and 0.21. That is, reviewers agreed on 10–21 % more manuscripts than could have been expected by chance. Thus, the more recent study shows an even lower degree of agreement among reviewers in the journal *Angewandte Chemie*.

Further results on interrater agreement in high paradigm disciplines can be derived from the analysis of Cole et al. (1981). The authors analyzed interrater reliability in a grant application process. Their data sample consisted of 150 proposals in chemical dynamics, solid-state physics, and economics—50 from each discipline—and about 12 reviewers for each proposal. In order to avoid making statistical assumptions, the authors used the analysis-of-variance approach to determine the level of consensus among referees. The variance in the ratings is decomposed into variation in the quality of a proposal, in the review procedure, and in the reviewers' judgments. The authors' results show that most of the variance in the ratings of the proposals is due to reviewer disagreement and not to differences in content or methodology. In fact, in chemistry dynamics the variance in the reviewers' ratings accounted for 53 % of the total variance. In solid-state physics this value reached 47 %. In both cases, the variance among the reviewers of the same proposal is almost twice as high as the variation in the quality of proposals. The authors conclude by saying that "[c]ontrary to a widely held belief that science is characterized by wide agreement [. . .] our research both in this and other studies in the sociology of science indicates that concerning work currently in process there is substantial disagreement in all scientific fields" (Cole et al. 1981, p. 885). Other articles presenting overviews of the reliability of peer review or meta-analysis also did not find "any effect of [the] discipline" (Bornmann et al. 2010, p. 7) on interrater agreement among reviewers (Cicchetti 1991; Weller 2001).

The results presented in Table 2 indicate that also in "high-paradigm" disciplines interrater agreement is low. In view of that, the argument that dissensus among reviewers is a consequence of the "low-paradigm" nature of management studies seems fragile. On the contrary, dissensus among reviewers who assess works submitted to academic journals appears to be a common issue in science.

Conclusion

Interrater reliability is a topic that "goes to the heart of peer review" (Miller 2006, p. 426). The advancement of knowledge would be impossible if scientists were not able to reach a certain degree of consensus (Pfeffer 1993, p. 611). If a work of research is rigorous, it can be expected that two independent reviewers will agree on its quality (Bornmann 2011, p. 207). Our overview on interrater reliability in peer review conducted for management journals shows that this is often not the case: there seems to be little consensus among reviewers in management studies. Some authors attribute this tendency to the "low-paradigm" and fragmented nature of management research, as a result of which, inconsistent judgment criteria may be applied in the review process (Beyer et al. 1995, p. 1255). However, a low degree of interrater agreement is not specific to management studies. Also "high-paradigm" fields exhibit a high degree of reviewer disagreement. Thus, the hope that consensus might grow as the paradigm of management studies develops seems delusive. Dissensus could even increase if, as some authors (e.g., Cohen 2007) suggest, management journals integrate more and more science-external audiences into the peer review process (see Nicolai et al. 2011 for a critical discussion).

A high degree of dissensus illustrates the central role of journal editors for two reasons. First, the editor's opinion is given greater prominence if the reviewers' recommendations point in different directions. Second, the editor chooses the referees, and the result of a review might depend more on the selected reviewer than on the quality of the submitted manuscript. Kravitz et al. (2010, p. 4) found that, "recommendations were more consistent for multiple manuscripts assigned to the same reviewer (intra-class correlation coefficient rho = 0.23) than for multiple reviewers assessing the same manuscript (rho = 0.17)." Overall, low reliability implies a certain randomness of the peer review process. Consequently, publication in journals should not serve as the only measure of scientific performance. Instead, researchers should triangulate different kinds of evidence about scientific quality (Starbuck 2005, p. 197). An alternative form of measuring scholarly performance is to use its communications or hyperlinks on the World Wide Web. Webometrics is based on "link analysis, web citation analysis, search engine evaluation and purely descriptive studies of the web" (Thelwall 2008, p. 611). Examples of such measures are hyperlinks between pages (Aguillo

(continued)

et al. 2006) or numbers of external inlinks received by one's website (Tang et al. 2012). A further development on webometrics is provided by the so called altmetrics indicators. Those additionally take into account social bookmarking services and the number of downloads (see for an overview of altmetrics indicators Weller (2015)). Those measures are mainly unattached by reviewer judgments, but are available on a large scale and mostly immediately after publications. Moreover, these indices additionally assess the impact of readers and not only citers (Thelwall et al. 2013).

The statistical tests that the studies included in this overview applied to examine reviewer agreement are the subject of an ongoing methodological debate (Kravitz et al. 2010; Weller 2001). For example, percentages of agreement and Pearson product–moment correlation present numerous problems, while raw frequency counts do not distinguish agreement by chance alone and thus include both true and chance agreement (Watkins 1979). The correlation approach corrects for random agreement; however, as a measure of association rather than reliability, this statistic does not provide any information on the level of disagreement among reviewers. In fact, it can obtain perfect correlation even if the referees never agreed exactly but disagreed proportionally (Hendrick 1976; Whitehurst 1984). Kravitz et al. (2010, p. 4) criticize the methods that many studies apply to analyze the assessments of reviewers: they assume that judgments vary along one latent dimension of publishability and merit but that this "can hardly be tested by calculating kappa or intraclass correlation coefficients." In a similar vein, Hargens and Herting (1990) also criticized the latent-dimension approach. As Hargens and Herting (1990, p. 14) argue, by using the row–column association model of Goodman (1984), it is possible to "derive information about the distances between recommendation categories empirically rather than requiring arbitrary assumptions about those distances." The authors recommend that researchers should analyze these issues before calculating any disagreement measures.

It should be noted, however, that low interrater agreement is not necessarily a sign that the review process is not working well. Editors might deliberately choose reviewers with complementary expertise and opposing perspectives to obtain recommendations on different aspects of a piece of research (Hargens and Herting 1990, p. 2). It is open to debate whether the choice of reviewers with complementary competences can explain the low agreement rates observed among reviewers. So far, very few studies have analyzed comparatively the content of reviewers' comments to identify the reasons behind their disagreement (Bornmann 2011, p. 226). An analysis of this issue could shed light on an interesting discrepancy. Reviews, which usually are not publicly available, are characterized by very different points of view and often harsh criticism (Miller 2006). The published debate in

(continued)

management studies is much more consensual. Negational or critical citations in scholarly management articles are very rare (Schulz and Nicolai 2014). A better understanding of why exactly the opinions of reviewers differ could contribute not only to the improvement of the reviewing process, but also to the progress of the scholarly management debate in general.

References

Aguillo IF, Granadino B, Ortega JL, Prieto JA (2006) Scientific research activity and communication measured with cybermetrics indicators. J Am Soc Inf Sci Technol 57(10):1296–1302

Bartko JJ (1976) On various intraclass correlation reliability coefficients. Psychol Bull 83 (5):762–765

Bedeian AG (2003) The manuscript review process: the proper roles of authors, referees, and editors. J Manag Inq 12(4):331–338

Beyer JM, Roland GC, Fox WB (1995) The review process and the fates of manuscripts submitted to amj. Acad Manage J 38(5):1219–1260

Bornmann L (2008) Scientific peer review. An analysis of the peer review process from the perspective of sociology of science theories. Hum Archit 6(2):23–38

Bornmann L (2011) Scientific peer review. Ann Rev Inf Sci Technol 45:199–245

Bornmann L, Daniel H-D (2008) The effectiveness of the peer review process: inter-referee agreement and predictive validity of manuscript refereeing at angewandte chemie. Angew Chem Int Ed 47(38):7173–7178

Bornmann L, Mutz R, Daniel H-D (2010) A reliability-generalization study of journal peer reviews: a multilevel meta-analysis of inter-rater reliability and its determinants. PLoS ONE 5(12):e14331

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

Cicchetti DV (1980) Reliability of reviews for the American psychologist: a biostatistical assessment of the data. Am Psychol 35(3):300–303

Cicchetti DV (1991) The reliability of peer review for manuscript and grant submissions: a cross-disciplinary investigation. Behav Brain Sci 14(01):119–135

Cohen DJ (2007) The very separate worlds of academic and practitioner publications in human resource management: reasons for the divide and concrete solutions for bridging the gap. Acad Manag J 50:1013–1019

Cole S, Cole RJ, Simon AG (1981) Chance and consensus in peer review. Science 214 (4523):881–886

Conger AJ, Ward DG (1984) Agreement among 2 × 2 agreement indices. Educ Psychol Meas 44 (2):301–314

Cummings LL, Frost PJ, Vakil TF (1985) The manuscript review process: a view from inside on coaches, critics, and special cases. In: Cummings LL, Frost PJ (eds) Publishing in the organizational sciences. Irwin, Homeland, pp 469–508

Daniel H-D (1993) Guardians of science: fairness and reliability of peer review. VCH, Weinheim

Fleiss JL, Cohen J (1973) The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. Educ Psychol Meas 33(3):613–619

Frey BS (2003) Publishing as prostitution? - choosing between one's own ideas and academic success. Public Choice 116(1–2):205–223

Gans JS, Shepherd GB (1994) How are the mighty fallen: rejected classic articles by leading economists. J Econ Perspect 8(1):165–179

Goodman LA (1984) The analysis of cross-classified data having ordered categories. Harvard University Press, Cambridge

Hargens LL, Herting JR (1990) A new approach to referees' assessments of manuscripts. Soc Sci Res 19(1):1–16

Hendrick C (1976) Editorial comment. Pers Soc Psychol Bull 2:207–208

Hubbard R, Vetter DE, Littel EL (1998) Replication in strategic management: scientific testing for validity, generalizability, and usefulness. Strateg Manage J 19:243–254

Hunt JG, Blair JD (1987) Content, process, and the Matthew effect among management academics. J Air Waste Manage Assoc 13(2):191–210

Ketchen D, Ireland RD (2010) From the editors upon further review: a survey of the academy of management journal's editorial board. Acad Manage J 53(2):208–217

Kravitz RL, Franks P, Feldman MD, Gerrity M, Byrne C, Tierney WM (2010) Editorial peer reviewers' recommendations at a general medical journal: are they reliable and do editors care? PLoS ONE 5(4):1–5

Kuhn T (1962) The structure of scientific revolutions, vol 2. University of Chicago Press, Chicago

Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33(1):159–174

Lazarus D (1982) Interreferee agreement and acceptance rates in physics. Behav Brain Sci 5 (2):219–219

Lodahl JB, Gordon G (1972) The structure of scientific fields and the functioning of university graduate departments. Am Soc Rev 37(1):57–72

Marsh HW, Jayasinghe UW, Bond NW (2008) Improving the peer-review process for grant applications. Reliability, validity, bias, and generalizability. Am Psychol 63(3):160–168

Merton RK (1968) The Matthew effect in science. Science 159(3810):56–60

Miller CC (2006) From the editors: peer review in the organizational and management sciences: prevalence and effects of reviewer hostility, bias, and dissensus. Acad Manage J 49 (3):425–431

Mutz R, Bornmann L, Daniel H-D (2012) Heterogeneity of inter-rater reliabilities of grant peer reviews and its determinants: a general estimating equations approach. PLoS ONE 7(10):1–10

Nicolai AT, Schulz A-C, Göbel M (2011) Between sweet harmony and a clash of cultures: does a joint academic–practitioner review reconcile rigor and relevance? J Appl Behav Sci 47 (1):53–75

Pfeffer J (1993) Barriers to the advance of organizational science: paradigm development as a dependent variable. Acad Manage Rev 18(4):599–620

Rowland F (2002) The peer-review process. Learned Publish 15(4):247–258

Schulz A-C, Nicolai A (2014) The intellectual link between management research and popularization media: a bibliometric analysis of the harvard business review. Acad Manage Learn Educ. forthcoming

Shrout PE, Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. Psychol Bull 86(2):420–428

Spitzer RL, Fleiss JL (1974) A re-analysis of the reliability of psychiatric diagnosis. Br J Psychiatry 125(587):341–347

Starbuck WH (2003) Turning lemons into lemonade: where is the value in peer reviews? J Manage Inq 12(4):344–351

Starbuck WH (2005) How much better are the most-prestigious journals? The statistics of academic publication. Organ Sci 16(2):180–200

Tang M-C, Wang C-M, Chen K-H, Hsiang J (2012) Exploring alternative cyberbibliometrics for evaluation of scholarly performance in the social sciences and humanities in Taiwan. Proc Am Soc Inf Sci Technol 49(1):1–1

Thelwall M (2008) Bibliometrics to webometrics. J Inf Sci 34(4):605–621

Thelwall M, Haustein S, Larivière V, Sugimoto CR (2013) Do altmetrics work? Twitter and ten other social web services. PLoS ONE 8(5):e64841

Tinsley HE, Weiss DJ (1975) Interrater reliability and agreement of subjective judgments. J Couns Psychol 22(4):358–376

Watkins MW (1979) Chance and interrater agreement on manuscripts. Am Psychol 34(9):796–798

Weller AC (2001) Editorial peer review: its strengths and weaknesses, Asist monograph series. Hampton Press, New Jersey

Weller K (2015) Social media and altmetrics: an overview of current alternative approaches to measuring scholarly impact. In: Welpe IM, Wollersheim J, Ringelhan S, Osterloh M (eds) Incentives and performance - governance of research organizations. Springer International Publishing AG, Cham

Whitehurst GJ (1984) Interrater agreement for journal manuscript reviews. Am Psychol 39 (1):22–28

Whitley R (1984) The development of management studies as a fragmented adhocracy. Soc Sci Inf 23(4–5):775–818

Zammuto RF (1984) Coping with disciplinary fragmentation. J Manage Educ 9(30):30–37

Zuckerman H, Merton RK (1971) Patterns of evaluation in science: institutionalisation, structure and functions of the referee system. Minerva 9(1):66–100

The Use of Bibliometrics for Assessing Research: Possibilities, Limitations and Adverse Effects

Stefanie Haustein and Vincent Larivière

Abstract Researchers are used to being evaluated: publications, hiring, tenure and funding decisions are all based on the evaluation of research. Traditionally, this evaluation relied on judgement of peers but, in the light of limited resources and increased bureaucratization of science, peer review is getting more and more replaced or complemented with bibliometric methods. Central to the introduction of bibliometrics in research evaluation was the creation of the Science Citation Index (SCI) in the 1960s, a citation database initially developed for the retrieval of scientific information. Embedded in this database was the Impact Factor, first used as a tool for the selection of journals to cover in the SCI, which then became a synonym for journal quality and academic prestige. Over the last 10 years, this indicator became powerful enough to influence researchers' publication patterns in so far as it became one of the most important criteria to select a publication venue. Regardless of its many flaws as a journal metric and its inadequacy as a predictor of citations on the paper level, it became the go-to indicator of research quality and was used and misused by authors, editors, publishers and research policy makers alike. The h-index, introduced as an indicator of both output and impact combined in one simple number, has experienced a similar fate, mainly due to simplicity and availability. Despite their massive use, these measures are too simple to capture the complexity and multiple dimensions of research output and impact. This chapter provides an overview of bibliometric methods, from the development of citation indexing as a tool for information retrieval to its application in research evaluation, and discusses their misuse and effects on researchers' scholarly communication behavior.

S. Haustein (✉)
École de bibliothéconomie et des sciences de l'information (EBSI), Université de Montréal, Montréal, QC, Canada
e-mail: stefanie.haustein@umontreal.ca

V. Larivière
École de bibliothéconomie et des sciences de l'information (EBSI), Université de Montréal, Observatoire des sciences et des technologies (OST), Centre interuniversitaire de recherche sur la science et la technologie (CIRST), Université du Québec à Montréal, Montréal, QC, Canada
e-mail: vincent.lariviere@umontreal.ca

I.M. Welpe et al. (eds.), *Incentives and Performance*,
DOI 10.1007/978-3-319-09785-5_8

1 Introduction

The evaluation of researchers' work and careers, which traditionally relied on peer review, is increasingly substituted or influenced by publication output and citation impact metrics (Seglen 1997b; Rogers 2002; Cameron 2005). Bibliometric indicators are more and more applied by governments and other funding organization mainly because of their large-scale applicability, lower costs and time as well as their perceived objectivity. The goal is to optimize research allocations and make funding both more efficient and effective (Moed 2005; Weingart 2005). Bibliometrics and citation analysis go back to the beginning of the twentieth century, when they were used by librarians as a tool for journal collection development. Although mentioned much earlier by Otlet (1934) in French (*bibliométrie*), it is Pritchard (1969) who is mostly associated with coining the term *bibliometrics* as a method

> to shed light on the processes of written communication and of the nature and course of development of a discipline (in so far as this is displayed through written communication), by means of counting and analyzing the various facets of written communication. (Pritchard 1969, pp. 348–349)

However, it is only with the creation of the Science Citation Index in the early 1960s that bibliometrics became a method that could be massively applied to analyze patterns of scholarly communication and evaluate research output. Over the last 20 years, the increasing importance of bibliometrics for research evaluation and planning led to an oversimplification of what scientific output and impact were which, in turn, lead to adverse effects such as *salami publishing*, *honorary authorships*, *citation cartels* and other unethical behavior to increase one's publication and citation scores, without actually increasing one's contribution to the advancement of science (Moed 2005).

The goal of this chapter is to inform the reader about bibliometrics in research assessment and explain possiblilities and limitations. The chapter starts with a brief historic summary of the field and the developments of its methods, which provides the context in which the measurement of scholarly communication developed. An overview of indicators and their limitations is then provided, followed by their adverse effects and influence on researchers' scholarly communication behavior. The chapter concludes by summarizing the possibilities and limitations of bibliometric methods in research evaluation.

2 Development of Bibliometrics and Citation Analysis

Bibliometric analyses are based on two major units: the scientific publication as an indicator of research output and citations received by them as a proxy of their scientific impact or influence on the scholarly community. Early bibliometric studies, at that time referred to as *statistical bibliography*, were mostly applied to investigate scientific progress and later library collection management. For example, Cole and Eales (1917), which can be considered the first bibliometric study, examined the scientific output of European countries in anatomy research based on literature published between 1543 and 1860. By defining publications as the main unit of measurement to assess scientific activity in certain research areas, they laid out the basis for future bibliometric studies (De Bellis 2009). Ten years later, Gross and Gross (1927) were the first ones to carry out a citation analysis of journals, in search for an objective method for the management of their library collection. Pressured by limited budgets and physical space in libraries opposed to the ever-growing volume of scholarly documents published, they extracted citations to journals from 3,663 references listed in the 1926 volume of the Journal of the American Chemical Society, thus compiling a list of most cited journals to subscribe to. In doing so they equated citations received by journals with their importance in the disciple, setting the stage for citation analysis in both library collection management and research evaluation. Bradford (1934) further influenced librarians and collection management through his famous law of scattering, stating that the majority of documents on a given subject are published in a small number of core journals. Together with Lotka's (1926) law on the skewed distributions of papers per author, Zipf's (1949) law on word frequencies in texts, as well as Price's (1963) work on the exponential growth of science, Bradford formed the basis of the mathematical foundations of the field of bibliometrics. It did, however, take the development of a global and interdisciplinary citation index, i.e., Garfield's Science Citation Index (SCI), for bibliometric methods—and citation analysis as its key aspect—to enter into the area of research policy and evaluation. Citation indexes were the key to evaluative bibliometrics and research evaluation because they provided the database and made global and large-scale analyses feasible. The development of the Institute of Scientific Information (ISI) and the SCI gave rise to both the practical application of bibliometrics in research evaluation and information retrieval and theoretical and empirical research of citation analysis and bibliometric indicators.

2.1 The Science Citation Index

After World War II it was believed that economic growth and scientific progress were intertwined and the latter could be controlled and steered towards specific goals, resulting in the era of big science and hyperspecialization (De Bellis 2009). It is in this context that citation indexing developed as a means to cope with the flood of scientific literature. With the growth of publication output, the scientific landscape had become complex and the amount of literature unmanageable. Garfield's citation indexes aimed at making the information overload manageable creating a "World Brain" (Garfield 1964) of scientific information through automatic indexing based on references. Garfield adopted this method from Shepard's Citation, a citatory service in the field of law established in 1873 to keep track of the application of legal decisions (Garfield 1979). Citations were assumed to be the better descriptors and indexing terms as symbols of a document's content than natural language such as terms derived from document titles (Garfield 1964). Garfield believed that the community of citing authors outperformed indexers in highlighting cognitive links between papers especially on the level of particular ideas and concepts (Garfield 1983, p. 9), an approach resembling the pheonomenon known today as *crowdsourcing*. This is the theoretical foundation that citation indexing is based on.

As a multidisciplinary citation index, the SCI was initially developed for information retrieval and not for research evaluation. Along these lines, the Impact Factor was initially developed by Garfield as a tool to select the most relevant journals for coverage in ISI's SCI, with a particular focus on cost efficiency. As the mean number of citations received in 1 year by papers published in a journal during the 2 previous years, the Impact Factor selects the most cited journals regardless of output size (Garfield 1972).

Garfield's law of concentration, a further development of Bradfield's law of scattering combining all fields of science, showed that the majority of cited references referred to as few as 500 to 1,000 journals, justifying a cost-efficient coverage approach (Garfield 1979). Based on the Impact Factor, 2,200 journals had been identified by 1969 as "the world's most important scientific and technical journals" (Garfield 1972, p. 471) and became fully indexed by the SCI. The ISI citation indexes fostered further developments of the field of bibliometrics in general and of citation analysis in particular, both empirically and theoretically. By enabling large-scale publication and citation analysis of different entities from micro (author) to macro (country) level, the SCI provided the basis for quantitative research evaluation. Garfield himself underlined the potential of citation analysis in research evaluation and outlined the usefulness of the Impact Factor for librarians, editors and individual scientists (Garfield 1972; Moed 2005).

2.2 *Theory of Publication and Citation Analysis*

Original research results are typically formally communicated through publications. Thus, publications can be regarded as proxies of scientific progress at the research front (Moed 2005). They do, however, not capture the entire spectrum of scientific activity. In most of the medical and natural sciences, the journal article is the main publication format used by researchers to disseminate and communicate their findings to the research community, claim priority of findings and make them permanent. Peer-review and editorial work ensure a certain level of quality control, and details on the methods provide the means for colleagues to replicate the findings. Given this central importance of articles for scholarly communication, sociological research has considered that, by counting papers, we obtain an indicator of research activity.

Citation analysis is based on the assumption that a document referenced in a subsequent paper marks the intellectual influence of the cited document on the citing paper. Since the number of cited items is usually restricted by the length of the publication, a reference lists should not be considered a complete list but a selection of the most influential sources related to a piece of work (Small 1987). The number of citations received is thus assumed to reflect the influence or scientific impact of scholarly documents and mark their contribution to progress and advancement of science. This assumption is based on the argument by sociologist Robert K. Merton (1977, pp. 54–55) that "if one's work is not being noticed and used by others in the system of science, doubts of its value are apt to rise." Merton's normative approach regards citations as the "pellets of peer recognition" (Merton 1988, p. 620) within the scientific reward system, a symbol of acknowledging the knowledge claim of the cited source. The work by Merton's students Stephen and Jonathan Cole (Cole and Cole 1973) and Harriet Zuckerman (1987), which analyzed Merton's theories from an empirical perspective, have shown positive but not perfect correlations between citation rates and qualitative judgment by peers, thus providing an early framework for the use of bibliometrics in assessing, albeit imperfect, scientific influence (Moed 2005).

2.3 Bibliometrics and Peer Review

Peer review is the most important instrument when it comes to judging or ensuring the quality of scientists or of their work. It is applied at each level of a researchers' career, from submitted manuscripts to the evaluation of grant proposals, as well as to their suitability for academic positions or scholarly awards. Based on Merton's set of norms and values conveyed by the *ethos of science*, in a perfect setting peer review should be based entirely on the scientific quality and disregard any personal interests (Merton 1973). In reality, judgements could be influenced by prejudices and conflicts of interest of the referee, are sometimes inconsistent or often contradict each other (Martin and Irvine 1983). In addition, peer review is time and cost intensive. This has generated, for some, the need for faster and more cost-efficient methods. Studies (e.g., Cole and Cole 1967; Martin and Irvine 1983; Rinia et al. 1998; Norris and Oppenheim 2003) correlating peer judgement with citations found positive but not perfect correlations, indicating that the two approaches reflect similar but not identical assessments. Given the limitations of both methods, none of them leads to a perfect, unbiased quality judgement. In the evaluation of research output, peer review and bibliometrics do thus not replace each other but are best used in combination.

3 Bibliometric Analyses

3.1 Basic Units and Levels of Aggregation

Aggregation levels of bibliometric studies range from micro (author) to macro (countries) with different kinds of meso levels in between such as institutions, journals or research fields and subfields. Regardless of the level of aggregation, publication activity of particular entities is determined through the author and author addresses listed in the article (De Lange and Glänzel 1997). A paper published by author A at Harvard University and author B at the University of Oxford in the Astrophysical Journal would thus count as a publication in astrophysics for authors A and B on the micro, Harvard and Oxford on the meso and the US and the UK on the macro level. Since co-publications serve as a proxy of scientific collaboration, the same publication would provide formal evidence of authors A and B, Harvard and Oxford and the US and UK collaborating in astrophysics research. Publications can be counted fully, i.e., each participating unit is credited with one publication, or fractionally, assigning an equal fraction of the paper to each entity (Price 1981), that is 0.5 to each of the two entities per aggregation level in the example above. The latter is particularly helpful to compare scientific productivity of research fields with different authorship patterns. While single-authored papers are quite common in the humanities, the list of authors in experimental physics can include more than a hundred names, because research is carried out at large international facilities (Larivière et al. 2006). The number of authors per article and international collaboration is increasing over time for all fields of research (Abt 1992; De Lange and Glänzel 1997). Hence, the probability that a researcher contributes to a paper is, thus, very different from one field to another.

Similarly, different fields have different rules when it comes to authorship (Pontille 2004; Birnholtz 2006). For example, the act of writing the paper is central to authorship in the social sciences and humanities, while, in the natural and medical sciences, the data analysis plays an important role. As a consequence, a research assistant who analyzed the data in sociology would not typically be an author of the paper—and thus be included in the bibliometrics measures—while in natural and medical sciences, this task could lead to authorship. This is exemplified by the very large number of authors found in high-energy physics, where all members of the research team—which can amount to several hundreds—will sign the paper, even several months after they left the experiment (Biagioli 2003). Thus, the adequation between 'research activity' and bibliometrics depends on the discipline.

3.2 Data Quality

Another aspect that is essential in bibliometric studies is the quality of data. This involves the selection of a suitable database and cleaning of bibliographic metadata. Author and institution names come in many different forms including first names and intitials, abbreviations and department names or spelling errors and they may change over time (*synonymy problem*). On the other hand, the same name might refer to more than one person or department (*homonymy problem*). Disambiguation and cleaning author names and institutions is fundamental to computing meaningful bibliometric indicators used in research evaluation.[1] It is most successful if either publication lists are verified by authors or on a larger scale if cleaning is carried out by experts supported by powerful rule-based algorithms computing the probabilities based on similarities of document metadata including author addresses, co-authors, research fields and cited references (Moed 2005; Smalheiser and Torvik 2009). It should be assumed that the individual author would be able to identify his or her own publications. As the authors themselves should know best which papers they have written, a registry with unique and persistent author identifiers which researchers can link with their publications, would solve author homonymy and synonymy problems. ResearcherID[2] represents such a registry within Web of Science and ORCID[3] has recently been launched as a platform-independent and non-profit approach to identify and manage records of research activities (including publications, datasets, patents etc.). For such a system to fully replace author name disambiguation for evaluative bibliometrics, it would have to be based on the participation of the entire population of authors during the period under analysis including those that are no longer active in the field or even alive. As this is not very likely such registries can be used to support the disambiguation process but cannot entirely replace data cleaning in bibliometric analyses.

3.3 Bibliometric Indicators

Provided a cleaned dataset, entities such as authors, institutions or countries can be compared regarding their publication activity and citation impact using bibliometric indicators. Among those frequently used in research evaluation, one can distinguish between *basic* and *normalized* metrics, *time-based* and *weighted indicators* are other kinds of citation indicators. Since they are more common in journal evaluation—e.g., cited half-life (Burton and Kebler 1960) or Eigenfactor

[1] It should be mentioned that, despite the fact that bibliometrics should not be used alone for research evaluation, individual level disambiguation is often needed in order to assess groups of researchers.

[2] http://www.researcherid.com

[3] http://orcid.org

metrics (Bergstrom 2007) and SCImago journal rank (Gonzalez-Pereira et al. 2010)—than in research assessment of authors, institutions and countries, they are not included in the following overview for reasons of space.

3.3.1 Basic Indicators

Basic or simple bibliometric indicators include the number of publications and citations, which are size-dependent measures. Mean citation rates try to account for output size by dividing the total number of citations received by an entity by the number of its publications. Basic citation rates can be limited to certain document types, include or exclude self-citations or have different citation and publication windows. That is, they can be calculated synchronously – citations received in 1 year for documents published during previous years – or diachronously – documents published in 1 year cited in subsequent years (Todorov and Glänzel 1988). Without fixed publication and citation windows or accurate normalization for publication age, an older document has higher probabilities of being cited.

Basic publication and citation indicators are influenced by different publication patterns of disciplines and also by the size or age of the measured entity. Using basic instead of normalized (see Sect. 3.3.2) metrics, a researcher from the medical sciences would thus seem more productive and to have higher citation impact than a mathematician, because medical scientists contribute to more papers and their papers contain a larger number of and more recent references than those in mathematics. Comparing the publication output and citation impact of authors, institutions, journals and countries without an accurate normalization is thus like comparing apples with oranges. A university with a large medical department would always seem more productive and impactful than those without.

Mean citation rates are the most commonly used size-independent indicator of scientific impact. Due to the highly skewed distribution of citations per paper—as a rule of thumb, 80 % of citations are received by 20 % of documents and many are never cited, especially in the humanities (Larivière et al. 2008)—the arithmetic mean is, however, not a very suitable indicator since other than in a normal distribution it is not representative of the majority of documents (Seglen 1992). The median has been suggested as more appropriate due to its robustness (Calver and Bradley 2009), but since it disregards the most frequently cited document, it cannot fully represent the citation impact of a set of papers. Providing the standard deviation with the mean and additional distribution-based indicators such as citation percentiles, for example, the share of top 1 or 5 % highly cited papers as a measure of excellence, seem more appropriate (Tijssen and van Leeuwen 2002; Bornmann and Mutz 2011).

Journal Impact Factor As noted above, the Impact Factor was developed out of the need to select the most relevant journals to include in the SCI regardless of output size:

> In view of the relation between size and citation frequency, it would seem desirable to discount the effect of size when using citation data to assess a journal's importance. We have attempted to do this by calculating a relative impact factor—that is, by dividing the number of times a journal has been cited by the number of articles it has published during some specific period of time. The journal impact factor will thus reflect an average citation rate per published article. (Garfield 1972, p. 476)

The journal Impact Factor is a certain type of mean citation rate, namely a synchronous one based on citations received in year *y* by papers published in the 2 previous years, i.e., *y-1* and *y-2*. As such the above-mentioned limitations of arithmetic means to represent non-normal distributions apply (Todorov and Glänzel 1988; Seglen 1992). An extreme example of the susceptibility of the Impact Factor to single highly-cited papers is that of Acta Crystallographica A, which increased 24-fold from 2.051 in 2008 to 49.926 in 2009 because of a software review of a popular program to analyze crystalline structure cited 5,868 times in 2009 (Haustein 2012).

In addition to being a mean citation rate, the Impact Factor has other limitations and shortcomings. It includes articles, reviews and notes as publication types while citations to all document types are considered, leading to an asymmetry between numerator and denominator (Moed and van Leeuwen 1995; Archambault and Larivière 2009). This asymmetry has led journal editors to "optimize" their journals' publication behavior (see Sect. 4). Another shortcoming of the journal Impact Factor is its short citation windows, which goes back to convenience and cost-efficiency decisions made in the early days of the SCI (Martyn and Gilchrist 1968; Garfield 1972). Garfield (1972) found that the majority of citations are received within the first 2 years after publication. For some disciplines 2 years are not long enough to attract a significant number of citations, thus leading to large distortions (Moed 2005). Since its 2007 edition, the Journal Citation Report (JCR) includes a 5-year Impact Factor but the 2-year version remains the standard. The asymmetry between numerator and denominator, which was caused by computational limitations in the 1960s and could easily be solved by document-based citation matching, however, still exists.

H-index Introduced by a researcher outside the bibliometric community physicist Jorge E. Hirsch, the h-index has had enormous impact on the scholarly and bibliometric community (Waltman and van Eck 2012) due to its attempt to reflect an author's publication output and citation impact with one simple integer. Although initially perceived as its strength, the oversimplification of the two orthogonal dimensions of publications and citations (Leydesdorff 2009), is actually its greatest weakness. Hirsch (2005) defined the h-index of an author as follows:

> A scientist has index h if h of his or her N_p papers have at least h citations each and the other $(N_p\ h)$ papers have h citations each. (Hirsch 2005, p. 16569)

In other words, for any set of papers ranked by the number of citations, *h* indicates the number of papers for which the number of citations is equal to or higher than the corresponding ranking position, i.e., an author has an h-index of 10, if 10 of his papers were each cited at least 10 times. The set of documents from the first to the h^{th} position are part of the so-called h-core. The indicator does not take into account the total number of publications or citations, so that two researchers with the same h-index could differ completely in terms of productivity and citation impact as long as they both published h papers with h citations.

Besides the fact that the metric tries to oversimplify a researcher's impact and is size-dependent, the h-index is also inconsistent. For example, if two authors gain the same number of citations even for the same co-authored document, their h-indexes increase by 1 only if the additional citations move the paper up in the ranking from a position outside to inside the h-core. Thus an identical citation increase even for the same paper can lead to different outcomes for two researchers (Waltman and van Eck 2012). Similarly, given that the maximum of the h-index is the entity's number of publication, the h-index is more strongly determined by the number of publications rather than the number of citations. Thus, the h-index cannot be considered as a valid indicator of research productivity and impact.

3.3.2 Normalized Indicators

As mentioned above, certain biases occur caused by differences in publication behavior between research fields, publication growth and speed, different document types, time frames and/or database coverage. To allow for a fair comparison of universities or researchers active in different subject areas, normalized citation indicators try to counterbalance these biases. A full normalization of biases is a difficult and so far not yet entirely solved task due to the complexity of processes involved in scholarly communication.

The most commonly used field-normalized indicators are based on the so-called *a posteriori*, *ex post facto* or *cited-side* normalization method, where normalization is applied after computing the actual citation score (Glänzel et al. 2011; Zitt 2010). The actual or observed citation value of a paper is compared with the expected discipline-specific world average based on all papers published in the same field in the same year and in some cases the same document type. Each paper thus obtains an observed vs. expected ratio, which above 1 indicates a relative citation impact above world average and below 1 the opposite (Schubert and Braun 1996). An average relative citation rate for a set of papers—for example, all papers by an author, university or country—is computed by calculating the mean of the relative citation rates, i.e., observed vs. expected citation ratios, of all papers (Gingras and Larivière 2011; Larivière and Gingras 2011). As observed citation impact is compared to field averages the cited-side normalization relies on a pre-established classification system to define a benchmark. A paper's relative impact thus depends on and varies with different definitions of research fields. Usually classification systems are journal based thus causing problems particularly for inter- and multidisciplinary journals. An alternative to the cited-side method is the *citing-side* or *a priori* normalization, which is independent of a pre-established classification system because the citation potential is defined through citing behavior, i.e., the number of references (Leydesdorff and Opthof 2010; Moed 2010; Zitt 2010; Waltman and van Eck 2010).

Although normalized indicators are the best way to compare citation impact of different entities in a fair way, the complex structures of scholarly communication are difficult to capture in one indicator of citation impact. It is thus preferable to triangulate methods and use normalized mean citation rates in combination with distribution-based metrics to provide a more complete picture.

4 Misuse and Adverse Effects of Bibliometrics

Adverse effects, misapplication and misuse of bibliometric indicators can be observed on the individual as well as the collective level. Researchers and journal editors look for ways to optimize or manipulate the outcomes of indicators targeted at assessing their success, resulting in changes of publication and citation behavior, while universities and countries reward publishing in high-impact journals. The more bibliometric indicators are used to evaluate research outputs and as a basis for funding and hiring decisions, the more they foster unethical behavior. The higher the pressure, the more academics are tempted to take shortcuts to inflate their publication and citation records. Misapplication and misuse of indicators such as the cumulative Impact Factor are often based on the uninformed use of bibliometric methods and data sources and develop in an environment where any number beats no number. The most common adverse effects and misuses of bibliometrics are described in the following.

Publishing in Journals That Count The importance of the Web of Science and the journal Impact Factor have led researchers to submit their papers to journals which are covered by the database and preferably to those with the highest Impact Factors, sometimes regardless of the audiences (Rowlands and Nicholas 2005). For example, results from the Research Assessment Exercise in the UK show that the share of publications in journals in the social sciences increased from 49.0 % in 1996 to 75.5 % in 2008.[4] At the same time, more and more publications are published in English instead of national languages (Engels et al. 2012; Hicks 2013). Although this can be seen as a positive outcome, it has adverse effects on the research system as it can lead to a change in scholars' research topics, especially in the social sciences and humanities. More specifically, given that journals with higher Impact Factors are typically Anglo-American journals that focus on Anglo-American research topics, scholars typically have to work on more international or Anglo-American topics in order for their research to be published in such journals and, as a consequence, perform much less research on topics of local relevance. For instance, at the Canadian level, the percentage of papers authored by Canadian authors that have "Canada" in the abstract drops from 19 % in Canadian journals to

[4] http://stadium.open.ac.uk/webcast-ou/documents/20130410_Jonathan_Adams_Presentation.ppt

6 % in American journals (compilation by the authors based on the Web of Science).

Salami Publishing and Self-Plagiarism Increasing the number of publications by distributing findings across several documents is known as *salami slicing, duplicate publishing* or *self-plagiarism*. This practice, where one paper is sliced into many small pieces to increase the number of countable output resulting in the smallest publishable unit, or when results from previously published papers are republished without proper acknowledgement, is regarded as unethical as it distorts scientific progress and wastes the time and resources of the scientific community (Martin 2013). The extent of such practices has been found to range between 1 and 28 % of papers depending on the level of the plagiarism (Larivière and Gingras 2010), which can span from the reuse of some data or figures to the duplication of the entire document. With the increase of the use of quantitative methods to assess researchers, we can expect such duplication to become more important and, thus, to artificially inflate the output of researchers.

Honorary Authorship and ghost authorship, i.e., listing individuals as authors who do not meet authorship criteria or not naming those who do, are forms of scientific misconduct which undermine the accountability of authorship and authorship as an indicator of scientific productivity. Flanagin et al. (1998) reported that honorary authors, also called guest or gift authors, appeared in 19.3 % of a sample of 809 medical journal articles with US corresponding authors published in 1996, while 11.5 % had ghost authors. A more recent study based on 2008 papers showed that the share of papers involving ghost authorship (7.9 %) had significantly decreased, while that with honorary authors (17.6 %) remained similar (Wislar et al. 2011). Honorary authorship represent one of the most unethical ways to increase publication output, since researchers are added to the author list who have not contributed substantially to the paper. In some extreme cases, authorship was even for sale. Hvistendahl (2013) reports about an academic black market in China, where authorship on papers accepted for publication in Impact Factor journals were offered for as much as US$ 14,800. Some journals try to prevent this unethical practice by publishing statements of author contributions. However, for the sample of papers in Wislar et al. (2011) publishing author contributions did not have a significant effect on misappropriate authorships. Similarly, within the context of international university rankings, universities in Saudi Arabia have been offering "part-time" contracts of more than US$ 70,000 a year to highly-cited researchers, whose task was simply to update their highly-cited profile on Thomson Reuters' with the additional affiliation, as well as sign the institution's name on their papers, with the sole purpose of increasing the institution ranking in the various university rankings (Bhattacharjee 2011). In a manner similar to salami publishing, these practices inflate scholars' research output and distort the adequacy of the indicator regarding the concept, i.e., the research activity of authors and institutions which it aims to measure.

Self-citations To a certain extent, author self-citations are natural, as researchers usually build on their own previous research. However, in the context of research evaluation, where citations are used as a proxy for impact on the scientific community, self-citations are problematic as they do in fact not mirror influence on the work of other researchers and thus distort citation rates (Aksnes 2003; Glänzel et al. 2006). They are also the most common and easiest way to artificially inflate one's scientific impact. Studies found that author self-citations can account for about one quarter to one third of the total number of citations received within the first 3 years, depending on the field, but generally decrease over time (e.g., Aksnes 2003; Glänzel et al. 2006; Costas et al. 2010). The common definition of author self-citations considers mutual (co-)authors of citing and the cited papers, i.e., a self-citation occurs if the two sets of authors are not disjoint (Snyder and Bonzi 1998). Self-citations can be removed from bibliometric studies to prevent distortions particularly on the micro and meso level. The correlations between citations and co-citations on a larger aggregation level were shown to be strong so that it is not necessary to control for self-citations on the country level (Glänzel and Thjis 2004). Another way to artificially increase one's citation counts, which cannot be easily detected, is citation cartels, where authors agree to cite each other's papers.

Increasing the Journal Impact Factor Due to its importance, the Impact Factor is probably the most misused and manipulated indicator. There are several ways how journal editors "optimize" the Impact Factor of their periodicals, a phenomenon referred to as the 'numbers game' (Rogers 2002), 'Impact Factor game' (The PLoS Medicine Editors 2006) or even 'Impact Factor wars' (Favaloro 2008). One method is to increase the number of citations to papers published in the journal in the last 2 years, i.e., journal self-citations, by pushing authors during the peer-review process to enlarge their reference lists (Seglen 1997a; Hemmingsson et al. 2002). Thomson Reuters monitors journal self-citations and suspends periodicals suspected of gaming. In the 2012 edition of the JCR, 65 titles were red-flagged[5] and thus not given an Impact Factor. Editors of four Brazilian journals went even a step further and formed a citation cartel to inflate each other's Impact Factors through *citation stacking*, which is not as easily detectable as journal self-citations (van Norden 2013). Another approach to manipulate the indicator is to foster the publication of non-citable items, which collect so-called 'free citations' to the journal by adding to the numerator but not the denominator (Moed and van Leeuwen 1995; Seglen 1997a).

Cumulative or Personal Impact Factors Aside from the Impact Factor being a flawed journal indicator, its worst application is that of cumulative or personal Impact Factors. Developed out of the need to obtain impact indicators for recent papers, which have not yet had time to accumulate citations, the journal Impact Factor is used as a proxy for the citations of papers published in the particular journal. The problem with using the journal Impact Factors as an expected citation rate is that due to the underlying skewed distributions, it is neither a predictor nor good representative of actual document citations (Seglen 1997a; Moed 2002). Recent research also provided evidence that this predictive power is actually decreasing since the 1990s (Lozano et al. 2012). Although the citations of papers

[5] http://admin-apps.webofknowledge.com/JCR/static_html/notices/notices.htm

determine the journal citation rate, the opposite does not apply. To go back to the Acta Crystallographica A example mentioned in Sect. 3.3.1, of the 122 articles and reviews published in 2007 and 2008, only the one highly cited document (49 citations) obtained as many citations than the Impact Factor value of 2009 (49.926), all other documents were cited less than 17 times, and 47 % were not cited at all during that year. Even though this shows that the Impact Factor is not at all a good predictor of citation impact, the cumulative Impact Factor, i.e., adding up Impact Factors for the papers published by a researcher, is frequently applied, most often in the biomedical fields, where grant committees ask cumulative Impact Factors of applicants and researchers list them in their CVs. Despite these deficiencies, the Impact Factor is still applied as a "cheap-and-cheerful" (Adam 2002, p. 727) surrogate for actual citations because it is available much faster. Although proven to be meaningless (Seglen 1997b), financial bonuses are awarded and hiring and tenure decisions based on the cumulative Impact Factors (Adam 2002; Rogers 2002; The PLoS Medicine editors 2006; Cameron 2005). It is hoped that the recent San Francisco Declaration on Research Assessment (DORA) can put an end to this malpractice.

Conclusions

This chapter has reviewed the framework, methods and indicators used in bibliometrics, focusing on its application in research evaluation, as well some of its adverse effects on researchers' scholarly communication behavior. It has argued that such indicators should be interpreted with caution, as they do not represent research activity—let alone scientific impact—but, rather, are indicators of such concepts. Also, they are far from representing the whole spectrum of research and scientific activities, as research does not necessarily lead to publication. Along these lines, bibliometric indicators do not provide any insights on the social or economic impact of research and are, thus, limited to assessing the impact of research within the scientific community. Hence, these indicators have to be triangulated and applied carefully, adapted to the units that are assessed. For example, while bibliometric data could be quite meaningful for assessing the research activity of a group of physicists in the United States, it would likely be much less relevant for historians in Germany, who typically publish in books and national journals. There is not a one-size-fits-all bibliometric method for research evaluation but, rather, several types of methods and indicators that can be applied to different contexts of evaluation and monitoring. On the whole, these indicators can certainly not offer a legitimate shortcut to replace traditional peer review assessments, especially at the level of individual researchers and small research groups. However, to assess the global output on the meso or macro level, bibliometrics can be quite useful, as it is perhaps the only method that can be used to compare and estimate strengths and weaknesses of institutions or

(continued)

countries. Most importantly, entities, such as researchers and institutions, should not be ranked by one indicator, but multiple metrics should be applied to mirror the complexity of scholarly communication. Moreover, quantitative metrics need to be validated and complemented with expert judgments. In addition to the importance of the application context of bibliometrics, attention should also be paid to data quality. This involves using adequate databases such as Web of Science or Scopus—instead of a blackbox like Google Scholar—and data cleaning in particular regarding author names and institutional addresses. Normalized indicators need to be used to balance out field, age and document type biases.

The Impact Factor has dominated research evaluation far too long[6] due to its availability and simplicity, and the h-index has been popular because of a similar reason: the promise to enable the ranking of scientists using only one number. For policy-makers—and, unfortunately, for researchers as well—it is much easier to count papers than to read them. Similarly, the fact that these indicators are readily available on the web interfaces of the Web of Science and Scopus add legitimacy to them in the eyes of the research community. Hence, in a context where any number beats no number, these indicators have prevailed, even though both of them have long been proven to be flawed. The same could in this book to new social-media based metrics, so-called altmetrics (see Weller 2015). The reason for the popularity of indicators such as the Impact Factor and h-index is that alternatives are not as simple and rightly available. Relative indicators that adequately normalize for field and age differences and percentile indicators that account for skewed citation distributions require access to local versions of the Web of Science, Scopus or other adequate citation indexes, and are much more difficult to understand. Multidimensional approaches are more complex than a simple ranking according to one number. Still, this is the only way fair and accurate evaluations can be performed.

The current use of simple indicators at the micro level—such as the Impact Factor and the h-index—has side effects. As evaluators reduce scientific success to numbers, researchers are changing their publication behavior to optimize these numbers through various unethical tactics. Moreover, the scientific community has been, since the beginning of the twentieth century, independent when it comes to research evaluation, which was performed through peer-review by colleagues who understood the content of the research. We are entering a system where numbers compiled by private firms are increasingly replacing this judgment. And that is the worst side

(continued)

[6] Only recently the San Francisco Declaration on Research Assessment (DORA) took a stand against the use of the Impact Factor in article and author evaluation.

effect of them all: the dispossession of researchers from their own evaluation methods which, in turn, lessens the independence of the scientific community.

References

Abt HA (1992) Publication practices in various sciences. Scientometrics 24(3):441–447. doi:10.1007/BF02051040

Adam D (2002) The counting house. Nature 415(6873):726–729

Aksnes DW (2003) A macro study of self-citation. Scientometrics 56:235–246

Archambault E, Larivière V (2009) History of the journal impact factor: contingencies and consequences. Scientometrics 79(3):635–649. doi:10.1007/s11192-007-2036-x

Bergstrom CT (2007) Eigenfactor: measuring the value and prestige of scholarly journals. Coll Res Libr News 68(5):314–316

Bhattacharjee Y (2011) Saudi universities offer cash in exchange for academic prestige. Science 334(6061):1344–1345

Biagioli M (2003) Rights or rewards? Changing frameworks of scientific authorship. In: Biagioli M, Galison P (eds) Scientific authorship: credit and intellectual property in science. Routledge, New York, pp 253–279

Birnholtz J (2006) What does it mean to be an author? The intersection of credit, contribution and collaboration in science. J Am Soc Inf Sci Technol 57(13):1758–1770

Bornmann L, Mutz R (2011) Further steps towards an ideal method of measuring citation performance: the avoidance of citation (ratio) averages in field-normalization. J Informetrics 5(1):228–230. doi:10.1016/j.joi.2010.10.009

Bradford SC (1934) Sources of information on specific subjects. Engineering 137:85–86. doi:10.1177/016555158501000407

Burton RE, Kebler RW (1960) The half-life of some scientific and technical literatures. Am Doc 11(1):18–22. doi:10.1002/asi.5090110105

Calver MC, Bradley JS (2009) Should we use the mean citations per paper to summarise a journal's impact or to rank journals in the same field? Scientometrics 81(3):611–615. doi:10.1007/s11192-008-2229-y

Cameron BD (2005) Trends in the usage of ISI bibliometric data: uses, abuses, and implications. Portal Libr Acad 5(1):105–125. doi:10.1353/pla.2005.0003

Cole JR, Cole S (1967) Scientific output and recognition: a study in the operation of the reward system in science. Am Soc Rev 32(3):377–390

Cole JR, Cole S (1973) Social stratification in science. University of Chicago Press, Chicago

Cole FJ, Eales NB (1917) The history of comparative anatomy. Part I: a statistical analysis of the literature. Sci Prog 11(43):578–596

Costas R, van Leeuwen TN, Bordóns M (2010) Self-citations at the meso and individual levels: effects of different calculation methods. Scientometrics 82(3):517–537

De Bellis N (2009) Bibliometrics and citation analysis. From the science citation index to cybermetrics. The Scarecrow Press, Lanham

De Lange C, Glänzel W (1997) Modelling and measuring multilateral coauthorship in international scientific collaboration. Part I. Development of a new model using a series expansion approach. Scientometrics 40(3):593–604

Engels TCE, Ossenblok TLB, Spruyt EHJ (2012) Changing publication patterns in the social sciences and humanities, 2000–2009. Scientometrics 93(2):373–390. doi:10.1007/s11192-012-0680-2

Favaloro EJ (2008) Measuring the quality of journals and journal articles: the impact factor tells but a portion of the story. Semin Thromb Hemost 34(1):7–25. doi:10.1055/s -2008-1066030

Flanagin A, Carey LA, Fontanarosa PB, Phillips SG, Pace BP, Lundberg GD, Rennie D (1998) Prevalance of articles with honorary authors and ghost authors in peer-reviewed medical journals. JAMA 280(3):222–224
Garfield E (1964) "Science citation index"—a new dimension in indexing. Science 144 (3619):649–654
Garfield E (1972) Citation analysis as a tool in journal evaluation. Journals can be ranked by frequency and impact of citations for science policy studies. Science 178(4060):471–479
Garfield E (1979) Citation indexing - its theory and application in science, technology, and humanities. Wiley, New York
Garfield E (1983) Citation indexing. Its theory and application in science, technology, and humanities. ISI Press, Philadelphia
Gingras Y, Larivière V (2011) There are neither "king" nor "crown" in scientometrics: comments on a supposed "alternative" method of normalization. J Informetrics 5(1):226–227
Glänzel W, Thjis B (2004) The influence of author self-citations on bibliometric macro indicators. Scientometrics 59(3):281–310
Glänzel W, Debackere K, Thijs B, Schubert A (2006) A concise review on the role of author self-citations in information science, bibliometrics and science policy. Scientometrics 67:263–277
Glänzel W, Schubert A, Thijs B, Debackere K (2011) A priori vs. a posteriori normalisation of citation indicators. The case of journal ranking. Scientometrics 87(2):415–424. doi:10.1007/s11192-011-0345-6
Gonzalez-Pereira B, Guerrero-Bote VP, de Moya-Anegon F (2010) A new approach to the metric of journals' scientific prestige: the SJR indicator. J Informetrics 4(3):379–391. doi:10.1016/j.joi.2010.03.002
Gross PLK, Gross EM (1927) College libraries and chemical education. Science 66(1713):385–389. doi:10.1126/science.66.1713.385
Haustein S (2012) Multidimensional journal evaluation. Analyzing scientific periodicals beyond the impact factor. De Gruyter Saur, Berlin
Hemmingsson A, Mygind T, Skjennald A, Edgren J (2002) Manipulation of impact factors by editors of scientific journals. Am J Roentgenol 178(3):767–767
Hicks D (2013) One size doesn't fit all: on the co-evolution of national evaluation systems and social science publishing. Confero 1(1):67–90
Hirsch JE (2005) An index to quantify an individual's scientific research output. Proc Natl Acad Sci U S A 102(46):16569–16572. doi:10.1073/pnas.0507655102
Hvistendahl M (2013) China's publication bazaar. Science 342(6162):1035–1039. doi:10.1126/science.342.6162.1035
Larivière V, Gingras Y (2010) On the prevalence and scientific impact of duplicate publications in different scientific fields (1980–2007). J Doc 66(2):179–190
Larivière V, Gingras Y (2011) Averages of ratios vs. ratios of averages: an empirical analysis of four levels of aggregation. J Informetrics 5(3):392–399. doi:10.1016/j.joi.2011.02.001
Larivière V, Gingras Y, Archambault E (2006) Canadian collaboration networks: a comparative analysis of the natural sciences, social sciences and the humanities. Scientometrics 68(3):519–533
Larivière V, Archambault É, Gingras Y (2008) Long-term variations in the aging of scientific literature: from exponential growth to steady-state science (1900–2004). J Am Soc Inf Sci Technol 59(2):288–296
Leydesdorff L (2009) How are new citation-based journal indicators adding to the bibliometric toolbox? J Am Soc Inf Sci Technol 60(7):1327–1336. doi:10.1002/asi.21024
Leydesdorff L, Opthof T (2010) Normalization at the field level: fractional counting of citations. J Informetrics 4(4):644–646. doi:10.1016/j.joi.2010.05.003
Lotka AJ (1926) The frequency distribution of scientific productivity. J Wash Acad Sci 16 (12):317–323
Lozano GA, Larivière V, Gingras Y (2012) The weakening relationship between the impact factor and papers' citations in the digital age. J Am Soc Inf Sci Technol 63(11):2140–2145

Martin BR (2013) Whither research integrity? Plagiarism, self-plagiarism and coercive citation in an age of research assessment. Res Policy 42:1005–1014. doi:10.1016/j.respol.2013.03.011
Martin BR, Irvine J (1983) Assessing basic research. Some partial indicators of scientific progress in radio astronomy. Res Policy 12:61–90
Martyn J, Gilchrist A (1968) An evaluation of British scientific journals. Aslib, London
Merton RK (1973) The sociology of science: theoretical and empirical investigations. University of Chicago Press, Chicago
Merton RK (1977) The sociology of science: an episodic memoir. In: Merton RK, Gaston J (eds) The sociology of science in Europe. Southern Illinois University Press, Carbondale, pp 3–141
Merton RK (1988) The Matthew effect in science. II. Cummulative advantage and the symbolism of intellectual property. ISIS 79:606–623
Moed HF (2002) The impact-factors debate: the ISI's uses and limits. Nature 415(6873):731–732
Moed HF (2005) Citation analysis in research evaluation. Springer, Dordrecht
Moed HF (2010) Measuring contextual citation impact of scientific journals. J Informetrics 4 (3):265–277. doi:10.1016/j.joi.2010.01.002
Moed HF, van Leeuwen TN (1995) Improving the accuracy of institute for scientific information's journal impact factors. J Am Soc Inf Sci 46(6):461–467
Norris M, Oppenheim C (2003) Citation counts and the research assessment exercise V: archaeology and the 2001 research assessment exercise. J Doc 59(6):709–730
Otlet P (1934) Traité de documentation: le livre sur le livre, théorie et pratique. Editiones Mundaneum, Brussels
Pontille D (2004) La Signature Scientifique: une Sociologie Pragmatique de l'Attribution. CNRS Éditions, Paris
Price DJDS (1963) Little science, big science. Columbia University Press, New York
Price DDS (1981) Letter to the editor: multiple authorship. Science 212(4498):984–1066
Pritchard A (1969) Statistical bibliography or bibliometrics? J Doc 25(4):348–349
Rinia EJ, van Leeuwen TN, van Vuren HG, van Raan AFJ (1998) Comparative analysis of a set of bibliometric indicators and central peer review criteria. Evaluation of condensed matter physics in the Netherlands. Res Policy 27:95–107
Rogers LF (2002) Impact factor: the numbers game. Am J Roentgenol 178(3):541–542. doi:10.2214/ajr.178.3.1780541
Rowlands I, Nicholas D (2005) New journal publishing models – an international survey of senior researchers. CIBER, University College London, London. Available from: http://www.homepages.ucl.ac.uk/~uczciro/pa_stm_final_report.pdf
Schubert A, Braun T (1996) Cross-field normalization of scientometric indicators. Scientometrics 36(3):311–324
Seglen PO (1992) The skewness of science. J Am Soc Inf Sci 43(9):628–638
Seglen PO (1997a) Citations and journal impact factors: questionable indicators of research quality. Allergy 52(11):1050–1056
Seglen PO (1997b) Why the impact factor of journals should not be used for evaluating research. BMJ 314(7079):498–502. doi:10.1136/bmj.314.7079.497
Smalheiser NR, Torvik VI (2009) Author name disambiguation. Annu Rev Inf Sci Technol 43:1–43. doi:10.1002/aris.2009.1440430113
Small H (1987) The significance of bibliographic references. Scientometrics 8:321–340
Snyder H, Bonzi S (1998) Patterns of self-citation across disciplines. J Inf Sci 24(6):431–435
The PLoS Medicine editors (2006) The impact factor game—it is time to find a better way to assess the scientific literature. PLoS Med 3(6):e291. doi:10.1371/journal.pmed.0030291
Tijssen RJW, Visser MS, van Leeuwen TN (2002) Benchmarking international scientific excellence: are highly cited research papers an appropriate frame of reference? Scientometrics 54 (3):381–397
Todorov R, Glänzel W (1988) Journal citation measures—a concise review. J Inf Sci 14(1):47–56. doi:10.1177/016555158801400106

Van Norden R (2013) Brazilian citation scheme outed. Thomson Reuters suspends journals from its rankings for 'citation stacking'. Nature 500(7464):510–511

Waltman L, van Eck NJ (2010) The relation between Eigenfactor, audience factor, and influence weight. J Am Soc Inf Sci Technol 61(7):1476–1486. doi:10.1002/asi.21354

Waltman L, van Eck NJ (2012) The inconsistency of the h-index. J Am Soc Inf Sci Technol 63 (2):406–415. doi:10.1002/asi.21678

Weingart P (2005) Impact of bibliometrics upon the science system: inadvertent consequences? Scientometrics 62(1):117–131. doi:10.1007/s11192-005-0007-7

Weller K (2015) Social media and altmetrics: an overview of current alternative approaches to measuring scholarly impact. In: Welpe IM, Wollersheim J, Ringelhan S, Osterloh M (eds) Incentives and performance - governance of research organizations. Springer, Cham

Wislar JS, Flanagin A, Fontanarosa PB, DeAngelis CD (2011) Honorary and ghost authorship in high impact biomedical journals: a cross sectional survey. BMJ 343:6128. doi:10.1136/bmj.d6128

Zipf GK (1949) Human behaviour and the principle of least effort. Addison-Wesley, Reading, MA

Zitt M (2010) Citing-side normalization of journal impact: a robust variant of the audience factor. J Informetrics 4(3):392–406

Zuckerman H (1987) Citation analysis and the complex problem of intellectual influence. Scientometrics 12:329–338

The Informative Value of International University Rankings: Some Methodological Remarks

Ulrich Schmoch

Abstract Global university rankings are critically discussed concerning their outcome, e.g., pointing to contradictory results. In contrast, this paper examines three issues with the example of three well known rankings (Shanghai, Times and Taiwan). First, on which methods, in particular on which indicators are they based? Second, are there fundamental deficits of concept and systematic biases? Third, can one predict good or bad ranking scores, based on the respective concept? In addition with the example of German universities there are two factors on performance measurement discussed, disciplinary cultures and size of the university. In sum, the usefulness of global rankings is considered to be highly questionable. Measures in science policy should not be legitimated with such rankings.

1 Introduction

In 2003 the first so called Shanghai ranking was published. It triggered an enormous resonance. As early as 2004, the international organization UNESCO-CEPES (Centre Européen pour l'Enseignement Supérieur) held a conference on international rankings and established an "International Expert Group Created to Improve Higher Education Rankings". A report about this event and a comparative analysis of ratings in different countries (Dill and So 2005) had a significant effect[1] reflecting the considerable interest in this topic. The different contributions can be classified into three major types (Teichler 2011, p. 59): (1) Publications by actors and sympathizers of the "ranking movement", (2) fundamental critics and sceptics of rankings, and (3) publications by higher education researchers.

[1] 320 citations in Google Scholar on 16 January 2014.

U. Schmoch (✉)
University of Administrative Sciences, Speyer, Germany

Fraunhofer Institute for Systems and Innovation Research, Karlsruhe, Germany
e-mail: schmoch@uni-speyer.de

I.M. Welpe et al. (eds.), *Incentives and Performance*,
DOI 10.1007/978-3-319-09785-5_9

The most frequent criticism regarding global rankings refers first to methodological shortcomings and second to the detrimental consequences for the higher education systems in different countries (e.g., Marginson 2007; Ishikawa 2009 or Rauhvargers 2011). The first discourse deals with a comparison of the results of different rankings (Mu-Hsuan 2011; Marginson 2007; Aguillo et al. 2010; Liu and Cheng 2005 or Marginson and Wende 2007). It finds differences, but also similarities, with specific issues. Examples are the consequences of the skewed distribution of the underlying data (Florian 2007), the inadequate use of bibliometric sub-indicators (Raan 2005) or the inappropriate definition of institutions in the bibliometric searches (Liu and Cheng 2005), thus the typical problems of clean and consistent data in large data sets. The second criticism as to the negative impacts on higher education systems focuses on the "push toward concentration of resources and quality" and "steeply stratified systems" (Teichler 2011, p. 65f, see also Kehm 2014) or the neglect of education (Hazelkorn 2011). The higher education institutions (HEIs) and the responsible policymakers are aware of rankings. Positive results are published in annual reports. The pursuit of an improvement in rankings is a major argument for radical measures such as the fusion of institutions. Kehm (2013) reports that some universities recruit highly cited professors and special ranking managers to boost their ranking position. To support the validity and usefulness of rankings, some authors refer to the similarity of the outcome of different rankings (e.g., Mu-Hsuan 2011). This is, however, no meaningful prove because similar rankings may have similar shortcomings. To investigate whether appropriate criteria to measure the performance of universities in an international comparison are chosen it is necessary to look at all partial indicators and their combination. Because of this methodological fussiness the claim that rankings improve the transparency as to the performance of universities must be critically considered.

To do so, in a first step, I will compile the criteria of the most popular rankings, focusing on those rankings which lead to one-dimensional league tables (Sect. 2). Based on this information, I identify several basic methodological problems (Sect. 3) implying quite different results of different rankings. Most important, one is able to predict which universities will have good or bad ranking positions beyond a closer look at their specific performance. Thus ranking positions largely depend on methodology, not on performance. In the next part, the often claimed relevance of "critical mass" and the comparability of different disciplines in rankings will be analyzed on an empirical basis (Sect. 4). The chapter concludes with a fundamental assessment with regard to the informative value and usefulness of international rankings of universities.

2 Criteria and Methodological Concerns of the Most Popular Rankings

2.1 *Criteria of the Most Popular Rankings*

The most popular international rankings of universities in the form of league tables are the

– Shanghai Academic Ranking of World Universities (ARWU)
– Times Higher Education Ranking (THE) and
– Ranking of the Higher Education Evaluation and Accreditation Council of Taiwan (HEEACT).

The U-Map, U-Multirank or the Leiden Ranking cover different clearly distinct dimensions and do not aim at simple league tables (Marginson 2007; Aguillo et al. 2010; Mu-Hsuan 2011; Vught and Ziegele 2012; Ziegele and Vught 2013). I will not discuss the latter in more detail in this chapter. The most influential global ranking, the Shanghai Ranking (ARWU), is based on the sub-indicators documented in the general description of ARWU (Table 1) as to the ranking of the total universities (covering all fields).

ARWU covers more than 1,000 universities worldwide and publishes the results for the 500 best universities for the activities in total and for the five sub-fields "natural sciences and mathematics", "engineering/technology and computer sciences", "life and agriculture sciences", "clinical medicine and pharmacy", and "social sciences". Furthermore, tables are offered for the five disciplines mathematics, physics, chemistry, computer, and economics, thus a further breakdown of

Table 1 Sub-indicators of the Shanghai Ranking as to the methodology linked to the ranking tables by fields[a]

Code	Weight (in %)	Sub-indicator
Alumni	10	Alumni of an institution winning Nobel Prices and Fields Medals
Award	15	Staff of an institution winning Nobel Prices and Fields Medals
HiCi	25	Highly cited researchers in the Web of Science (WoS)
PUB	25	Papers indexed in Science Citation Index-Expanded (SCI) and the Social science Citation Index (SSCI)
TOP	25	Percentage of papers published in top 20 % journals of SCI or SSCI fields to that in all SCI or SSCI journals

[a]The sub-indicators in engineering are defined in a specific way, the sub-indicators for the totals slightly differ from those as to fields
Source http://www.shanghairanking.com/ARWU-FIELD-Methodology-2013.html

Table 2 Areas of analysis of the Times Ranking

No	Area of analysis	Weight (in %)
1	Teaching: the learning environment	30
2	Research: volume, income, reputation	30
3	Citations: research influence	30
4	Industry income: innovation	2.5
5	International outlook: staff, students and research	7.5

Source http://www.timeshighereducation.co.uk/world-university-rankings/2013-14/subject-ranking/subject/arts-and-humanities/methodology

Table 3 Criteria and sub-indicators of the Taiwan Ranking

Criteria	Weight (in %)	Sub-indicators	Weight (in %)
Research productivity	25	Articles of the last 11 years	10
		Articles of the current year	15
Research impact	35	Citations of the last 11 years	15
		Citations of the last 2 years	10
		Average citations of the last 11 years	10
Research excellence	40	h-index of the last 2 years	10
		Highly cited papers	15
		Articles of the current year in highly cited journals	15

Source http://nturanking.lis.ntu.edu.tw/BackgroundMethodology/Methodology-enus.aspx#2

selected sub-fields is made. The Times Higher Education Ranking (THE) comprises the areas of analysis listed in Table 2[2]:

These five sub-areas of assessment are based on 13 sub-indicators. Besides the totals, the Times Ranking differentiates by the six sub-fields “arts and humanities”, “clinical pre-clinical and health”, “engineering and technology”, ”life sciences”, “physical sciences”, and “social sciences”. Thus, the sub-fields are similar to those of the Shanghai ranking, but not identical. On the Times website, the number of included universities is not indicated; but the ranking obviously covers several hundred universities, and the data for the best 400 universities are documented. The Taiwan Ranking of the Higher Education Evaluation and Accreditation Council of Taiwan (HEEACT) covers the sub-indicators listed in Table 3.

Thus, in the case of the Taiwan Ranking the indicators are exclusively bibliometric ones. The ranking is differentiated by the six sub-fields “agriculture”, “clinical medicine”, “engineering”, “life sciences”, “natural science”, and “social sciences”. In addition, a finer disaggregation by 13 subjects/disciplines is made. 500 universities are included in the ranking system.

[2] http://www.timeshighereducation.co.uk/world-university-rankings/2013-14/subject-ranking/subject/arts-and-humanities/methodology

2.2 Methodological Concerns Regarding the Most Popular Rankings

A closer look at these indicator sets of the different rankings points to various methodological concerns. The most important of them will be discussed in the following section.

First, all three rankings are based on a set of sub-indicators which are combined—with more or less arbitrary weights—to one integrated indicator, so that a simple ranking and the generation of league tables are possible. A major disadvantage of such a so-called composite indicator is that in the final value the underlying sub-indicators are not visible. A change of the sub-indicator weights can imply a change of the ranks (Grupp and Mogee 2004). The different indicator sets—in particular the Taiwan one—suggest that all performance indicators proposed in the science policy arena are mixed together to prevent every type of critique. In the Taiwan Ranking topical and longtime indicators are combined, highly cited papers and papers in highly cited journals are associated to the popular h-index, productivity, impact and excellence are put together. In the end, the reasons for a good or bad position in the ranking cannot be determined. Instead of the often claimed increase of transparency (Marginson 2007), the composite indicators increase obscurity.

Second, some of the sub-indicators prove to be less appropriate. For example, the sub-indicator of Nobel Prizes in the Shanghai ranking refers to extremely rare events, a characteristic which is not consistent with the target of the regular graduation between the different rank positions. Thus, it is not surprising that the distribution of the Shanghai scores is quite skewed and that the universities on the upper ranks have also high values of the Alumni sub-indicator linked to the Nobel Prizes.

In the case of the Taiwan ranking, the criteria of research impact and research excellence address the same aspect of quality, so that a double counting of quality is induced again, implying a skewed distribution.

In the Times Ranking, the industrial income has a weight of 2.5 %, thus it has no real impact on the final score. In the Shanghai Ranking, a similar indicator (“FUND”) is only considered for engineering, so that again the impact on the total score is negligible. These sub-indicators obviously have an alibi function to illustrate that all relevant aspects are considered.

Third, the Shanghai and the Taiwan Rankings do not account for teaching at all. Thus, the performance measure is exclusively oriented toward research. With regard to the broad spectrum of university activities, this approach is one-sided and too narrow. For students who primarily look at universities with excellent teaching the restriction on research is inadequate. Of course, the assessment of the performance in teaching is quite difficult, in particular in an objectively comparative way at an international level. In the case of the Times Ranking, the teaching sub-indicator is made up of further five sub-indicators:

Reputation survey (15 %)
Ratio of students to staff (4.5 %)
Doctoral to bachelor's degrees (2.25 %)
Doctorates per academic staff (6 %)
Institutional income scaled against academic staff numbers (2.25 %)

The survey of the Times Ranking with reference to teaching is performed by Thomson Reuters and refers to the reputation in research and teaching. Thus, the survey is not exclusively oriented toward teaching. Furthermore, it is questionable whether any senior expert is able to objectively assess the quality of teaching. He or she studied in a specific field some decades earlier. Consequently, any judgment about the own university is arguable, even more so than that about other universities, in particular those in foreign countries. In addition, Bowman and Bastedo (2011) show that the survey results of subsequent years are not independent. It is also questionable, if sub-indicators such as "doctorates per academic staff" really reflect teaching quality.

In the Shanghai Ranking, four out of the five sub-indicators are based on absolute numbers and have a substantial weight. This means that the ranking implicitly assumes that the size of a university is linked to the level of performance. This assumption is often formulated as the need for a "critical mass" for achieving relevant scientific research performance. The disputable validity of this supposition will be discussed further below. The Times Ranking scales all absolute values against the staff numbers. In principle, this approach appears to be more reasonable. However, the Times collect all data from the universities themselves. Then one cannot be sure that the universities collect their data according to comparable standards. Scholars with experiences of different national academic systems, e.g., the French versus the US-American one, will have doubts whether these systems are really comparable. Also it is well known that the OECD invests enormous cleaning efforts in making the numbers on research staff internationally comparable. So it is quite probable that the staff numbers in the Times Ranking are not comparable, but it is not clear whether there are systematic country effects and in which direction they influence the outcome. In the Taiwan Ranking, the number and impact of the articles and citations are normalized by the full-time faculty obtained from the websites of the universities. Again, the question of international comparability comes up. In the excellence category, the number of highly cited papers and that of articles in highly cited journals are not normalized.

Fourth, all rankings refer to the publication database Web of Science (WoS) or its sub-databases SCI and SSCI in a substantial way (Shanghai: three sub-indicators with a weight of 75 %; Times: the sub-indicator of research influence is completely based on the WoS and has a weight of 30 %; 6 % of the sub-indicators of the research category are based on the WoS as well; Taiwan: All sub-indicators are based on WoS-data, i.e., 100 %). The relevant use of the WoS implies that the final results of the rankings are significantly dependent on the WoS. All distortions of the WoS are directly transmitted to the outcomes of the rankings. Major distortions of the WoS are:

Language bias: The majority of journals indexed in the WoS are US-American.[3] More than half of the publications in the WoS are published in US journals. This implies that authors from non-US-countries are less represented in WoS than US ones. This observation was discussed as language bias, English versus non-English journals. So it is not surprising that Raan et al. (2011) diagnose severe language effects in the WoS with substantial disadvantages for French and German authors. This observation primarily affects countries with a large domestic language area such as France, Germany, Italy, Spain or Japan (Schmoch et al. 2011). Authors from smaller countries generally publish more in English-language journals, as few domestic journals with a high international reputation exist. The United States represents a large domestic market so that articles in US journals have a broad readership and in consequence the citations of articles in US journals are quite high. Thus, the language bias is also a bias towards the United States (Michels and Schmoch 2014; Schubert and Michels 2013; Mueller et al. 2006). In any case, US universities will have a substantial advantage in rankings due to the input from the WoS.

Bias in favour of medicine: The historic core of the WoS is medicine, and to this day, medicine is the largest field in the WoS by far. Almost a quarter of all articles in WoS is classified in medicine. As a result, universities with a medical department have better pre-conditions to achieve a good rank.

Bias to the disadvantage of social sciences: Social sciences and the humanities account for less than 10 % of the WoS articles. This low number is due to the fact that many articles in the social sciences and humanities are not published in journals, but as contributions to books (Hicks 2004). Furthermore, in the social sciences and humanities, many papers are published in domestic journals in the native language and are not covered by the WoS. Again, countries with large domestic language areas are more affected by this phenomenon (cf. the above explanations as to the language bias). The publication and citation cultures in different disciplines differ considerably. In particular the average number of papers per scientist and the number of citations per paper vary substantially. For example, in 2010, the average citation rate of German papers in medicine was about 8.0, in mathematics 2.0. In standard bibliometrics, this phenomenon has been known for many years. Various methods are applied to normalize by field in order to make different disciplines comparable.[4] In none of the rankings such standardizations are used. The critique by Raan (2005) on the Shanghai Ranking highlights this special shortcoming. At a university, there is always a mixture of many disciplines and its accidental strengths in fields with high paper per scientist ratios may imply a good rank. This mixture of disciplines may be the major reason why Ioannidis et al. (2007) find a decreasing precision with growing institutional entities. The consequences of normalization of citations are shown further below. At first sight the problem of field differences seems to be solved by field-specific tables which

[3] Defined by the location of the headquarter of the publisher.

[4] Examples of a normalization are discussed in Sect. 3.1.

are offered by each of the three rankings. But the differentiation is very coarse and the variation within a field is still rather large so that the rankings at the field level are an improvement, but not a satisfying solution.

3 Why Different Rankings Lead to Different Results

3.1 Comparing the Three Most Popular Rankings

At first sight, it may be assumed that the global rankings by different providers lead to similar results. A more detailed look at the sub-indicators and their combination to final scores clearly shows that the results must be quite different. The following example of three German universities illustrates that the positions in the rankings vary considerably (Table 4). For example, the position of the FU Berlin in the total ranking according to Shanghai is extremely bad, according to the Times quite good, and according to Taiwan again bad. The bad ranks in Shanghai and Taiwan may be linked to the considerable weight of bibliometric indicators in these rankings and the low coverage of the social sciences and humanities in WoS. About half of the academic staff of the FU Berlin works in the social sciences and humanities. But there are also analogies between the three rankings, e.g., referring to the good position of the University Heidelberg which has a large medical department in the total assessment.

The contingency of the ranking positions becomes even more obvious in an international comparison. We again take TU Munich and University of Heidelberg as examples of Germany. We compare them to the well-reputed US-American universities Harvard and MIT, as well as to the University of Lund in Sweden and the well-known Swiss Federal Institute of Technology in Zurich (ETH Zurich) as examples for smaller countries. In addition to the three rankings Shanghai, Taiwan,

Table 4 Position of three selected German universities according to different global rankings, 2013

	Shanghai		Times		Taiwan	
University	Total rank	Medicine	Total rank	Medicine	Total rank	Medicine
TU Munich	50	101–150	87	>100	89	107
FU Berlin	>500	>200	86	>100	231	254
University of Heidelberg	54	76–100	63	40	56	39

Sources Websites of the Shanghai, Times, and Taiwan rankings, own compilation

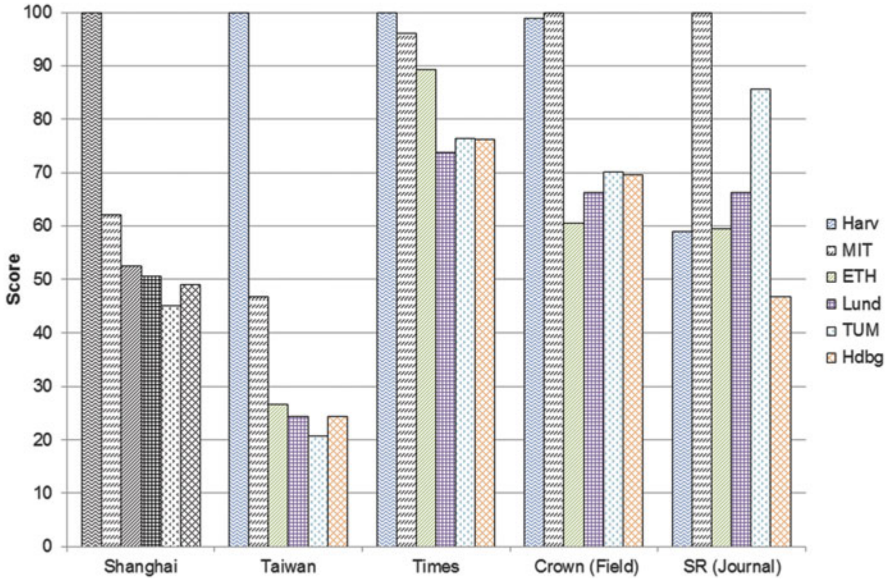


Fig. 1 Position of selected universities according to different ranking approaches (*Sources* Websites of the Shanghai, Times, and Taiwan rankings, Web of Science, own compilation)

and the Times, I computed the rank positions based on to two bibliometric citation-based indicators: The Crown Indicator[5] suggested by the CWTS group at the University Leiden[6] and the Scientific Regard (SR) suggested by Grupp et al. (2001). In the case of the Crown Indicator, the citations to each publication are normalized by the average citation rate in the field of the publication with almost 250 fields (subject categories) defined in the WoS. This differentiation is much more fine-grained than the few fields suggested in the global rankings. However, the average rates are still dominated by publications in US journals, as the latter are the most frequent ones in the WoS. The Scientific Regard normalizes the citations per publication by the average citation rates of the journal where the publication is included. This measure implies an even finer disaggregation of fields and compensates for the dominance of US journals.[7] Figure 1 is organized in such a way that the highest value for each approach is put to 100. As a first observation, the skewed distributions in the Shanghai and Taiwan Rankings are illustrated. In the Times Ranking the distance between the positions of Harvard and MIT is less large, according to the Crown Indicator both are almost at the same level, and according to the SR index, the level of Harvard is much lower. Second, the position of the other

[5] In its latest version.

[6] CWTS (Centrum voor Wetenshap- en Techologie Studies, engl. Center for Science and Technology Studies), Leiden, Netherlands is a leading research center in bibliometrics.

[7] For a more detailed explanation of the SR index cf. Michels and Schmoch (2014) and Schmoch et al. (2012).

four universities is generally at a similar level, but is different for each approach. E.g., the ETH Zurich holds a strong position in the Times Ranking, but less strong as to the Crown Indicator or the SR index. To sum up, the rankings lead to arbitrary results. In the case of Shanghai, the Times and Taiwan, the underlying reasons are diffuse. In the case of the two citation-based indicators, one is able to follow the reasons for the ranking. However, major performance dimensions, are not covered, in particular teaching.

3.2 The Relation of Size and Performance

In the Shanghai ranking, the size of a university is implicitly linked to its performance. To check this assumption, we collected the number of publications in WoS for the natural sciences for almost 80 German universities and in addition the referring staff numbers in the natural sciences from the so-called EUMIDA data set.[8] The scatter plot of the number of publications per staff versus the staff number as proxy for the size of the university shows a positive moderate correlation of $r = 0.46$. In the lower part the correlation is quite strong up to a size of about 300 staff members, then the relation between size and publication intensity is quite flat—even slightly negative (Fig. 2). As a consequence, for the whole range a

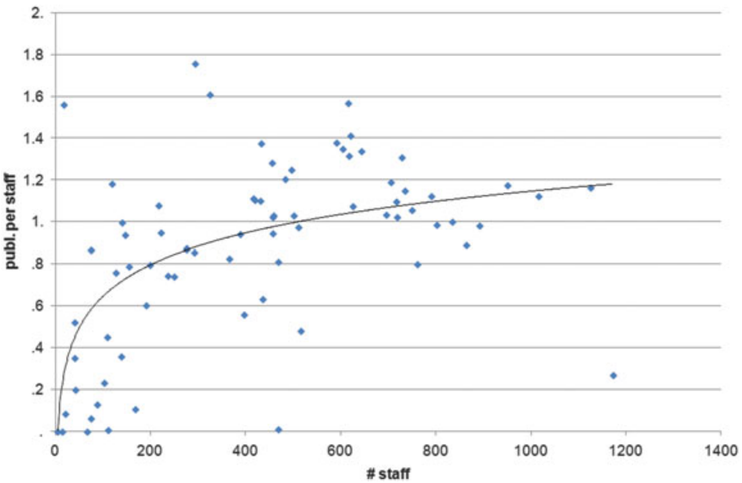


Fig. 2 Publications per staff in the natural sciences indexed in WoS by staff number for German universities, 2008 (*Sources* WoS, own compilations)

[8] Download from http://datahub.io/de/dataset/eumida

Table 5 Ranking of German universities in the natural sciences as to the number of publications in WoS, 2008

Rank	University	Publications
1	München U	1,309
2	Bonn U	1,142
3	Heidelberg U	1,117
4	Göttingen U	965
5	Frankfurt/M. U	953
6	Münster U	887
7	Bochum U	876
8	Hamburg U	875
9	Tübingen U	860
10	Erlangen-N. U	844
11	Köln U	840
12	KIT	835
13	Würzburg U	815
14	Berlin HU	814
15	Mainz U	810

Sources WoS, own compilations

logarithmic relation appears to be more appropriate than a linear one. It is plausible for the natural sciences that larger universities can provide better laboratory equipment and that, after a certain threshold, further improvement is almost impossible. The relation between staff size and publication intensity might be different for the social sciences. In any case, the assumption that larger universities achieve a higher performance is appropriate only up to a certain threshold. As a consequence, in the Shanghai Ranking very large universities are overrated.

A further observation is the considerable variation of the publication intensity by university. The main reason is the high variation of the publication habits in the different sub-fields of larger field such as the natural sciences. This finding illustrates that the differentiation by about five large fields in the three rankings considered is insufficient. The data set for German universities allows demonstrating that rankings based on absolute or relative data lead to different outcomes. On the basis of absolute publication numbers, the University of Munich (LMU) is in first position (Table 5), according to the publication intensity, the University of Gießen comes first (Table 6). The University of Munich does not appear on the relative list, some universities, such as the University of Göttingen, are on both. In any case, the ranking criteria have a severe influence.

Table 6 Ranking of German universities in the natural sciences as to the number of publications per staff in WoS, 2008

Rank	University	Publ./staff
1	Gießen U	1.757
2	Rostock U	1.611
3	Göttingen U	1.571
4	Witten/Herdecke U	1.564
5	Bochum U	1.415
6	Berlin HU	1.380
7	Marburg U	1.377
8	Würzburg U	1.352
9	Tübingen U	1.339
10	Mainz U	1.316
11	Frankfurt am Main U	1.310
12	Kiel U	1.284
13	Leipzig U	1.252
14	Regensburg U	1.205
15	Köln U	1.192

Sources WoS, EUMIDA, own compilations

Conclusions

International rankings offering league tables of universities as to "performance" lead to arbitrary results where the reasons for a specific rank position cannot be reproduced due to the mix of the underlying sub-indicators. The results of the different rankings differ, as the underlying methodology differs.

There is evidence that the performance in research increases with the size of a university in areas requiring substantial laboratory equipment, but only at the lower end. There is no evidence that the performance of very large universities is higher than that of medium ones. In any case, it is decisive whether absolute or relative measures are used for ranking indicators.

The positions in the Shanghai Ranking are predictable to a certain extent: small universities with no medical department, from a country with a large domestic non-English language area, and with a focus on the social sciences or humanities will have low ranks and vice versa.

The detailed positions have the character of a lottery, and a substantial re-organization of universities such as institutional fusions should not be justified by an improvement of the position in international rankings.

A basic requirement for rankings of high methodological quality is that the different dimensions of their performance are clearly visible. Instead of a simple composite indicator a profile of the different dimensions is much more meaningful. Such profiles would reflect the various activities of universities such as teaching, research, knowledge transfer to enterprises, knowledge transfer to the civil society, internationalization, regional engagement, policy

(continued)

advice etc. beyond research. If a ranking is intended, this is only useful between universities with similar profiles. Furthermore, rankings should never refer to total universities encompassing all disciplinary fields. They should rather focus on specific fields or disciplines.

The comparison of universities with similar profiles may be used as benchmark tool. E. g., a university with a clear focus on transfer to enterprises could look at the research level of other universities with a similar focus.

The analysis of profiles would open up the possibility to change the discussion with science departments and donors of base funds. The universities could bring in more performance dimensions than research performance. As a consequence, a science ministry could decide on a more rational basis which types of universities are useful for a region (Country, state . . .).

The most serious activities in this direction are presently undertaken by the projects U-Map and U-Multirank (Vught and Ziegele 2012). Of course, it is possible to raise various concerns against these approaches. But the participating universities have a broad voice in the selection of appropriate indicators and the methods of their analysis. All in all, my recommendation is that universities refuse any further participation in simple global rankings aiming at league tables and engage in more sophisticate activities such as U-Multirank instead. The major limitation of this approach will be the workload that the participating universities have to invest for providing appropriate data as to the different dimensions of performance.

Acknowledgement I thank Nicole Schulze for the careful conduct of the bibliometric analysis and for various helpful comments.

Certain data included herein is derived from the Science Citation Index, Social Science Citation Index and Arts & Humanities Citation Index databases prepared by Thomson Reuters (Scientific) Inc. (TR®), Philadelphia, PA, USA:

References

Aguillo IF, Bar-Ilan J, Levene M et al (2010) Comparing university rankings. Scientometrics 85 (1):243–256. doi:10.1007/s11192-010-0190-z

Bowman NA, Bastedo MN (2011) Anchoring effects in world university rankings: exploring biases in reputation scores. Higher Educ 61(4):431–444

Dill DD, So M (2005) Academic quality, league tables, and public policy: a cross-national analysis of university ranking systems. Higher Educ 49(4):495–533

Florian RV (2007) Irreproducibility of the results of the Shanghai academic ranking of world universities. Scientometrics 72(1):25–32

Grupp H, Mogee ME (2004) Indicators for national science and technology policy: how robust are composite indicators? Res Policy 33(9):1373–1384

Grupp H, Schmoch U, Hinze S (2001) International alignment and scientific regard as macro-indicators for international comparisons of publications. Scientometrics 51(2):359–380

Hazelkorn E (2011) How rankings are reshaping higher education. In: Climent V, Michavila F, Ripolles M (eds) Los rankings univeritarios: Mitos y Realidades, Ed. Tecnos

Hicks D (2004) The four literatures of social science. In: Moed H, Glaenzel W, Schmoch U (eds) Handbook of quantitative science and technology research. Kluwer, Dordrecht, pp 473–496

Ioannidis JPA, Nikolaos A, Patsopoulos NA, Fotini K et al (2007) International ranking systems for universities and institutions: a critical appraisal. BMC Med 5:30. doi:10.1186/1741-7015-5-30

Ishikawa M (2009) University rankings, global models, and emerging hegemony. Critical analysis from Japan. J Stud Int Educ 13(2):159–173. doi:10.1177/1028315308330853

Kehm BM (2013) Der Kampf ums Treppchen. Wie Hochschulen bei der Lieferung von Daten für internationale Rankings schummeln. Paper presented at the Fraunhofer ISI workshop “Perspektiven das deutsche Hochschulsystem”, Berlin, 21–22 Nov 2013. Acessible through http://www.isi.fraunhofer.de/isi-de/p/veranstaltungen/ws_ho-forschung_11-2013.php

Kehm BM (2014) Global university rankings—impacts and unintended side effects. Eur J Educ 49 (1):102–112

Liu NC, Cheng Y (2005) The academic ranking of world universities. Higher Educ Eur 30(2):127–136. doi:10.1080/03797720500260116 Scopus 76

Marginson S (2007) Global university rankings: implications in general and for Australia. J Higher Educ Policy Manage 29(2):131–142

Marginson S, van der Wende M (2007) To rank or to be ranked: the impact of global rankings in higher education. J Stud Int Educ 11(3–4):306–329. doi:10.1177/1028315307303544

Michels C, Schmoch U (2014) Impact of bibliometric studies on the publication behaviour of authors. Scientometrics 98(1):369–385

Mueller PS, Murali NS, Cha SS, Erwin PJ, Ghosh AK (2006) The association between impact factors and language of general internal medicine journals. Swiss Med Wkly 136:441–443

Mu-Hsuan H (2011) A comparison of three major academic rankings for world universities: from a research evaluation perspective. J Libr Inf Stud 9(1):1–25

Rauhvargers A (2011) Global university rankings and their impact. Leadership for WorldClass Universities. Challenges for Developing Countries, No. European University Association, Brussels

Schmoch U, Mallig N, Michels C, Neuhäusler P, Schulze N (2011) Performance and structures in the German science system in an international comparison 2010. Studien zum deutschen Innovationssystem 8-2011. EFI, Berlin

Schmoch U, Michels C, Schulze N, Neuhäusler P (2012) Performance and structures of the German science system 2011. Germany in an international comparison, China’s profile, behaviour of German authors, comparison of Web of Science and Scopus. Studien zum deutschen Innovationssystem Nr. 9–2012. EFI, Berlin

Schubert T, Michels C (2013) Placing articles in the large publisher nations: is there a “free lunch” in terms of higher impact? J Am Soc Inf Sci Technol 64(3):596–611

Teichler U (2011) Social contexts and systemic consequences of university rankings: a meta-analysis of the ranking literature. In: Shin JC, Toutkoushian RK, Teichler U (eds) University rankings. Theoretical basis, methodology and impact on global higher education. Springer, Dordrecht, pp 55–69

van Raan AFJ (2005) Fatal attraction: conceptual and methodological problems in the ranking of universities by bibliometric methods. Scientometrics 62(1):133–143. doi:10.1007/s11192-005-0008-6

van Raan AFJ, van Leeuwen TN, Martijn S, Visser MS (2011) Severe language effect in university rankings: particularly Germany and France are wronged in citation-based rankings. Scientometrics 88(2):495–498

van Vught FA, Ziegele F (eds) (2012) Multidimensional ranking. The design and development of U-multirank. Springer, Dordrecht

Ziegele F, van Vught F (2013) “U-Multirank” und “U-Map” als Ansätze zur Schaffung von Transparenz im europäischen und globalen Hochschulsystem – Konzepte und Erfahrungen. Beiträge zur Hochschulforschung 35(2):50–74

Frontier Efficiency Analysis in Higher Education

Stefano Nigsch and Andrea Schenker-Wicki

Abstract Over the last 20–30 years, many European governments have implemented reforms to improve the efficiency and competitiveness of their national higher education and research systems. They have granted universities more autonomy while introducing new accountability tools and fostering competition through performance-based funding schemes. The growing emphasis on productivity and efficiency has led to the diffusion of a variety of performance indicators, including publication and citation counts, and university rankings. Another approach increasingly applied in the higher education sector is frontier efficiency analysis. Similarly to university rankings, efficiency analyses include several indicators for research and teaching in order to assess the performance of a university or a university department. However, as opposed to most rankings, they relate the outputs to the inputs used and do not necessarily favor larger or richer institutions. Moreover, estimation techniques such as Data Envelopment Analysis (DEA) do not require any assumption about the form of the production function and allow for different factor combinations to achieve efficiency. The method thus accounts for the diversity among universities and does not necessarily penalize more teaching-oriented institutions as compared to research-oriented ones. In this contribution we present the frontier efficiency approach and its application to higher education, highlighting the main estimation techniques and methodological specifications. We provide an overview of studies that have applied DEA to the higher education sector and discuss their results, methodological contributions, and shortcomings. We conclude by identifying the advantages and limitations of frontier efficiency approaches as compared to other performance measures in higher education and delineating possible areas for further research.

S. Nigsch (✉) • A. Schenker-Wicki
Department of Business Administration, University of Zurich, Zurich, Switzerland
e-mail: stefano.nigsch@business.uzh.ch

I.M. Welpe et al. (eds.), *Incentives and Performance*,
DOI 10.1007/978-3-319-09785-5_10

1 Introduction

Over the last 30 years, the overall structure of the European higher education system and the way universities are organized have fundamentally changed. Growing student numbers and the high pace of technological development have led to a cost explosion that could not be fully covered by public funding. Among other things, universities must provide expensive facilities and services to their students and researchers in order for them to succeed in the modern economy and in global scientific competition (see, for example, Archibald and Feldman 2011). For these reasons, many governments have implemented reforms aimed at increasing the cost efficiency of universities. Following the example of the North American system, they have granted universities more autonomy while introducing accountability tools and fostering competition through performance-based funding schemes. Market-like mechanisms were expected to increase the efficiency and responsiveness of universities, while the newly created quality assurance and accreditation agencies were to assure the quality and comparability of the educational services offered. Moreover, the limited availability of funds and the increasing competition at national and international level have forced many universities to introduce new management approaches in order to increase their productivity and fulfil their tasks as efficiently as possible (see, for example, Deem and Brehony 2005).

Both the need for external accountability and the marketization of higher education have led to the introduction of a variety of indicators for the assessment of the performances of universities, single departments, or individual scholars (Ollsen and Peters 2005). In teaching, rather simple proxies for quality such as the student-to-teacher ratio or the availability of specific services and equipment—such as libraries and computers—are often complemented by surveys of students, parents, alumni, or prospective employers. In research, quantitative bibliometric indicators such as publication and citation counts or standardized measures like the ratio of publications to faculty members have become relevant. Moreover, analysts often collect information regarding the amount of external funding or the number of patents and collaborations with industry, and carry out international reputation surveys among scholars. In order to get an overall picture of a university's performance with respect to other institutions, these indicators can be combined to create university rankings. Much attention has been given to such rankings both within and outside academia, but their methodology and their results have also been extensively criticized. Among others, university rankings are perceived as favoring larger, older, and research-oriented institutions focused on the natural sciences rather than the humanities and the social sciences, and penalizing universities not located in an English-speaking country (for a further discussion on university rankings see Hazelkorn 2011).

Another instrument that can be used to measure the performance of higher education institutions is frontier efficiency analysis. This approach stems from economics and has been widely applied in industrial analysis and in public domains such as the energy and transportation sectors. Since the late 1980s and early 1990s, it has increasingly been applied to assess the efficiency of universities. Like university rankings, frontier efficiency analysis allows for the inclusion of several indicators for research and teaching to account for the multiple tasks of universities. However, as opposed to most rankings, it always relates the inputs universities use to the outputs they generate and does not necessarily favor larger and richer institutions. The most common technique used to estimate efficiency in higher education, called Data Envelopment Analysis (DEA), compares only universities with similar combinations of input or output factors when assessing their relative performance. DEA thus accounts for the fact that different institutions may have different foci on either teaching or research and be better in some activities than in others.

Despite the advantages of frontier efficiency analysis and the evident interest of governmental institutions in such analyses, this approach has played only a minor role in the recent debate on the transformation of higher education, and has been given much less attention than university rankings both within and outside academia. Rankings are often used as indicators for the prestige and reputation of universities, but if analysts want to evaluate performance they also need to account for efficiency. The aim of this chapter is to provide an overview and discussion of the current literature on frontier efficiency analysis in higher education. First, we present the basic concepts of efficiency analysis and the main methods used. We focus on Data Envelopment Analysis and illustrate how it can be applied, indicating different methodological specifications. In a second step, we review contributions from the main scholars that have used DEA in the higher education sector and present their results. We also point to methodological issues and illustrate the solutions that have been proposed to address the technique's main shortcomings. We conclude by assessing the advantages and disadvantages of frontier efficiency analysis as compared to other performance indicators in higher education and identifying possible directions for further research.

2 Efficiency Analysis: Basics

From an economic perspective, universities can be seen as decision-making units that use several inputs (e.g., academic and administrative staff, facilities) to create different outputs (mainly graduates and research results). Given their limited financial resources and the increasing competition, they are under pressure to operate at the most efficient level. In economics, the measurement of efficiency or productivity (in this chapter we use the two terms as synonyms) goes back to Farrell (1957), who distinguished three types of efficiency. First, *technical efficiency* relates to the maximum output that can be achieved using a determinate set of inputs (output-oriented efficiency), or, put differently, the minimum amount of inputs needed to produce a certain level of output (input-oriented efficiency). Second, *allocative efficiency* describes the choice between (technically efficient) combinations of inputs given their relative cost. Third, *total economic efficiency* (or *productive efficiency*) occurs when the conditions of both technical and allocative efficiency are satisfied. Because input factors in higher education cannot be easily substituted and their costs are difficult to estimate, most studies on university efficiency focus on technical rather than allocative and total economic efficiency.

2.1 Definition of Inputs and Outputs

In order to calculate a university's efficiency, inputs and outputs need to be correctly identified. In higher education, finding appropriate indicators is a difficult task (Johnes 2004). The inputs most commonly used are the number of undergraduate and graduate students—including the average scores of the entrants—and the number of academic staff (professors, assistant professors, and scientific assistants). In some cases, technical and administrative staff is also considered. Other common input measures are total revenues, operating budget (total expenses minus wages), university expenditures on centralized academic and administrative services, or the value of non-current assets (Abbott and Doucouliagos 2003). Another possible input measure is research income acquired through funds. However, as the acquisition of third-party funds reflects the quality of research conducted at universities, research income may also be used as an output measure.

Because universities have two main tasks, at least two outputs measures need to be considered: one for teaching and one for research. As an output measure for teaching, most studies include the number of students that have successfully completed their studies (see, for example, Johnes 2008; Kempkes and Pohl 2010). The number of postgraduate and PhD completions or the number of formative credits obtained may also be considered. In a few cases, the overall number of students—usually treated as an input factor—has also been used as an output measure for teaching (Avkiran 2001). To capture research output, many scholars have included the amount of third-party funds acquired for research, or bibliometric indicators such as the number of publications in international peer-reviewed journals and citation indexes. Another possibility is to use national research performance indexes such as the ratings of the Research Assessment Exercise in the UK (Athanassopoulos and Shale 1997).

2.2 Main Approaches and Estimation Techniques

Two main approaches exist to analyze the productivity of decision-making units: estimating either a *production function* or a *production frontier* (the latter approach is often called *efficiency analysis*). The production function approach specifies a functional relation between a set of inputs and a maximum level of outputs (Bonaccorsi and Daraio 2004). By means of average relations it estimates coefficients that give information on marginal costs and allocative efficiency. When applied to higher education, this approach has specific limitations. Its main problem is that the production processes within universities are difficult to model and cannot be compared to those in industry. It is questionable whether the assumed functional relations between inputs and outputs truly apply to teaching and research. For these reasons, many higher education scholars prefer the production frontier approach, which estimates a frontier that envelops all decision-making units in the dataset. Units located on this frontier are operating efficiently, while the others exhibit inefficiencies. The larger a unit's distance to the production frontier, the higher its inefficiency as compared to the best performers.

The two most common estimation techniques used to compute the production frontier are Stochastic Frontier Analysis (SFA) and Data Envelopment Analysis (DEA). SFA is a parametric approach developed by Aigner et al. (1977) that specifies a production function and identifies two components of deviation from that function: random errors and inefficiency. DEA, first proposed by Charnes et al. (1978), is a nonparametric linear modelling approach that estimates the production frontier by assigning weights to inputs and outputs and then comparing similar decision-making units. Its main advantage is that it is solely based on observed values and does not require any prior assumptions regarding the form of the production function. By comparing only similar decision-making units in terms of their input and output combinations, it also accounts for existing diversity—which may be particularly relevant in higher education. The technique's main limitation is that it assumes all deviations from the benchmark are caused by inefficiencies and does not distinguish random errors. Issues of data quality are thus of central importance. Nevertheless, because production functions in teaching and research are particularly difficult to model and any prior assumptions can be challenged, DEA has become the most-used efficiency analysis approach in higher education.

DEA can calculate both input- and output-oriented efficiencies. Because universities can hardly minimize inputs such as student numbers in the short run, output-oriented approaches are more common in higher education (Bonaccorsi et al. 2006). As an example of an output-oriented DEA model, consider the case of six decision-making units *A, B, C, D, E,* and *F* producing two outputs y_1 and y_2, using one input x (a similar example is given by Johnes 2006). In a graph that plots the ratio of output y_1 to input x against the ratio of output y_2 to input x, one can easily represent the production possibility frontier by connecting the efficient units (see Fig. 1). Every decision-making unit tries to maximize its outputs using fixed inputs. *A, B, C,* and *D* are all efficient because none of them produces more of both outputs than the others. In contrast, units *E* and *F* are not efficient and lie below the production possibility frontier. Unit *E* can be compared to unit *B* in terms of its combination of outputs, while unit *F* can be related to a hypothetical unit F' on the production frontier. The ratio $0E/0B$ or $0F/0F'$ measures the two units' efficiency relative to the best performers in the sample.

Another important feature of DEA concerns the possibility of economies of scale. Scale economies occur when a university's efficiency depends on its size. For

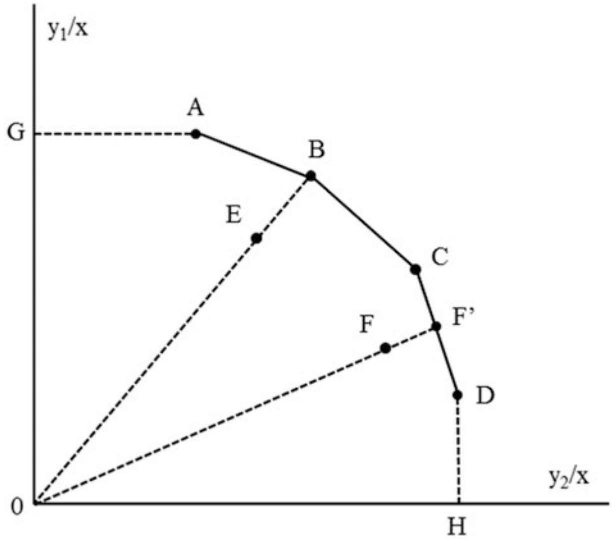


Fig. 1 Representation of an output-oriented DEA model showing the production frontier [based on Johnes (2006)]

example, larger research groups may be more productive due to their intellectual and technical resources. DEA can estimate efficiency under the assumption of constant (i.e. no scale economies) and variable returns to scale. Because the existence of scale economies in higher education is disputed and often depends on the specific situation, many scholars use both specifications in their analyses (see, for example, Agasisti and Salerno 2007; Flegg et al. 2004). Moreover, by computing the ratio of efficiency scores under constant and variable returns to scale they assess scale efficiency, or the extent to which a decision-making unit is operating at an optimal level of inputs and outputs.

2.3 Efficiency Changes over Time

When analyzing university efficiency, scholars are often interested in observing how efficiency has changed over time—for example, before and after the implementation of a higher education reform. To do so, they use a measure called Malmquist index first introduced by Caves et al. (1982). The Malmquist index compares the efficiency of a single observation at two points of time against the production probability frontier at one of these two points of time, and then computes the geometric mean of the two measures. A Malmquist index greater than the value one indicates an increase in efficiency, while Malmquist indexes below one point to a decrease. It is important to note that two main types of efficiency changes exist (see Coelli et al. 2005). First, a single decision-making unit may improve (or worsen) its efficiency compared to the other units, decreasing (or increasing) its distance from the production possibility frontier (relative efficiency change). Second, the production function itself may shift due to a general increase or decrease of productivity in the whole sector (technological change). In this case, new policies, technologies, and management approaches affect all universities in the same way. A useful advantage of the Malmquist index is that it allows for the separation of these two effects, differentiating efficiency changes of one university relative to the others from efficiency changes experienced by all institutions.

3 Main Studies and Results

In order to get an overview of existing studies on frontier efficiency analysis in higher education, we searched for studies that applied DEA to universities or university departments. We focused on articles published in international peer-reviewed journals, as we assume them to be the most considered contributions. In order to limit the scope of the review, we present the results from four countries that exhibited a particularly high number of studies: the United Kingdom, Australia, Germany, and Italy. In recent years, all of these countries have implemented important reforms to enhance the efficiency of their universities, which explains the high interest in this topic. Given that the international dimension of higher education is becoming important, we also consider some cross-country analyses.

3.1 Studies in the United Kingdom

In the late 1980s and early 1990s—about a decade after the seminal work of Charnes et al. (1978)—several scholars in the UK started applying frontier efficiency analysis to the higher education sector. At that time, discussions on university performance became relevant due to political initiatives such as the Research Assessment Exercise. Tomkins and Green (1988) conducted a first explorative study measuring the efficiency of twenty accounting departments using DEA, in order to assess what the new method might reveal. They found that DEA produced results that largely differ from more simple performance measures such as student to staff ratio and that the new technique provided valuable insights. Beasley (1990) analyzed the efficiency of chemistry and physics departments in the UK. He improved the basic model by introducing constraints for every input and output measure based on a judgment regarding its relative importance. In a later study, Beasley (1995) expanded his own model to jointly determine teaching and research efficiency of university departments.

Johnes and Johnes (1993) analyzed the efficiency of 36 economics departments. They had a very rich database at their disposal that included, among other things, several measures for research output such as papers published in academic journals, articles in popular journals, authored books, reports, and contributions to edited works. Given this multitude of inputs and outputs, they calculated 192 different DEA specifications and identified two clusters of efficiency scores. Within the clusters, efficiency scores were not sensitive to particular changes of input and outputs. The difference between the two clusters was caused by the inclusion of research grants as an input factor in the second group of estimations. Because the efficiency scores of the second cluster differed strongly from the results of the research selectivity exercise conducted by the Universities Funding Council, the authors argued that the exercise should also account for universities that achieve comparatively high levels of research output with less research grants as inputs. Athanassopoulos and Shale (1997) applied DEA to 45 whole universities in the UK. They identified cost and outcome efficiencies by including different sets of inputs in the analysis. Using the approach proposed by Beasley (1990) the authors extended the model by including restrictions reflecting the relative importance of inputs and outputs, but found similar results as in the basic DEA model. Overall, they identified six institutions that performed well both in terms of cost and outcome efficiencies, which to some extent challenges the view that being cost-efficient conflicts with outcome efficiency.

Later studies in the UK have analyzed efficiency changes over time by applying Malmquist indexes. Flegg et al. (2004) were among the first to do so, using data on 45 British universities. They found a substantial increase in efficiency over the period from 1981/1982 to 1992/1993, resulting mainly from technological change, i.e., an outward shift of the production frontier. In a later study on the efficiency of 112 British universities between 1996/1997 and 2004/2005, Johnes (2008) used a similar approach and obtained comparable results. She showed that higher education institutions experienced an annual productivity increase of 1 % that was attributable to an increase in technology of 6 % combined with a decrease in relative efficiency of 5 %. New policies and technological innovation have therefore led to an overall increase in university efficiency. However, not all universities profited equally from these changes, suggesting that some institutions need more time to adapt to the changing technology (Johnes 2008).

An important limitation of DEA is that it does not allow for statistical inference. One cannot assess whether efficiency differences among decision-making units are statistically significant. To address this problem, Johnes (2006) proposed the use of bootstrapping procedures (developed by Simar and Wilson 1998) to estimate 95 % confidence intervals. She found that most English universities were highly efficient, with efficiency scores varying from 93 to 95 % across different DEA specifications. However, some institutions exhibited rather low scores of around 60 % that proved to differ significantly from the best results.

3.2 Studies in Australia

Similarly to the UK, Australia has been reforming its higher education system in order to make it more efficient. One important set of reforms were introduced by former Labor Education Minister John Dawkins, who abolished the binary system by converting the Colleges of Higher Education into universities and introduced a funding scheme based on single institutional teaching and research profiles. Madden et al. (1997) analyzed the effects of these reforms on the efficiency of economic departments and found that while the "new universities" performed worse than the traditional ones, they managed to improve their relative efficiency in the period between 1987 and 1991. Avkiran (2001) investigated technical and scale efficiency of whole universities using three performance models: overall performance, performance on delivery of educational services, and performance on fee-paying enrollments. She found that most universities where operating at a technical- and scale-efficient level but that there was potential for improvements on fee-paying enrollments. Moreover, some universities were operating at decreasing returns to scale and could thus become more efficient by downsizing. In a similar study, Abbott and Doucouliagos (2003) found that Australian universities where performing well when compared among themselves.

Another study on Australian universities applied a two-stage approach, regressing efficiency scores obtained through DEA against various environmental variables such as academic research capacity and characteristics of the student body (Carrington et al. 2005). The authors showed that a non-metropolitan location had a positive impact on efficiency, while the proportion of students from rural and remote regions, or from a low socio-economic background, had a negative impact. Surprisingly, the proportion of academics at the level of associate professor and above was negatively correlated with efficiency, which may indicate that senior staff is less productive than junior staff. Using Malmquist indexes, the authors also found that, on average, university productivity grew 1.8 % per year from 1996 to 2000 and that this increase was mainly achieved through technical change. Worthington and Lee (2008) confirmed these findings over the period 1998–2003. They also conducted separate analyses for research-only and teaching-only productivity and concluded that efficiency gains were to a large extent attributable to improvements in research productivity.

3.3 Studies in Germany

After frontier efficiency approaches were successfully applied to British and Australian universities, they received attention in continental Europe. In Germany, Warning (2004) analyzed universities using the concept of strategic groups. She divided universities into different performance categories (low, middle, and high performance) and assumed that institutions within the same category would be characterized by similar strategic orientations. While she did not find any evidence that research-oriented universities were more efficient than teaching-oriented ones, she could show that universities focusing on the humanities and social sciences performed better. This result is probably due to the higher costs of teaching and research in the sciences. The author also found that universities in Western Germany were more efficient than those located in Eastern Germany, and that research grants were negatively related to efficiency among the best-performing universities.

Fandel (2007) analyzed the efficiency of universities in the German State of North Rhine-Westphalia and compared his results to the redistribution key for public funds proposed by the local government. He showed that most gains and losses produced by the new funding formula could not be justified in terms of efficiency. Kempkes and Pohl (2010) analyzed the efficiency of 72 public German universities using both DEA and SFA approaches and obtained similar results. Although universities in Western Germany were performing better, East German institutions had been increasing their productivity more rapidly between 1998 and 2003. By regressing efficiency scores against university-internal and -external factors, the authors found that GDP per capita had a small but significant positive impact on efficiency. Moreover, as the presence of a medical or engineering department was significantly correlated with higher efficiency, they concluded that accounting for a university's departmental composition is fundamental for obtaining unbiased efficiency scores. Entire universities are too heterogeneous to be compared using nonparametric methods.

3.4 Studies in Italy

In Italy, Agasisti and Dal Bianco (2006) analyzed 26 different DEA specifications and identified a small group of best-practice universities mostly located in Northern Italy. Moreover, they found decreasing returns to scale for large universities. In a later study, Agasisti and Salerno (2007) analyzed the efficiency of 52 Italian universities, accounting for their internal cost structure (i.e. the presence of a medical faculty) and including specific measures for education and research quality. While most institutions without medical faculties were found to be operating at increasing returns to scale, most medical universities were operating at decreasing returns to scale.

Because extreme values and outliers in a dataset may lead to biased efficiency scores, Bonaccorsi et al. (2006) studied Italian universities applying a robust nonparametric method based on order-m efficiency scores (for further methodological information see Daraio and Simar 2005). Instead of DEA they used another nonparametric estimation technique called Free Disposal Hull (FDH), and compared efficiency conditional on factors external to unconditional efficiency. By doing so, they were able to understand the impact of external variables on efficiency along the entire distribution and identify individual and localized effects. The authors found no evidence for any economies of scale and economies of scope (i.e. efficiency gains through specialization). They also analyzed the existence of trade-offs between teaching and research and showed that high scientific quality improved educational efficiency and universities that performed well in education where not necessarily less efficient in terms of research.

Abramo et al. (2008) developed and validated a benchmarking tool for research performance by combining DEA and advanced bibliometric methods. They focused on the sciences and differentiated several disciplines such as Mathematics, Physics, Chemistry, and Engineering. They found strong heterogeneity in average performance among disciplines and concluded that analyses of whole universities should account for such diversity. Moreover, they computed global efficiency scores for universities by weighting the results of single disciplines and compared them with a simple ratio of publications to researchers, finding significant differences.

In a recent study, Agasisti et al. (2012) investigated whether research departments experience a trade-off between different types of research outputs, namely the quantity of publications, quality of publications (citation indexes), public research grants, and research contracts with industry. They analyzed 69 science departments located in Lombardy and found that efficiency rankings changed considerably when considering different research-related outputs. Their results highlighted different research strategies among academic departments. While departments that produced a high number of publications also achieved high citation indexes, they were not necessarily among the best performers when it came to acquiring public and private funds. Moreover, those departments that obtained many research grants where not very productive in terms of research contracts with industry, which indicates that a trade-off exists between different types of research.

3.5 Cross-Country Studies

The diffusion of frontier efficiency techniques in Europe coincided with the implementation of the Bologna process and the emergence of a European dimension in the higher education discourse. The growing interest in this European dimension is reflected in the increasing number of cross-country efficiency analyses. Tommaso Agasisti for example, in cooperation with other researchers, has compared the efficiency of Italian universities to that of institutions in several other European countries. When compared to a common production frontier, Italian universities were shown to be less efficient than English and German institutions (Agasisti and Johnes 2009; Agasisti and Pohl 2012). However, they were increasing efficiency at a higher pace and thus catching up the lag from their neighbors in Northern Europe. Agasisti and Pohl (2012) also found that a high unemployment rate was negatively associated with university efficiency, while a positive relationship existed between efficiency and the regional share of workers employed in science and technology. In another cross-country study, Italian universities proved more efficient than those in Spain (Agasisti and Pérez-Esparrells 2010). Both systems managed to increase their efficiency over time, but while this was achieved through technological change in the Italian case, Spanish universities mainly experienced an increase in relative efficiency.

A small number of studies have included several countries in their analyses. Joumady and Ris (2005) analyzed the efficiency of universities across eight European countries. In contrast to previous studies, they focused solely on teaching and included survey data on the competencies of graduates and their employability. They observed that universities in the United Kingdom, the Netherlands, and Austria performed well in all model specifications. France and Germany were located at the average level of efficiency, while Spain, Italy and Finland achieved the lowest efficiency scores. Wolszczak-Derlacz and Parteka (2011) examined the efficiency of 259 public universities from seven European countries across the time period 2001–2005. They found rather low levels of efficiency and considerable variability in efficiency both within and across countries. Using a two-stage approach, they showed that unit size, number of and qualification of academic staff, sources of funding, and staff gender composition were among the crucial determinants of performance. Specifically, a higher share of third party funds and a higher number of women among academic staff were positively related to an institution's efficiency.

4 Conclusions

4.1 Strengths and Weaknesses of DEA

Frontier efficiency analysis in general and DEA in particular have proved to be useful tools for analyzing the performance of universities or university departments. They provide a single efficiency score that integrates a large amount of information and thus facilitates assessments and comparisons. As compared to other common performance indicators, DEA allows for more differentiation among universities, taking into account existing diversity in terms of relative size and resources or with respect to an institution's focus on specific tasks in teaching and research. By comparing similar institutions, analysts can identify specific areas where efficiency gains are possible. This may be of interest to university managers aiming to increase the performance of their institutions.

In order to overcome some of the methodological limitations of DEA, scholars have proposed and validated several extensions, including value judgments about the relative weights of inputs and outputs, bootstrapping methods, and order-m frontiers (see, for example, Bonaccorsi and Daraio 2004; Johnes 2006). However, frontier efficiency analysis only delivers information about the variables included in the analysis, and strongly depends on the data used. It does not give a comprehensive reflection of the overall performance of such complex organizations as universities. For example, valid indicators for teaching quality or other forms of university output such as consultancy are often missing. One should be very careful when drawing conclusions or policy implications based on purely quantitative indicators. It is advisable to combine efficiency results with qualitative data on universities and the context in which they are operating (Tomkings and Green 1988). Moreover, efficiency scores are always computed in relation to the best performers in the sample. DEA results do not say much about the *absolute* efficiency of a county's higher education sector—a benchmark that is itself almost impossible to define.

Finally, the level of analysis is another important issue when applying DEA. Several studies have showed that efficiency scores are likely to vary across different scientific fields and may be biased when compared to a common frontier (Kempkes and Pohl 2010; Abramo et al. 2008). For example, teaching and research in medicine or the natural sciences are more expensive than in the social sciences and the humanities, but may also acquire higher amounts of third party funds. Therefore, depending on the variables used as input and output factors, they could appear to be more or less efficient than they truly are with respect to other scientific disciplines. Accounting for these differences is fundamental, especially when analyzing whole universities. One possibility to partly overcome this problem is to analyze natural science departments separately from the social sciences and humanities or to control, for example, for the presence of a medical department (Agasisti and Pohl 2012).

4.2 Trends in University Efficiency

The studies under review have used frontier efficiency analysis to assess the efficiency of universities or university departments within and across countries. They have shown that the overall level of university efficiency has increased over the last 20–30 years and that most of the efficiency gains were due to technological change. Policy reforms, technological innovations and new management approaches seem to have had their desired effects. However, these changes have often favored already-efficient universities, widening the gap between the best and the worst performers. The effects of many reforms remain controversial and more studies directly relating political initiatives to university efficiency are needed. Given that efficiency scores are always relative to the best performers in the sample, cross-country analyses that allow for comparisons across different systems are very useful. Further research in this area seems important but is often hampered by the lack of internationally comparable micro-level data.

Further research is also needed to optimize the application of DEA and enhance the validity of its results. This may be achieved by using more valid indicators for research performance, for example, publications weighted by the journals' impact factors and by the number of authors from the institution under review. Measures for teaching quality can be updated to include items such as student satisfaction, employability, or future earnings, while other forms of university output like consultancy may also be useful. With respect to economies of scale and scope in higher education, the effects of external research funds on efficiency, or trade-offs between universities' different tasks, research results are often ambiguous and many open questions remain. The solution seems to lie in greater differentiation, whether it be in terms of specific disciplines, types of activities, or the form of the functional relationship under analysis. Efficiency analysis is most valuable when applied to smaller units that are comparable to one another, such as scientific disciplines, departments, or even research groups within a specific field.

To date, most scholars have not given practical policy recommendations regarding the use of efficiency analysis for incentive-setting and the reallocation of funds. They have limited themselves to presenting descriptive findings, noting that the results from common performance assessments often differ from the ones based on efficiency analysis (Johnes and Johnes 1993; Fandel 2007). Normative statements based on efficiency analysis are problematic because efficiency scores do not capture the whole performance of a university or a university department. Approaches such as DEA may offer precious insights, but need to be combined with other performance measures to obtain a result that is maximally accurate. Moreover, as long as large units are analyzed, efficient subunits will offset the performance of inefficient ones and vice versa, reducing the explanatory power of the results. In order to exploit the whole potential of frontier efficiency analysis, more sophisticated indicators for inputs and outputs are necessary. Valid and comparable data needs to be collected, both at the national and international levels.

References

Abbott M, Doucouliagos C (2003) The efficiency of Australian universities: a data envelopment analysis. Econ Educ Rev 22(1):89–97

Abramo G, D'Angelo CA, Pugini F (2008) The measurement of Italian universities' research productivity by a non parametric bibliometric methodology. Scientometrics 76(2):225–244

Agasisti T, Dal Bianco A (2006) Data envelopment analysis to the Italian university system: theoretical issues and policy implications. Int J Bus Perform Meas 8(4):344–367

Agasisti T, Johnes G (2009) Beyond frontiers: comparing the efficiency of higher education decision making units across more than one country. Educ Econ 17(1):59–79

Agasisti T, Pérez-Esparrells C (2010) Comparing efficiency in a cross-country perspective: the case of Italian and Spanish state universities. Higher Educ 59:85–103

Agasisti T, Pohl C (2012) Comparing German and Italian public universities: convergence or divergence in the higher education landscape? Manage Dec Econ 33:71–85

Agasisti T, Salerno C (2007) Assessing the cost efficiency of Italian universities. Educ Econ 15 (4):455–471

Agasisti T, Catalano G, Landoni P, Verganti R (2012) Evaluating the performance of academic departments: an analysis of research-related output efficiency. Res Eval 21:2–14

Aigner DC, Lovell CAK, Schmidt P (1977) Formulation and estimation of stochastic frontier production function models. J Econ 6:21–38

Archibald RB, Feldman DH (2011) Why does college cost so much? Oxford University Press, New York

Athanassopoulos AD, Shale E (1997) Assessing the comparative efficiency of higher education institutions in the UK by the means of data envelopment analysis. Educ Econ 5(2):117–134
Avkiran NK (2001) Investigating technical and scale efficiencies of Australian universities through data envelopment analysis. Socio-Econ Plann Sci 35(1):57–80
Beasley J (1990) Comparing university departments. Omega Int J Manage Sci 18(2):171–183
Beasley J (1995) Determining teaching and research efficiencies. J Oper Res Soc 46:441–452
Bonaccorsi A, Daraio C (2004) Econometric approaches to the analysis of productivity of R&D systems: production functions and production frontiers. In: Moed HF, Glänzel W, Schmoch U (eds) Handbook of quantitative science and technology research: the use of publication and patent statistics in studies on S&T systems. Kluwer, Dordrecht, pp 51–74
Bonaccorsi A, Daraio C, Simar L (2006) Advanced indicators of productivity of universities: an application of robust nonparametric methods to Italian data. Scientometrics 66(2):389–410
Carrington R, Coelli T, Prasada Rao DS (2005) The performance of Australian universities: conceptual issues and preliminary results. Econ Pap 24(2):145–163
Caves DW, Christensen LR, Diewert WE (1982) The economic theory of index numbers and the measurement of input, output and productivity. Econometrica 50(6):1393–1414
Charnes A, Cooper W, Rhodes E (1978) Measuring the efficiency of decision making units. Eur J Oper Res 2(6):429–444
Coelli TJ, Prasada Rao DS, O'Donnell CJ, Battese GE (2005) An introduction to efficiency and productivity analysis, 2nd edn. Springer, New York
Daraio C, Simar L (2005) Introducing environmental variables in nonparametric frontier models: a probabilistic approach. J Prod Anal 24:93–121
Deem R, Brehony KJ (2005) Management as ideology: the case of 'new managerialism' in higher education. Oxf Rev Educ 31(2):217–235
Fandel G (2007) On the performance of universities in North Rhine-Westphalia, Germany: government's redistribution of funds judged using DEA efficiency measures. Eur J Oper Res 176:521–533
Farrell MJ (1957) The measurement of productive efficiency. J R Stat Soc A (Gen) 120:253–289
Flegg AT, Allen DO, Field K, Thurlow TW (2004) Measuring the efficiency of British universities: a multi-period data envelopment analysis. Educ Econ 12(3):231–249
Hazelkorn E (2011) Rankings and the reshaping of higher education: the battle for world-class excellence. Macmillan, Basingstoke
Johnes J (2004) Efficiency measurement. In: Johnes G, Johnes J (eds) The international handbook on the economics of education. Edward Elgar, Cheltenham, pp 613–742
Johnes J (2006) Data envelopment analysis and its application to the measurement of efficiency in higher education. Econ Educ Rev 25:273–288
Johnes J (2008) Efficiency and productivity change in the English higher education sector from 1996/97 to 2004/05. Manchester Sch 76(6):653–674
Johnes G, Johnes J (1993) Measuring the research performance of UK economics departments: an application of data envelopment analysis. Oxf Econ Pap 45:332–347
Joumady O, Ris C (2005) Performance in European higher education: a non-parametric production frontier approach. Educ Econ 12(2):189–205
Kempkes G, Pohl C (2010) The efficiency of German universities: some evidence from nonparametric and parametric methods. Appl Econ 42:2063–2079
Madden G, Savage S, Kemp S (1997) Measuring public sector efficiency: a study of economic departments in Australian universities. Educ Econ 5(2):153–168
Ollsen M, Peters MA (2005) Neoliberalism, higher education and the knowledge economy: from the free market to knowledge capitalism. J Educ Policy 20(3):313–345
Simar L, Wilson PW (1998) Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric frontier models. Manage Sci 44(1):49–61
Tomkins C, Green R (1988) An experiment in the use of data envelopment analysis for evaluating the efficiency of UK university departments of accounting. Financ Account Manag 4(2): 147–164

Warning S (2004) Performance differences in German higher education: empirical analysis of strategic groups. Rev Ind Organ 24:393–408

Wolszczak-Derlacz J, Parteka A (2011) Efficiency of European public higher education institutions: a two-stage multicountry approach. Scientometrics 89:887–917

Worthington A, Lee BL (2008) Efficiency, technology and productivity change in Australian universities, 1998-2003. Econ Educ Rev 27(3):285–298

Part III
Incentives: Monetary or Non-Monetary? Extrinsic or Intrinsic?

Academic Scientists and Knowledge Commercialization: Self-Determination and Diverse Motivations

Alice Lam

Abstract This chapter draws on self-determination theory to explain the mix of pecuniary and non-pecuniary motivational drivers underlying academic scientists' commercial pursuits. It examines the diversity of their personal motivations for knowledge commercialization and how this is influenced by their values and beliefs about the science-business relationship. It argues that scientists can be extrinsically or intrinsically motivated to different degrees in their pursuit of knowledge commercialization, depending how far they have internalized the values associated with it. Beyond reputational and financial rewards, intrinsic motivations (e.g., pro-social norms and hedonic motivation) are also powerful drivers of commercial engagement. The conventional assumption that scientists are motivated by reputational rewards and the intrinsic satisfaction of puzzle-solving in academic research while commercial engagement is driven primarily by the pursuit of the financial rewards builds on a false dichotomy and polarized view of human motivation.

1 Introduction

The rise of the entrepreneurial university (Clark 1998; Etzkowitz et al. 2000) with its emphasis on knowledge commercialization has generated an intense debate about the changing work norms of university scientists (Slaughter and Leslie 1997; Vallas and Kleinman 2008). Central to this is a growing concern that academics are becoming captured by the ethos of commercialism and the growth of a 'for-profit' motive among the new school entrepreneurial scientist (Jacob 2003; Slaughter and Rhoades 2004). While there is ample evidence of increased academic engagement in commercial activities such as patenting and spin-off company formation (D'Este and Patel 2007; Siegel et al. 2007), what remains unclear is whether or not this reflects the growth of a uniform category of entrepreneurial scientists driven by a common motive. Empirical research on the impact of financial incentives on academic scientists' propensity to engage in commercialization has

A. Lam (✉)
School of Management, Royal Holloway University of London, Egham, UK
e-mail: alice.lam@rhul.ac.uk

I.M. Welpe et al. (eds.), *Incentives and Performance*,
DOI 10.1007/978-3-319-09785-5_11

provided mixed evidence. While some studies find a positive link between financial incentives and the motivations of inventors to patent (Lach and Schankerman 2008; Thursby et al. 2001; Owen-Smith and Powell 2001a), others conclude that monetary rewards offered by universities play no role (Colyvas et al. 2002; Markman et al. 2004). Much of this work, however, has adopted a narrow conception of human motivation based on an economic model and is concerned primarily with the effects of financial incentives on the behavior of academic institutions rather than individual scientists. More recently, some authors have explored the personal motives behind scientists' transition to academic entrepreneurialism (Göktepe-Hulten and Mahagaonkar 2010; Goethner et al. 2012) but none has given adequate attention to the intrinsic aspect of their motivation (Deci 1975; Lindenberg 2001). Moreover, the question of how scientists' motives for commercial engagement vary according to their values and beliefs remains little understood.

This chapter draws on self-determination theory (Ryan and Deci 2000; Gagne and Deci 2005; Deci and Ryan 2000) to explain the mix of pecuniary and non-pecuniary motivational drivers underlying scientists' commercial pursuits. It examines the diversity of their personal motivations for knowledge commercialization and how this is influenced by their values and beliefs about the science-business relationship. It argues that scientists can be extrinsically or intrinsically motivated to different degrees in their pursuit of knowledge commercialization, depending on how far they have internalized the values associated with it. Beyond reputational and financial rewards, intrinsic motivations (e.g., pro-social norms and hedonic motivation) are also powerful drivers of commercial engagement.

2 Background and Theory

2.1 The Entrepreneurial University and Sociological Ambivalence of Scientists

In the UK and also elsewhere, government science and technology policy over the past two decades has sought to exploit the scientific knowledge base for innovation and economic competitiveness by promoting stronger collaboration between university and industry, and stimulating academic entrepreneurship (Lambert 2003; OECD 2003). At the same time, universities themselves have become willing actors in a range of markets in response to growing constraints on public funding and to adapt to a more competitive environment (Slaughter and Leslie 1997; Henkel 2007). Many are experimenting with new modes of governance and institutional practices to engage in commercial exploitation of research. The institutional transformation associated with the entrepreneurial academic paradigm has broadened the acceptable roles of academic researchers to accommodate commercial engagement (e.g., patenting, licensing and spin-off company formation) (Stuart and Ding 2006; Bercovitz and Feldman 2008). In parallel with the traditional 'Mertonian' model that emphasizes disinterested research, an alternative model of academic entrepreneurialism that encourages commercial exploitation of research has gained prominence in recent years. In the face of a normative duality that now governs university research, scientists may find themselves torn between the traditional Mertonian ideals of basic science and the reality of an encroaching market-oriented logic. While some may seek to resolve the tension by making choices between the dichotomous alternatives, others may attempt to reconcile the polar positions by re-negotiating their roles at the intersection of the two domains (Owen-Smith and Powell 2001b; Lam 2010; Murray 2010). Scientists' engagement in commercial activities needs to be interpreted within this shifting institutional context in which individual action often reflects the contradiction experienced rather than necessarily signaling unequivocal acceptance of a particular set of norms or values.

Early research in the sociology of science drew attention to the 'sociological ambivalence' of scientists and the frequent deviation of their actual behavior from the Mertonian norms (Mitroff 1974; Merton and Barber 1963). Gieryn (1983), for example, coined the term 'boundary work' to denote the active agency role of scientists in drawing and redrawing the boundaries of their work in pursuit of professional autonomy and increased resources. He stressed the power of scientists' interpretative strategies in constructing a space for science for 'strategic practical action'. His analysis showed that the boundary between the production of knowledge and its exploitation was clearly established when the scientific community wanted to protect its professional autonomy. However, it often became obscure, if not dissolved, when scientists sought to secure increased resources and public support for scientific research. This ambiguity is a source of internal tension, as well as giving scientists much opportunity for choice and variation. The contemporary transformation in the relationship between science and business has brought the sociological ambivalence of science to the forefront and opened up new opportunities for individual action.

Recognizing this marked ambivalence in scientific work is essential for understanding the complex relationship between values and behaviors because it implies that scientists' adherence to traditional 'Mertonian' norms does not preclude involvement in commercial activities and commercial engagement does not necessarily signal their acceptance of its underlying (market) ethos.

The same outward behavior of commercial engagement may be underpinned by diverse attitudes and motives. Shinn and Lamy (2006), for example, identify three categories of academic researchers who pursue divergent paths of commercialization: the 'academics' are those who weakly identify with the firm but may create a business venture for instrumental reasons; the 'pioneers' are driven by economic as well as scientific considerations; and the 'janus' are the hybrid type driven primarily by their passion for scientific knowledge production. In similar vein, Lam (2010) finds that academic scientists adopt different value orientations towards commercial science. Whilst traditionally-oriented scientists believe that science and business should remain distinct, entrepreneurially-oriented scientists fuse academic and entrepreneurial role identities and embed commercial practices in their work routines. Between the 'traditional' and 'entrepreneurial', some maintain a 'hybrid' orientation which involves the creative management of a fuzzy boundary between science and business. Scientists who participate in commercial activities do not constitute a homogeneous category. They may be driven by diverse and mixed motives and hence the need to adopt a differentiated approach for understanding their motivations.

2.2 *The Social Psychology of Human Motivation: Self-Determination Theory*

The work of social psychologists on motivation, notably self-determination theory (SDT) (Ryan and Deci 2000; Deci and Ryan 2000; Gagne and Deci 2005), provides a useful lens for examining the multifaceted nature of human motivation and its relationship with social values and norms. It treats motivation as the outcome of interaction between external regulatory processes and individuals' internal psychological needs for autonomy and self-determination. Taking the view that people are moved to act when they believe that the behaviors will lead to desired outcomes, SDT differentiates the content of outcomes and the regulatory processes through which they are pursued and thus predicting variation in the motivation underlying behaviors. Moreover, its emphasis on self-regulation in the motivational process is particularly germane to the case of academics who enjoy considerable freedom in their work.

SDT distinguishes three main states: intrinsic motivation, extrinsic motivation and amotivation. Intrinsic motivation refers to doing something for its inherent pleasure and satisfaction, whereas extrinsic motivation refers to doing something for a separable outcome or external rewards. Amotivation means having no intention to act because of lack of interest or not valuing the activity (Ryan 1995). SDT posits that an individual's motivation for behavior can be placed on a continuum of self-determination. It ranges from amotivation, which is wholly lacking in self-determination to intrinsic motivation which is an archetypal self-determined behavior because it arises from the individual's spontaneous interest rather than driven by external control. Between the two poles, extrinsic motivation can vary in its degree of self-determination, ranging from behavior that is fully externally regulated to one that is partially or fully internally integrated which approximates intrinsic motivation. Central to SDT is the argument that extrinsically motivated behavior can be transformed into intrinsically motivated one as individuals internalize the values and behavioral regulation that underlies it. When this occurs, the behavior becomes autonomous and no longer requires the presence of an external reward.

Building on the basic tenet that human beings have innate psychological needs for autonomy, SDT sees internalization as "an active, natural process in which individuals attempt to transform socially sanctioned mores or requests into personally endorsed values and self-regulations" (Deci and Ryan 2000, p. 235). As such, SDT stresses individual agency in the internalization process in that it is not just something that the socializing environment does to individuals but also represents the means through which individuals actively assimilate and reconstitute external regulations into inner values so that the individuals can be self-determined while enacting them (Ryan 1993). SDT identifies three distinct processes of internalization: introjection, identification and integration, which represent different degrees or forms of regulation associated with the different motivational types (mix) on the continuum of extrinsic-intrinsic motivation (Deci and Ryan 2000). *Introjection* occurs when individuals partially take in an external regulation but do not accept it as their own and therefore the behavior is not congruent with their values and is not self-determined: it is a partially 'controlled' activity and is predominately extrinsically motivated. *Identification* occurs when individuals identify with the value of behavior for their own self-selected goals and they experience greater freedom and volition because the behavior is more congruent with their personal goals and identities. Identification makes the behavior more autonomous and moves it towards the intrinsic end of the continuum. The most complete form of internalization is *integration* which occurs when individuals completely identify with the importance of social regulations or values, assimilate them into their sense of self and accept them as their own. As the behavior becomes fully congruent with the individuals' values and identity, they can be intrinsically motivated by it in the absence of an external regulation.

By focusing on the variation in the level of internalization and its relationship with the extrinsic and intrinsic aspects of motivation, SDT suggests that there are different kinds and degrees of motivation between the two polar types. There are three broad categories of outcomes associated with the different types of motivation: material, social and affective. While material outcomes are primarily related to extrinsic motivation, affective outcomes are closely related to intrinsic motivation. Social outcomes, however, are related to the 'in-between' types of motivation such as introjection (to fit in or feel worthy) and identification (to act appropriately). In contrast to the canonical economic model of human motivation and behavior which stresses the efficacy of extrinsic financial rewards, social psychologists argue that social and affective outcomes are equally salient. In fact, by postulating that human beings have a general organismic tendency towards self-regulation (Ryan 1995), social psychologists stress the potency of intrinsic motivation in driving behaviors. Although the concept of intrinsic motivation is often linked to affective outcomes, it has recently been broadened to incorporate a social, normative dimension (Grant 2008). Lindenberg (2001), for example, makes a distinction between enjoyment-based and obligation-based intrinsic motivation. The former is tied to the emotion for the improvement of one's condition whereas the latter refers to the satisfaction derived from acting according to a rule, norm or principle. In both cases, the motivation driving the behavior can be said to be intrinsic because it arises in the absence of material rewards or external constraints. It should be noted, however, that behaviors often lead to a combination of different outcomes. An intrinsic interest in the activity does not necessarily rule out the salience of extrinsic rewards insofar as the perceived locus of causality of the activity lies within the individual (Deci 1975). Amabile et al. (1994) argue that that some highly autonomous individuals (e.g., creative artists or scientists) may be strongly intrinsically interested in the activity and, at the same time, be strongly motivated to acquire extrinsic rewards (e.g., recognition, careers and money) for that activity.

3 Scientific Motivation and Knowledge Commercialization: 'Ribbon', 'Gold' and 'Puzzle'

Scientists may be motivated by a complex array of pecuniary and non-pecuniary factors in their commercial pursuits. A characteristic feature of the scientific reward system is its multidimensional nature, comprising the three components of the 'ribbon' (reputational/career rewards), 'gold' (financial rewards) and 'puzzle' (intrinsic satisfaction) (Stephan and Levin 1992). In the Mertonian world of academic science, the ribbon is the most substantial part of scientists' rewards. This is not only because scientists are strongly motivated by the recognition and prestige bestowed by their professional peers but also, other rewards such as salary and research funds are usually graduated in accordance with the degree of recognition achieved (Stephan 1996; Mulkay and Turner 1971). The ribbon is a deeply institutionalized feature in the scientific reward system and scientists feel the effects of the drive (Hagstrom 1974; Hong and Walsh 2009). Within the traditional model, publication is the main currency in the exchange relationship for the ribbon. The growing influence of the entrepreneurial paradigm has been subtly changing the ribbon exchange relationship to incorporate certain forms of commercial science. Several authors point out that contemporary academics can use patents as an alternative currency for building cycles of credit for obtaining resources to further the traditional rewards (Murray 2010; Owen-Smith 2003). Others suggest that the increasing reputational returns associated with patenting may prompt some scientists to use commercial activities as a means to further their academic careers (Krabel and Mueller 2009).

Although personal pecuniary gain, the 'gold', is also a component of the scientific reward system, it is predominately a consequence of the ribbon, and not a direct incentive for research in academic science (Stephan 1996). The rise of entrepreneurial science may well have opened up opportunities for scientists to reap financial rewards from commercial activities. It is, however, not entirely clear whether, and to what extent, the 'gold' is a motivational driver in the first place. There is a longstanding controversy in motivation theory about the role of money as a motivator (Sachau 2007). Herzberg's (1966) 'motivation–hygiene theory' and more recently authors in the field of positive psychology (Seligman and Csikszentmihalyi 2000) argue that money is a hygiene factor, not a motivator. It contributes to individual satisfaction or dissatisfaction but may not have the power to motivate on its own.

Beyond the extrinsic rewards of the ribbon and the gold, the majority of academic scientists are intrinsically motivated to advance knowledge, and they also derive immense satisfaction from engaging in challenging and creative activities. Indeed, the desire to engage in creative puzzle solving is the hallmark of a dedicated scientist (Eiduson 1962; Cotgrove 1970). In the Mertonian world of basic science, scientists derive satisfaction and enjoyment from seeking and finding vital truths within a relatively bounded scientific community. According to this view, there is no reason why the pursuit of creativity and puzzle solving should not take place in the context of an orientation towards knowledge application and entrepreneurial engagement.

The different motivational drivers can co-exist and scientists may be extrinsically or intrinsically motivated to different degrees in their pursuit of commercialization. The university is a professional bureaucracy where academics are accorded a relatively high degree of autonomy and they can choose whether to engage with industry. Few would be doing it as a result of external compulsion but the individuals' sense of pressure or willingness to participate in commercial activities may vary according to their beliefs about the values and potential benefits of such activities. In a study of academic scientists from UK research universities working in the fields of biological sciences, medicine, computer science and engineering, and physical sciences, Lam (2011) shows the relationships between their value orientations and motivations for engaging in knowledge commercialization.[1] Her analysis identifies three categories of 'commercializers' driven by different mix of motives: (a) the traditional scientists as 'reluctant' commercializers driven predominantly by the 'ribbon'; (b) the 'entrepreneurial' scientists as 'committed' commercializers motivated by the 'gold' as well as the 'puzzle'; and (c) the 'hybrid' scientists as 'strategic' commercializers motivated by the 'ribbon' and the 'puzzle'. Their characteristics are summarized below.

3.1 The Traditional Scientists as 'Reluctant' Commercializers: The 'Ribbon'

The 'traditional' scientists are those who adhere to the Mertonian norms of basic science and perceive commercialization to be at odds with their personal values and goals. While the majority of this category defies the growing pressures for knowledge commercialization, some are prepared to experiment with commercial practices in anticipation of possible benefits. They are 'reluctant' commercializers who pursue commercial activities in order to obtain the much needed funding for research in an increasingly resource constrained environment. The reply of a professor, who took part in Lam's study, to the question of what drove his group to form a spin-off company is illustrative: "*We just wanted to test our ideas. We*

[1] The study was based on 36 in-depth individual interviews and an on-line questionnaire survey of 735 from five major UK research universities. For further details of the quantitative results, see Lam (2011).

were desperate to get funding... Well, so none of us are born again entrepreneurs. We were driven by the idea we wanted to do this research and to use it". Others recognize that commercial engagement has gained increased institutional legitimacy and are motivated by the possibility of using it as a currency for building scientific credit and enhancing their careers. For example, one traditional scientist interviewed in the study pointed out that commercial engagement was "*a risk worth taking*' because "*it was the culture of the department*' and that "*if you were going to be a top academic, that's one of the things that you had to cover"*.

The 'traditionalists' are motivated primarily by the 'ribbon'. To these scientists, knowledge commercialization represents an 'introjected' and extrinsically motivated behavior. Introjection, according to social psychologists (Koestner et al. 1996), is associated with emotional incongruence and ambiguity and is a relatively unstable form of regulation. The position of the traditionalists is somewhat indeterminate and they may change directions based on evaluations of the success or failures of their trial efforts. For example, several of the traditional scientists interviewed in the study mentioned how their own attitudes and the 'culture' of their Departments had shifted from away from the entrepreneurial pull towards more a basic research orientation as a result of the unsuccessful ventures. Some expressed regret at their commercial involvement because of the diversion of time and resources away from fundamental research. These accounts suggest that scientists' transition to academic entrepreneurialism is not necessarily a linear process but can be halted or even reversed when commercial engagement proved to be of limited value for furthering their quest for the ribbon.

3.2 The Entrepreneurial Scientists as 'Committed' Commercializers: The 'Gold' and 'Puzzle'

In contrast, the 'entrepreneurial' scientists are those who believe in the fundamental importance of science-business collaboration for knowledge application and exploitation. These scientists often pursue commercial activities volitionally out of a sense of personal commitment or interest because they have fully *integrated* the norm of entrepreneurialism. They are motivated to do what is actually in their own interest and the desired outcomes could be both material and affective. 'Cashing in' on their scientific expertise, in this case, is seen as a legitimate means of topping up their academic salaries and also an achievement. While openly acknowledging the importance of the gold, the majority interviewed in the study talked about the money reward in a somewhat negative manner in that it was portrayed as a source of discontent, what Herzberg (1966) refers to as a 'hygiene' factor, rather than a positive motivator. Complaints about being underpaid and lagging academic salaries permeate the conversations in the interviews when the scientists responded to the question about the money incentive. For example, one company founder said, "*the university pays absolute peanuts and therefore you'd be totally mad not to do it if you are in an area where you can do it*".

Academic scientists, like everyone else, need to earn a decent living. Some may well be 'cashing in' on their scientific expertise. However, what is clear from the evidence presented in the study is that even for the apparently most market-oriented entrepreneurial scientists, the gold appears to be only one of the motivational drivers underlying their commercial endeavors. Besides the gold, these entrepreneurial scientists also derive immense psychic satisfaction from taking part in commercial ventures. To some, the sense of achievement that they experienced in starting up a business venture is no less intense than the satisfaction of solving a scientific puzzle. The interviews reveal the salience of an enjoyment-based (hedonic) intrinsic motivation (Lindenberg 2001) among the entrepreneurial scientists. Many used words such as 'excitement', 'fun' and 'thrill' to describe their experience. The entrepreneurial scientists are 'committed' commercializers who are driven by what Shane et al (2003) refer to as an 'egoistic passion' for achievement. This manifests in their love for the activity as well as the fortune that may come along with it. They have autonomous reasons for pursuing the puzzle as well as the gold, and external regulation may have limited effect on their behavior. They can be placed at the intrinsic end of the motivational continuum.

3.3 The Hybrid Scientists as 'Strategic' Commercializers: The 'Ribbon' and 'Puzzle'

Between the 'traditional' and 'entrepreneurial', there are the 'hybrids' who share the entrepreneurial scientists' belief about the importance of science-business collaboration, while maintaining the traditionalists' commitment to the core scientific values. Unlike the traditionalists, the hybrids participate in commercial activities more autonomously, supported by feelings of identification. They perceive commercial activities as largely legitimate and desirable for their scientific pursuits. Forming a spin-off company, for example, is seen as a way of maintaining their scientific autonomy and asserting control over the knowledge exploitation process. They seek *identification* with commercial activity by reconstituting its meaning so that it becomes more congruent with their professional goals and values. Besides obtaining funding to support their research, the hybrid scientists are also strongly intrinsically motivated in their commercial activities. The majority interviewed in the study believed in the positive benefits of knowledge application (e.g., testing new ideas) and saw commercialization as an extension of their knowledge search activities. Many emphasized 'the challenge' of solving complex industrial problems. To these scientists, knowledge application through taking part in commercial ventures represents a kind of puzzle-solving activity that satisfies their intellectual curiosity. The assumption that scientists derive the pleasure of puzzle-solving only from basic research is based on a narrow conception of the full range of creative scientific activity in their work.

The hybrid scientists' personal interest in knowledge application also bolsters a strong professional conviction to make their knowledge socially relevant. This is particularly notable among those researching in the life-sciences. The following reply of a professor in biosciences to the question of why forming a spin-off company is indicative: "*I think we as academics have a responsibility to the nation really, we're in a very privileged position. And our money comes from the state or from charities*". Another biomedical professor made a similar comment: "*we wanted to see if, you know, there was a potential new drug there for, you know, treating people who can't get treated with anything else*". Grant (2008) argues that personal interest in an activity can reinforce pro-social intrinsic motivation which is a particular form of intrinsic motivation based on a desire to help others resulting from identified regulation. The hybrid scientists can be described as 'strategic' commercializers in that they incorporate commercial practices into their repertoire of behavior without sacrificing their focal academic identity. They can be placed mid-point on the motivational continuum—they are extrinsically motivated ('ribbon') somewhat while at the same time intrinsically motivated ('puzzle') in their commercial pursuits.

Figure 1 summarizes the main points presented above in a conceptual model.

The model postulates that commercial engagement can be either a 'controlled' or an 'autonomous' activity depending how far scientists have 'internalized' the values associated with it. Scientists who adhere strongly to the traditional Mertonian norms of basic science are likely to be *amotivated*. However, some may take part in the activity as a result of *introjection*. These 'traditional-oriented' scientists use commercialization as a means to obtain resources (e.g., research funding) to support their pursuit for the ribbon, and are placed at the extrinsic end of the motivational continuum. In contrast, other scientists pursue the activity volitionally because they have fully *integrated* the norm of entrepreneurialism and the desired outcomes are both affective and material. The financial returns in this case represent both achievement and profit. This type of 'entrepreneurial-oriented' scientists is placed at the intrinsic end of the continuum. Between the polar opposites, some scientists adopt a 'hybrid' position encompassing characteristics of the 'traditional' and 'entrepreneurial'. Their transition to academic entrepreneurialism typically involves crafting a hybrid role identity which 'overlays elements of a commercial orientation onto an academic one' (Jain et al. 2009, p. 927). This is similar to the process of *identification* described in SDT through which people identify with the value and importance of behavior for their self-selected values (Ryan and Deci 2000). The hybrids occupy the mid-position on the extrinsic and intrinsic motivational continuum.

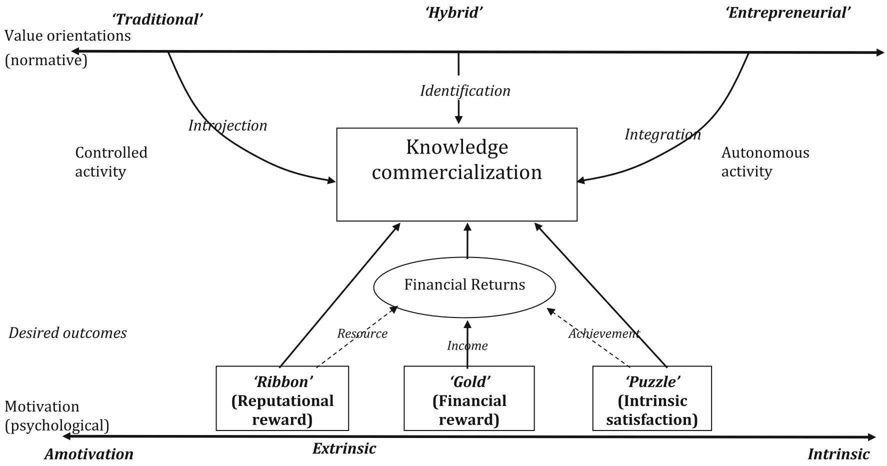


Fig. 1 Scientific motivation and knowledge commercialization: a conceptual model

Discussion and Conclusion

The assumption that scientists are motivated by reputational rewards and the intrinsic satisfaction of puzzle-solving in academic research while commercial engagement is driven primarily by the pursuit of financial rewards builds on a false dichotomy and polarized view of human motivation. This chapter highlights the diverse motives driving the commercial behavior of academic scientists and the primacy of their self-motivation. Intrinsic motivation has long been recognized by social psychologists as a pervasive and powerful driver of human action but is neglected in much of the research on scientists' transition to academic entrepreneurialism. This chapter argues that a fuller explanation of scientists' engagement in knowledge commercialization will need to consider a broader mix of motives, beyond the narrow confines of extrinsic rewards, to include the social and affective aspects related to intrinsic motivation. Scientists, like other professionals, have the desire to expend effort to benefit others and society in the context of both academic and entrepreneurial science. This 'pro-social' motivation is a specific form of intrinsic motivation (Grant 2008). Moreover, having fun or the joy of achievement is at the heart of Lindenberg's (2001) idea of 'enjoyment-based' intrinsic motivation. The idea that fun of play is an important motivation underpinning creative and inventive behavior is a longstanding one. Taussig, in his book *Inventors and Money Makers* (1930), argues that the 'instinct of contrivance', namely the desire to create and recombine things, is a powerful motivational underpinning of human inventive behavior. Other studies also show the importance of intrinsic enjoyment as the creative driver of inventors and researchers alike (Rossman 1931; Loewenstein 1994).

(continued)

Gustin (1973) argues that creative scientists are motivated to do charismatic things and that science is charismatic because of its puzzle-solving nature. One might argue that for some scientists, commercial engagement represents a kind of puzzle that satisfies their desire for pursuing ‘charismatic’ activities.

Academic scientists have diverse motives for pursuing knowledge commercialization and there is no one single type of entrepreneurial scientists driven by a common motive. The analysis in this chapter suggests that academics’ value preferences influence their motives for commercial engagement and the relative importance of reputational, financial and affective rewards as desired outcomes. It draws attention to internalization of values and external regulation as key factors differentiating the types of motives driving the commercial behavior of scientists. Values are not fixed but may evolve over time which could result in changes in motives and behaviors. SDT argues that human beings have an organismic tendency towards autonomy and self-regulation in their behavior (Ryan 1995; Deci and Ryan 2000). However, this by no means suggests that an introjected behavior will gradually become identified or integrated over time. Normative change, more often than not, involves the paradoxical combination of opposing values in an ambivalent manner (Colyvas and Powell 2006; Owen-Smith and Powell 2001b; Murray 2010). The preponderance of the ‘hybrid’ category in Lam’s study (2011) is a case in point. The salience of enjoyment-based intrinsic motivation among the ‘entrepreneurial’ scientists illustrates the primacy of self-motivation rather than external regulation in driving their commercial behavior.

Policies designed to promote research commercialization often assume that academics respond to financial incentives tied to successful exploitation of their knowledge. However, if academics are motivated by a complex mix of extrinsic and intrinsic rewards, then policy initiatives focusing narrowly on providing financial rewards might be inadequate or even misplaced. Motivation theorists (Frey 1997; Frey and Jegen 2001) also warn of the troubling possibility that management practices based on economic models could undermine intrinsic motivation by transforming a non-monetary relationship into an explicitly monetary one. Moreover, given the diverse values and motives underlying scientists’ commercial activities, it is unlikely that an undifferentiated approach will be effective in eliciting the requisite effort across the board. Some authors (Krabel and Mueller 2009; Hoye and Pries 2009) propose that policies to facilitate academic entrepreneurialism should target the subpopulation of academic researchers with commercialization-friendly attitudes such as the ‘habitual entrepreneurs’. The analysis presented in this chapter suggests that additional incentives may have only limited effect on those who are already deeply engaged in the activity as in the case of the ‘entrepreneurial’ scientists. These scientists have autonomous

(continued)

reasons for pursuing commercial science and they may follow what is professionally challenging and personally interesting rather than anything else. On the contrary, it seems that it is the 'traditionalists' who would be amenable to behavioral change in response to external rewards linked to the ribbon. In particular, rewards in the form of additional funding for research and ascription of academic status to commercial success may have high motivating power for inducing some traditionalists to go down the commercial path. These ribbon-related rewards may also reinforce the commercial behavior of the 'hybrids' and strengthen their perception of the positive benefits of the activity. There is, however, always a potential danger that top-down engineering of knowledge commercialization may undermine scientists' sense of self-determination and the intrinsic, puzzle-solving aspect of their motivation which is the ultimate driver of creativity. While intrinsic motivation cannot be enforced, it can be enabled through socialization and competence enhancing provisions to strengthen feelings of autonomy and the culture of creativity and prosocial motivation (Osterloh 2006).

Finally, it is worthy of note that the analysis presented in this chapter has focused on the relationship between the motivating factors and scientists' engagement in knowledge commercialization in terms of their behavioral choice (that is, whether they had been involved in any commercial activities). The motivational implications for behavioral intensity or persistence have not been examined. Future research might include indicators such as the amount of time spent on the activity and duration of engagement to capture behavioral intensity and persistence. Relatedly, the question of how scientists' value orientations and motives might change over the course of their careers, and the possible causes also merit further research.

References

Amabile TM, Hill KG, Hennessey BA, Tighe EM (1994) The work preference inventory: assessing intrinsic and extrinsic motivational orientations. J Pers Soc Psychol 66(5):750–967

Bercovitz J, Feldman M (2008) Academic entrepreneurs: organizational change at the individual level. Organ Sci 19(1):69–89

Clark BR (1998) Creating entrepreneurial universities. Pergamon, Oxford

Colyvas J, Powell WW (2006) Roads to institutionalization. In: Staw BM (ed) Research in organizational behavior. Elsevier, New York

Colyvas J, Crow M, Gelijns A, Mazzoleni R, Nelson RR, Rosenberg N, Sampat BN (2002) How do university inventions get into practice? Manag Sci 48(1):61–72

Cotgrove S (1970) The sociology of science and technology. Br J Soc 21(1):1–15

Deci EL (1975) Intrinsic motivation. Plenum, New York

Deci EL, Ryan RM (2000) The "what" and "why" of goal pursuits: human needs and the self-determination behavior. Psychol Inq 11(4):227–268

D'Este P, Patel P (2007) University-industry linkages in the UK: what are the factors underlying the variety of interactions with industry? Res Policy 36(9):1295–1313

Eiduson BT (1962) Scientists: their psychological world. Basic Books, New York

Etzkowitz H, Webster A, Gebhardt C, Terra B (2000) The future of university and the university of the future: evolution of ivory tower to entrepreneurial paradigm. Res Policy 29(2):313–330
Frey BS (1997) Not just for the money: an economic theory of personal motivation. Edward Elgar, Cheltenham
Frey BS, Jegen R (2001) Motivation crowding theory. J Econ Surv 15(5):589–611
Gagne M, Deci EL (2005) Self-determination theory and work motivation. J Organ Behav 26 (4):331–362
Gieryn T (1983) Boundary-work and the demarcation of science from non-science: strains and interests in professional ideologies of scientists. Am Sociol Rev 48(6):781–795
Goethner M, Obschonka M, Silbereisen RK, Cantner U (2012) Scientists' transition to academic entrepreneurship: economic and psychological determinants. J Econ Psychol 33(3):628–641
Göktepe-Hulten D, Mahagaonkar P (2010) Inventing and patenting activities of scientists: in the expectation of money or reputation? J Technol Transf 35(4):401–423
Grant AM (2008) Does intrinsic motivation fuel the prosocial fire? Motivational synergy in predicting persistence, performance, and productivity. J Appl Psychol 93(1):48–58
Gustin BH (1973) Charisma, recognition and the motivation of scientists. Am J Sociol 78 (5):1119–1133
Hagstrom W (1974) Competition in science. Am Sociol Rev 39(1):1–18
Henkel M (2007) Can academic autonomy survive in the knowledge society? A perspective from Britain. Higher Educ Res Dev 26(1):87–99
Herzberg F (1966) Work and the nature of man. World Publishing, Cleveland
Hong W, Walsh JP (2009) For money or glory? Commercialization, competition, and secrecy in the entrepreneurial university. Soc Q 50(1):145–171
Hoye K, Pries F (2009) Repeat commercializers, 'the habitual entrepreneurs' of university-industry technology transfer. Technovation 29(10):682–689
Jacob M (2003) Rethinking science and commodifying knowledge. Policy Future Educ 1(1):125–142
Jain S, George G, Maltarich M (2009) Academics or entrepreneurs? Investigating role identity modification of university scientists involved in commercialization activity. Res Policy 38 (6):922–935
Koestner R, Losier GF, Vallerand RJ, Carducci D (1996) Identified and introjected forms of political internalization: extending self-determination theory. J Pers Soc Psychol 70(5):1025–1036
Krabel S, Mueller P (2009) What drives scientists to start their own company? An empirical investigation of Max Planck society scientists. Res Policy 38(6):947–956
Lach S, Schankerman M (2008) Incentives and invention in universities. RAND J Econ 39 (2):403–433
Lam A (2010) From 'ivory tower traditionalists' to 'entrepreneurial scientists'? Academic scientists in fuzzy university-industry boundaries. Soc Stud Sci 40(2):307–340
Lam A (2011) What motivates academic scientists to engage in research commercialization: 'gold', 'ribbon' or 'puzzle'? Res Policy 40(10):1354–1368
Lambert R (2003) Lambert review of business-university collaboration. HMSO, London
Lindenberg S (2001) Intrinsic motivation in a new light. Kyklos 54(2–3):317–342
Loewenstein G (1994) The psychology of creativity: a review and reinterpretation. Psychol Bull 116(1):75–98
Markman GD, Gianiodis PT, Phan PH, Balkin DB (2004) Entrepreneurship from the ivory tower: do incentive systems matter? J Technol Transfer 29(3–4):353–364
Merton RK, Barber E (1963) Sociological ambivalence. In: Tiryakian EA (ed) Sociological theory, values and sociocultural changes. The Free Press, Glencoe
Mitroff II (1974) Norms and counter-norms in a select group of the Apollo Moon Scientists: a case study of the ambivalence of scientists. Am Sociol Rev 39(4):579–595
Mulkay MJ, Turner BS (1971) Over-production of personnel and innovation in three social settings. Sociology 5(1):47–61

Murray F (2010) The oncomouse that roared: hybrid exchange strategies as a source of distinction at the boundary of overlapping institutions1. Am J Sociol 116(2):341–388
OECD (2003) Turning science into business: patenting and licensing in public research. OECD, Paris
Osterloh M (2006) Human resources management and knowledge creation. In: Ichijo K, Nonaka I (eds) Knowledge creation and management. Oxford University Press, Oxford
Owen-Smith J (2003) From separate systems to a hybrid order: accumulative advantage across public and private science at research one universities. Res Policy 32(6):1081–1104
Owen-Smith J, Powell WW (2001a) To patent or not: faculty decisions and institutional success at technology transfer. J Technol Transfer 26:99–114
Owen-Smith J, Powell W (2001b) Careers and contradictions: faculty responses to the transformation of knowledge and its uses in the life sciences. Res Sociol Work 10:109–140
Rossman J (1931) The motives of inventors. Q J Econ 45(3):522–528
Ryan RM (1993) Agency and organization: intrinsic motivation, autonomy and the self in psychological development. In: Jacobs J (ed) Nebraska symposium on motivation: developmental perspectives on motivation. University of Nebraska Press, Lincoln
Ryan RM (1995) Psychological needs and the facilitation of integrative processes. J Pers 63 (3):397–427
Ryan RM, Deci EL (2000) Intrinsic and extrinsic motivations: classic definitions and new directions. Contemp Educ Psychol 25(1):54–67
Sachau DA (2007) Resurrecting the motivation-hygiene theory: Herzberg and the positive psychology movement. Hum Resour Dev Rev 6(4):377–393
Seligman ME, Csikszentmihalyi M (2000) Positive psychology: an introduction. Am Psychol 55 (1):5–14
Shane S, Locke EA, Collins CJ (2003) Entrepreneurial motivation. Hum Resour Manag Rev 13 (2):257–279
Shinn T, Lamy E (2006) Paths of commercial knowledge: forms and consequences of university-enterprise synergy in scientist-sponsored firms. Res Policy 35(10):1465–1476
Siegel DS, Wright M, Lockett A (2007) The rise of entrepreneurial activity at universities: organizational and societal implications. Ind Corp Chang 16(4):489–504
Slaughter S, Leslie L (1997) Academic capitalism: politics, policies and the entrepreneurial university. John Hopkins University Press, Baltimore
Slaughter S, Rhoades G (2004) Academic capitalism and the new economy: markets, state, and higher education. Johns Hopkins University Press, Baltimore
Stephan P (1996) The economics of science. J Econ Lit 34(3):1199–1235
Stephan P, Levin S (1992) Striking the mother lode in science: the importance of age, place and time. Oxford University Press, New York
Stuart TE, Ding WW (2006) When do scientists become entrepreneurs? The social structural antecedents of commercial activity in the academic life sciences. Am J Sociol 112(1):97–144
Taussig FW (1930) Inventors and money makers. Macmillan, New York
Thursby JG, Jensen R, Thursby MC (2001) Objectives, characteristics and outcomes of university licensing: a survey of major U.S. universities. J Technol Transfer 26(1–2):59–72
Vallas SP, Kleinman LD (2008) Contradiction, convergence and the knowledge economy: the confluence of academic and commercial biotechnology. Soc Econ Rev 6(2):283–311

Imaginary Contradictions of University Governance

Uwe Wilkesmann

Abstract New modes of managerial governance have caused universities to function more like companies and produce non-intended effects as well as imaginary contradictions. In this article, four of these contradictions are discussed to provide answers to the following research questions: Do professors have a higher commitment to their organization or to their peers in the scientific community? Which factors strengthen the affective organizational commitment? Which work environment supports intrinsic motivation at universities? Can universities provide incentives that do not crowd out intrinsic motivation? A theoretical underpinning of hypotheses will be provided, and these hypotheses will be tested using two nationwide surveys of German professors. The empirical results demonstrate that commitment to professional peers increases affective organizational commitment. In the perception of German professors, there is no contradiction between profession and organization, but the newly implemented steering instruments increase organizational commitment. In addition, the results also provide evidence that autonomy, relatedness, and perceived competence increase intrinsic teaching motivation. These findings support the Self-Determination Theory. The results also provide some evidence of a crowding-out effect of the new steering instruments and that teaching awards do not crowd out intrinsic motivation.

1 Introduction

In the last decade, the higher education system in Europe has shifted to New Public Management (NPM) and established new modes of governance (de Boer et al. 2007). These modes of managerial governance have caused universities to function more like companies, producing non-intended effects and non-intended contradictions of governance. In this article, four of these contradictions, which can be described as imaginary contradictions and are closely related to each other, are discussed. These contradictions are imaginary because an inherent solution exists

U. Wilkesmann (✉)
Center for Higher Education, TU Dortmund University, Dortmund, Germany
e-mail: uwe.wilkesmann@tu-dortmund.de

I.M. Welpe et al. (eds.), *Incentives and Performance*,
DOI 10.1007/978-3-319-09785-5_12

and only a first consideration characterizes them as contradictions. The following four contradictions are particularly relevant to higher education institutions in Germany and therefore discussed in this article:

Profession vs. organization: Professors are professionals who work in a loosely coupled system, and their scientific communities provide relevant career resources (Hüther and Krücken 2011). However, due to NPM reforms, universities are shifting in the direction of "complete organizations" (Ahrne and Brunsson 2011). Consequently, universities establish more principal agent relationships, replacing the influence of professions. Nevertheless, professionals must be organized in organizations, and organizations must manage professionals. The relevant research questions are the following: do professors have a higher commitment to their organization or to their peers in the scientific community? Which factors strengthen affective organizational commitment?

Monitoring vs. autonomy: Professors may perceive development toward NPM as a new form of monitoring and, most likely, as an alienating experience. New steering instruments are formal regulations that increase the distance between rectors and deans in the role of superiors and professors in the role of subordinates. However, it is difficult to monitor and manage scientific work according to formal rules. Professors need autonomy in their work (Frey et al. 2013; Osterloh and Frey 2013). How can monitoring and autonomy be balanced?

Intrinsic vs. extrinsic motivation: For academic work, intrinsic motivation is absolutely necessary (Lam 2011). Intrinsic motivation occurs only in work environments with a high degree of self-determination (Ryan and Deci 2000b). Selective incentives—such as NPM instruments at universities (see Sect. 4)—crowd out intrinsic motivation (Frey and Osterloh 2002; Osterloh and Frey 2013; Wilkesmann and Schmid 2014). However, some selective incentives are necessary for managing an organization. The relevant questions in this field are the following: which work environment supports intrinsic motivation at universities? Can universities provide incentives that do not crowd out intrinsic motivation?

Transactional vs. transformational governance: Selective incentives are an example of transactional governance, in which each behavior of a member is related to an organizational exchange. Transformational governance creates leeway for intellectual innovation and common visions. However, both transactional and transformational governance are necessary for a university (Wilkesmann 2013). How can these two types of governance function in concert?

In this paper, imaginary contradictions one and three will be proven using two surveys of German professors at research universities and professors at universities of applied sciences (empirical evidence for the other two contradictions can be found in Wilkesmann 2013).

2 Profession Versus Organization

Professors are all members of a profession; as physicians, sociologists, engineers, chemists or business economists, they belong to their specialist profession (Starbuck 1992). The word 'professionalism' refers "[. . .] to the institutional circumstances in which the members of occupations rather than consumers or managers control work. [. . .] While few, if any occupations can be said to *fully* control their own work, those that come close are called 'professions' in English" (Freidson 2001, p. 12). Professors feel more committed to their profession than to their organization. All feedback and all career-relevant evaluations (i.e., peer reviews of submitted articles or research proposals) are under the control of the profession and not the organization. The organization of universities before NPM did not enable domination over professors because universities did not monitor or support the careers of academic staff in the German higher education system (Hüther and Krücken 2015). This model is described, for example, in Mintzberg's professional bureaucracy or in models that characterize universities as "specific organizations" (Musselin 2006). NPM changes the power relationship of the organization. Under NPM, the rectorate can sanction professors via its ability to distribute or refuse resources. When a professor establishes a new research program or a new Master's program, he or she can be rewarded with additional research assistants. Performance-related budgets or Management by Objectives (MbO) are the new steering instruments that strengthen the hierarchy. Due to these new steering instruments, the power distance within the organization increases, shifting the university to a "complete organization" (Ahrne and Brunsson 2011). Ahrne and Brunsson (2011) define a complete organization as having the following five characteristics:

Membership: The organization university is defined by two different groups of members, academic staff and students. The former group is paid by the organization because they have an employment agreement. The latter group must pay the organization or is not paid by the organization. Their conditions of membership are comparable to those of a club (Wilkesmann et al. 2011).

Hierarchy: Formerly, the hierarchy was limited in a university, but NPM has strengthened the roles of rectorates, vice-chancellors, and deans. The new steering instruments (pay-for-performance, performance-related budgets and MbO) are all instruments that constitute a principal agent relationship. The superior (rectorate or dean) can use these instruments as selective incentives (Wilkesmann and Schmid 2012). In addition, in Germany, superiors gained more legal rights, thereby strengthening organizational roles.

Rules: Ahrne and Brunsson (2011) refer to rules as explicit decisions and not social norms that members follow implicitly. In Germany, professors at research universities must adhere to the following explicit rules: they must have a teaching load of 9 h/week and a budget that is related to explicit performance criteria, such as the amount of money collected for third-party-funded research projects (in €) and the number of published peer-reviewed articles.

Monitors: Compliance with the rules must be monitored. Teaching and research assessment measures include the amount of money collected from third parties, the number of publications, and the number of delivered classes and lectures.

Sanctions: The new steering instruments of NPM allow superiors to distribute resources as selective incentives. Granting or refusing monetary resources is a powerful sanction system. The rectorate can, for example, reward a professor with two more research associates to increase his/her research group or punish him/her by reducing the budget for the laboratory.

Thus, to varying extents, universities fulfill all five characteristics of a complete organization. These characteristics describe a formal organization. In summary, universities have shifted toward becoming complete organizations.

Does this shift imply that professors now have a higher commitment to the organization than to the profession? Meyer and Allen (1991) distinguish between three types of (organizational) commitment: affective, normative, and continuance commitment.

> "Affective commitment refers to the employee's emotional attachment to, identification with, and involvement in the organization. Employees with a strong affective commitment continue employment with the organization because they *want* to do so. Continuance commitment refers to an awareness of the costs associated with leaving the organization. Employees whose primary link to the organization is based on continuance commitment remain because they *need* to do so. Finally, normative commitment reflects a feeling of obligation to continue employment. Employees with a high level of normative commitment feel that they *ought* to remain with the organization" (Meyer and Allen 1991, p. 67).

Due to space limitations, we will focus on affective commitment in this chapter and provide empirical evidence for the following two research questions. Which factors strengthen or weaken affective organizational commitment? Do professors have a higher commitment to their organization or to their peers in the scientific community? The new steering instruments (pay-for-performance, performance-related budgets and MbO) result in a utilitarian calculation, much like that described by Barnard's (1938) theory: a member of an organization contributes to the organizational goals as long as his or her perception of the given inducement is greater than the costs of his or her contribution. This calculus reduces the affective commitment (but increases the continuance commitment) because the only cause for a behavior that is in line with the organizational goals is a monetary or other selective incentive. The new steering instruments established a difference between the principals who distributed the incentives and the agents who received the incentives. If professors are treated like agents, they behave (in the long run) like agents, i.e., continuance commitment increases and affective commitment decreases. We can summarize these findings in hypothesis 1:

H1 The new steering instruments reduce the affective organizational commitment of professors.

Barnard (1938) emphasized that organizations could not be efficient when all members only used the above calculus. Members must also fulfill an extra behavioral role (Matiaske and Weller 2003) linked to an inherent motivation based on internalized social norms. Professionals were socialized within such vocational norms. As mentioned above, German universities have little control over the careers of professors because the profession manages the peer-review process, e.g., scholarly peers review short lists for appointments, project applications, or articles submitted to journals (Hüther and Krücken 2015). Furthermore, the social norms that govern the behavior of professors are professional norms that are internalized during a long education process as a student and during the assistantship, during which each researcher learns what constitutes good research and teaching and scientific behavior. Although professors receive their resources and salary from the organization university, their behavior is governed by professional norms. Due to this socialization process, professors are highly committed to their peers in the scientific community. Therefore, we predict that professors have a low affective commitment to the organization and a high commitment to scholarly peers. We summarize this in hypothesis 2:

H2 The higher the commitment to the peers in the scientific community, the lower the affective organizational commitment.

The introduction of new steering instruments based on NPM results in the transfer of the principles of organizing a private company to a public organization, particularly universities. The NPM conflicts with the academic habitus (Bourdieu 1988), which emphasizes the freedom and autonomy of intellectual work. Therefore, many, particular older, professors who were socialized into the classic homo academicus have an attitude against NPM. Attitudes are independent from behavior because attitudes have no expensive or painful consequences, in contrast to behavior. However, attitudes guide behavior and can particularly govern affective commitment. Affective commitment is reduced when the organization of the university is changing in a direction opposite of the attitudes of professors.

H3 The higher the attitude against New Public Management, the lower the affective organizational commitment.

3 Monitoring Versus Autonomy

The second imaginary contradiction is much like the first one. When an organization shifts from a loose-coupled professional bureaucracy (Mintzberg 1989) to a hierarchical organization, such as a principal agent relationship, professors feel like agents. Due to performance measurement, assessment, and evaluation, professors may perceive themselves as monitored agents. The new steering instruments establish a new relationship between the superiors that allocate resources in response to performance indicators and the subordinates that receive these resources. The former relationship was one of dependency with unbalanced power. If every behavior of a professor was measured by performance criteria, they could perceive their relationship with the organization (in the long run) as alienated because they would no longer have "*full* control (over) their own work" (Freidson 2001, p. 12). This is in line with the Principal-Agent Theory (Eisenhardt 1989), in which superiors monitor and motivate agents with the help of selective incentives.

Many studies have analyzed the unintended effects of selective incentives in academia (Osterloh and Frey 2013). Frey et al. (2013) argued that only income control and not output control was suitable in the academic world because output reassurance and knowledge of cause-effect relationships are both low. However, if the academic world was monitored only with the help of output control, such as rankings, evaluations, and performance-related budgets, academics would have an incentive for "gaming the system" (van Thiel and Leeuw 2002).

Nevertheless, management requires measurement. Even in universities in which professionals work, the organization needs numbers to help assess the achievement of their objectives. To overcome the contradiction between monitoring and autonomy, numbers (from performance measurement, evaluations and rankings) should only be used for a collective reflection upon goal attainment. If numbers are not used as an ineluctable rule for distributing resources, they will not develop an 'independent existence' that ultimately results in an institution that is perceived as an alienated object. In organizational terms, personal contact and leadership are more relevant than performance-related rules. While numbers are relevant for the legitimacy of decisions, they also serve as an origin of organizational reflection. The organization should enable an "Initiative-Freeing Radical Organizational Form" (Carney and Getz 2009; Getz 2009), but to control the achievement of the collective agreed objectives, numbers are used as a reflection of development. Even collective decision-making needs legitimacy and an origin for underpinning arguments. Therefore, numbers are helpful. Perhaps, the handling of numbers and not numbers per se is what is important for their perception by professors. The main questions in this field are the following. Do numbers serve as instruments for self-governance of the organization or as instruments for punishment by a superior? Do the members have the freedom to influence their own behavior and the organizational objectives or do they perceive the organization as a strange institution? The perception of organizational autonomy or monitoring and punishment are also influencing factors for the motivation of the members.

4 Intrinsic Versus Extrinsic Motivation

Autonomy and monitoring are also closely related to work motivation. Intrinsic motivation is necessary for academic work. Their professional habitus motivates professors to pursue innovative, non-standardized work (see chapters two and three in Wilkesmann and Schmid (2014) for more details about the theoretical underpinning of the nexus between work environment and motivation). Traditionally, professors were considered highly intrinsically motivated because otherwise they would not endure the pressures and imponderables of accomplishing successful academic careers: “We may say that it is this intrinsic motivation which makes academics commit themselves to their scholarly activities not as a job but as a vocation, profession and hobby; which sustains them despite deteriorating working conditions and salaries” (Moses and Ramsden 1992, p. 105). As mentioned above, the main research question is the following: which work environment supports intrinsic motivation at universities? According to Self-Determination Theory (SDT) (Ryan and Deci 2000a, b), intrinsically motivated action encompasses any action that is performed for pure enjoyment and satisfaction. By contrast, if an action is accomplished for separable outcomes, the motivation is extrinsic (Ryan and Deci 2000a, p. 56; Ryan and Deci 2006, p. 1562). Intrinsically motivated behavior satisfies three basic psychological needs: relatedness, competence, and autonomy (Reeve et al. 2004; Ryan and Deci 2000b). SDT establishes a theoretical framework that relates these primary human needs to intrinsic motivation. Research in the tradition of SDT emphasizes the autonomy-supportive work environment as a relevant prerequisite to foster intrinsic motivation (Ryan and Deci 2000a, p. 58; see Lam 2015).

SDT also encompasses amotivation, which is any behavior that is not valuable or any compulsory task performed by actors who feel absolutely incompetent (Wilkesmann and Schmid 2014). The SDT model differentiates types of extrinsic motivation according to different levels of internalization of social norms and values: external, introjected, identified and integrated. Ryan and Deci (2000a, pp. 61–62) define external motivation as behavior that is rewarded and/or punished by others. The three other types that follow involve increasing levels of internalization of goals or external punishment. Introjected motivation “[. . .] describes a type of internal regulation that is still quite controlling because people perform such actions with the feeling of pressure in order to avoid guilt or anxiety, or to attain ego-enhancements or pride” (Ryan and Deci 2000a, p. 62). Identified motivation reflects a higher level of internalization in which the individual identifies him or herself with the behavior by valuing it as personally important. The highest level of internalization is integrated motivation. “Integration occurs when identified regulations have been fully assimilated to the self” (Ryan and Deci 2000a, p. 62). Action is in alignment with self-perception, and professors behave like professional academics.

Autonomy, as one of the three basic needs, is important for motivation because an internal locus of control is only possible when an individual inwardly grasps the meaning and worth of the regulation (Ryan and Deci 2000a). Otherwise, the regulation would be more external because a person would be following a rule only to avoid punishment.

H1 The more an academic work environment is perceived as autonomy supportive, the more intrinsic the motivation.

In addition, individuals internalize a social norm only when they feel related to the agent (a person, group or institution) of that norm (Ryan and Deci 2000b; Pelletier et al. 2002). Social relatedness is understood in this context as a social mechanism of appreciation, which fosters self-esteem and encourages individual initiative.

H2 The higher the perception of relatedness to an agent, the higher the intrinsic motivation.

Competence, as the third basic need, is a prerequisite for the internalization process. Only when a person is not over challenged and is acknowledged as competent can he or she internalize external expectations. The ascription as competent is necessary because otherwise an individual could not interact effectively with the environment and would therefore feel helpless.

H3 The higher the perception of acknowledged competence, the higher the intrinsic motivation.

The new steering instruments, such as pay-for-performance, performance-related budgets or MbO, crowd out intrinsic motivation if professors perceive them as control mechanisms (Frey 1997). All performance-related incentives require measurement; otherwise, behavior and bonuses cannot be related. Measurement is a monitoring capacity that generates an external rule. This externality could be perceived as an alienating institution.

H4 Selective managerial incentives at universities crowd out intrinsic motivation.

Regarding our second research question in this field (Can universities provide incentives that do not crowd out intrinsic motivation?), we must ask the following: how can the incentive system be structured such that intrinsic motivation is not crowded out? According to Frey and Neckermann (2008), academic rewards will not crowd out intrinsic motivation. Therefore, our fifth hypothesis is the following:

H5 Academic rewards do not crowd out intrinsic motivation.

We will provide empirical evidence for these hypotheses in the case of academic teaching.

5 Transactional Versus Transformational Governance

The fourth imaginary contradiction can be understood as an encompassing model of the first three contradictions. The terms 'transactional' and 'transformational' are based on the 'full range leadership model' (Bass and Avolio 1993). We will transform them to the governance discourse to describe different types of governance. NPM, which includes selective incentives, monitoring and sanction capacity, can be described as a form of transactional governance (Bass and Avolio 1993; Frost et al. 2010). Bass and Avolio (1993) defined transactional governance as follows: "There is a price on everything. Commitments are short-term. Self-interests are stressed" (Bass and Avolio 1993, p. 116). Conversely, transformational governance enables flexibility, autonomy for intellectual innovation and the ability to perceive employees as humans and individuals, take them seriously and be respectful. Bass and Avolio described transformational behavior as follows: “There is a rich set of norms which covers a wide range of behaviors; norms that will adapt to and change with external changes in the organizations’ environment. There is much talk at all levels in the organization about purposes, visions, and meeting challenges” (Bass and Avolio 1993, p. 118).

Transactional governance encompasses monitoring and sanction capacity, whereas transformational governance covers social norms that exist within organizations (Elster 1989; Inauen et al. 2010), such as the norms that guide the quality of research or approaches to teaching (Trigwell and Prosser 2004), organizational culture (Wilkesmann et al. 2009), and shared visions (Bass and Avolio 1993). There is empirical evidence that transactional governance has no impact on the perception of the significance of academic teaching but that transformational governance may have an effect on teaching (Wilkesmann 2013).

Nevertheless, a university cannot function without transactional governance. For some aspects, (e.g., a high number of examinations or additional administrative functions), an extra bonus could be justified. In this case, the incentive provides recognition for extra work that is time consuming and does not support the academic career.

6 Empirical Evidence

6.1 Survey Design

We provide empirical evidence based on two surveys. The first survey was conducted at research universities in Germany between May and July 2009 (Wilkesmann and Schmid 2012), and the second survey was conducted at universities of applied sciences in Germany between March and April 2011. The target population was all German professors at both types of universities. Both surveys were designed to analyze professors’ academic teaching behavior and are used here for a secondary data analysis.

For the first survey (Wilkesmann and Schmid 2012), we selected 8,000 research professors from the email distribution list of the German Association of University Professors (DHV). Professors paid within the framework of the new pay-for-performance salary (W-salary) scale were of special interest for the study; thus, we opted for a disproportionate stratified sampling approach that differentiated between two strata according to salary categories (merit pay vs. the age-related seniority scheme). A total of 1,119 professors completed the survey, constituting a response rate of 14 %; 58.5 % received pay-for-performance, and 41.5 % received the old seniority wages. Among the sample, 77.7 % were male, and 22.3 % were female. The mean age in our sample was 49.0 years.

The second survey was based on a list of emails of the German Association of University of Applied Science Professors (HLB). The HLB organizes all professors at universities of applied sciences, but the address list included only the deans of all German universities of applied sciences. We checked all email addresses and sent an email with a link to the online questionnaire that requested that the email be forwarded to all professors in their faculty. In total, 942 professors completed the questionnaire. In the sample, 47.8 % of professors received a performance-based salary, and 52.2 % received the old seniority wages; 87.7 % were male, and 21.3 % were female. The mean age of the professors in our sample was 50.3 years. Due to the distribution method, the response rate cannot be determined, but the sample covers 5.95 % of the population of all professors at universities of applied sciences.

The samples of both surveys were representative with respect to faculties, gender and age but not payment scheme. There was no need to weight the disproportionate strata for the purpose of multivariate analysis because we integrated the respective variables into the model.

For a more detailed measurement description, see Wilkesmann (2012, 2013) and Wilkesmann and Schmid (2012, 2014).

6.2 Empirical Results for Profession Versus Organization

We estimated an OLS regression with affective commitment as a dependent variable (see Table 1). The scale for affective commitment was an index (Cronbach's $\alpha = 0.78$) with the following four variables: "I perceive a strong sense of belonging to my university"; "I'm proud to tell other people that I'm a member of this university"; "I perceive the problems of my university as my own problems"; "Actually, I can work just as well at another university, when the general conditions are the same (recoded)". All items were measured on a five-point Likert scale ranging from 1 'I totally disagree' to 5 'I totally agree'.

The independent variables were the following:

new steering instruments, which were operationalized with the four dummy variables shown in Table 1;

commitment to the peers in the scientific community on a scale (Cronbach's $\alpha = 0.62$) of two items: "My colleagues and I are on the same wavelength" and "I'm highly appreciated by my colleagues";

attitude against NPM; to measure this attitude, we developed a four-item scale (Cronbach's $\alpha = 0.81$) comprising general reactance toward managerial governance, non-feasibility of measuring academic performance, inadequacy of managerial governance for professors, and awareness of managerial instruments as restricting control mechanisms.

We also added five control variables (see Table 1).

Table 1 Influence of NPM and professional recognition on affective commitment

		Affective Commitment α = .78 (beta)
H 1 new steering instruments	Pay-for-performance at the university (1 = yes; 0 = no)	0.077**
	Agreement on objectives including teaching (1 = yes; 0 = no)	0.008
	Teaching award winner (1 = yes; 0 = no)	−0.047*
	Performance related budgets at the university (1 = yes; 0 = no)	0.061**
H 2 commitment peers	Commitment to the peers in the scientific community	0.370**
H 3 attitude against NPM	Attitude against NPM	−0.114**
Control variables	Duration at the current university	0.081**
	Type of university (1 = university of applied sciences; 0 = research university)	0.020
	Discipline (1 = engineering; 0 = all others)	0.101**
	Gender (1 = male; 0 = female)	0.027
	Payment scheme (1 = pay-for-performance; 0 = old seniority pay)	0.046
	N	1,838
	Adjusted r^2	0.188

Level of significance 1 % (**); 5 % (*)

Hypothesis 1 was mostly rejected. Pay-for-performance and performance-related budgets increased affective commitment, but the impact was small. Affective commitment was reduced only for teaching award winners.

The same was true for hypothesis 2. Commitment to peers had the strongest impact on the dependent variable but in the direction opposite to that assumed. A high commitment to peers increased affective organizational commitment. The first two hypotheses were not confirmed. By contrast, hypothesis 3 was supported; an attitude against NPM reduced affective organizational commitment. There were two interesting results regarding the control variables: duration increased organizational commitment, a relatively straightforward result, and discipline affected organizational commitment. Engineers had higher affective commitment than members of other disciplines.

In summary, the new steering instruments and an attitude toward (not against) NPM increased affective organizational commitment. The shift to managerial governance supported the development toward a complete organization. The new managerial instruments that strengthened the hierarchy supported the university as a complete organization. Simultaneously, recognition from colleagues in the profession increased organizational commitment. We could conclude that, in the perception of German professors, there was no contradiction between profession and organization.

6.3 *Empirical Results for Intrinsic Motivation*

To measure SDT, we used items from the Work Tasks Motivation Scale for Teachers of Fernet et al. (2008) and the Academic Motivation Scale developed by Vallerand et al. (1992) (for a more detailed description, see Wilkesmann and Schmid 2014). All items were measured on a five-point Likert scale ranging from 1 ‘I totally disagree’ to 5 ‘I totally agree’. We used a principal component analysis (PCA) to test the dimensionality of our translated and modified motivation scale. The PCA with varimax rotation revealed four latent variables (KMO-value .830; explained variance 57 %): intrinsic motivation (Cronbach’s $\alpha = 0.79$), introjected motivation (Cronbach’s $\alpha = 0.65$), extrinsic motivation (Cronbach’s $\alpha = 0.68$), and amotivation (Cronbach’s $\alpha = 0.61$). The empirical merger of the intrinsic and the identified motivation subscales explained the difficulty of analytically differentiating between these two levels of internalization for our sample or task of academic teaching. We used only intrinsic motivation as a dependent variable for the OLS regression (see Table 2).

Table 2 Influences of the three basic needs and crowding-out and awards on intrinsic teaching motivation

		Intrinsic teaching motivation $\alpha = 0.79$ (beta)
H1 autonomy	More autonomy in comparison with private companies	0.076**
H 2 relatedness	Support from the dean	0.027
	Students actively participate in teaching	0.062**
H 3 competence	Approach to teach was a central criteria for my appointment	0.255**
H 4 crowding-out effect	Receiver of merit pay for teaching (1 = yes; 0 = no)	0.011
	Agreement on objectives includes teaching (1 = yes; 0 = no)	−0.078*
	Extrinsic teaching motivation	−0.299**
H 5 awards	Teaching award winner (1 = yes; 0 = no)	0.054*
Control variables	Gender (1 = male; 0 = female)	−0.006
	Age	−0.033
	Payment scheme (1 = pay for performance W; 0 = old wage system C)	0.059
	Duration of employment at the current university	0.037
	N	1,787
	Adjusted r^2	0.193

Note Level of significance 1 % (**); 5 % (*)

The independent variables were:

autonomy: To measure perceived autonomy, we used the item "To work autonomously is a value in itself which cannot be compensated with all the known incentives exclusively provided by private sector companies (e.g., higher income, company car, etc.)".

relatedness: We operationalized perceived relatedness with the following two items: "The dean provides active support for the enhancement of teaching activities" and "My students are eager to actively participate in teaching".

competence: Perceived competence was operationalized with the following item: "My approach to teaching was a central criterion for my [successful] appointment".

crowding-out effect: We used two dummy variables that could be answered with 'yes' (=1) or 'no' (=0) "Are you receiving merit pay [bonuses] for teaching?" and "Does your agreement on objectives [with the dean/rectorate] include any statements on the advancement of teaching activities?" In addition, we integrated the index of extrinsic motivation as an independent variable.

awards: We used the dummy-variable "Have you ever won a teaching award?"

Furthermore, we controlled for age, gender, payment scheme, and the duration of employment at the current university.

Hypotheses 1–5 were supported. Autonomy and perceived competence increased intrinsic motivation. For relatedness, we only found evidence when the professors perceived support from students. There was no effect of perceived support from the deans. A plausible interpretation of this result is that, in Germany, deans are not known to intervene in teaching activities. For a crowding-out effect, we found some indication that agreement on objectives that included teaching had a negative impact on intrinsic teaching motivation. In addition, extrinsic teaching motivation had a negative impact on the dependent variable. An appropriate empirical validation of this hypothesis would require longitudinal data. Teaching awards appear to increase, not crowd out, intrinsic motivation (Frey and Neckermann 2008). This increase can be attributed to the nature of the awards, which did not qualify as selective incentives because they had no effect on the distribution of monetary and personal funds within universities. Professors perceive teaching awards not as a monitoring but rather as an appreciation tool. None of the control variables had an effect on intrinsic teaching motivation.

In the case of academic teaching, we found empirical evidence for the basic assumption of SDT. Autonomy, relatedness, and competence were relevant for intrinsic motivation. All three factors describe, coincidentally, transformational governance. In addition, we found some indication that extrinsic rewards could crowd out intrinsic motivation. At least in the case of teaching, the different regulatory styles in the SDT model were in conflict with each other. Professors were not simultaneously intrinsically and extrinsically motivated to the same extent.

Discussion and Conclusion

Regarding the imaginary contradiction between profession and organization, we found empirical evidence that NPM supports the development of universities toward a complete organization. In the perception of German professors, there is no conflict between organization and profession. Like professional service organizations (PSO), universities must manage more or less deviant members (autonomous working researchers and teachers) to ensure that they were working together toward a common goal, at least in terms of academic teaching.

The empirical evidence for the imaginary contradiction between intrinsic vs. extrinsic motivation demonstrates that intrinsic teaching motivation is necessary for innovative academic work. However, intrinsic motivation only occurs when professors perceive an autonomous, supportive environment. NPM launched selective incentives in the university, which can crowd out intrinsic motivation. Our data about academic teaching indicate one exception: teaching awards. Awards are most likely extrinsic rewards that do not crowd out intrinsic motivation at universities (Frey and Neckermann 2008).

In summary, universities need transformational governance as well as some transactional governance. Strengthening the organizational hierarchy with the help of the new steering instruments (performance-related budgets, MbO) increases the affective commitment of professors to the organization university. Simultaneously, professors need autonomy for intellectual innovation and respectful treatment by the rectorate. The new selective incentives established by the governments mislead academics to "game the system" (van Thiel and Leeuw 2002). National science foundations and politicians must counteract this development. For innovation in teaching as well as scientific development, professors must become 'institutional entrepreneurs' who are creative and who change the organization and their scientific field. These 'institutional entrepreneurs' need 'opportunity structures' that provide opportunities for success. 'Opportunity structures' include autonomy in organizations or the support of individual projects by national research foundations. If professors are guided only by a carrot-and-stick policy, they will not be innovative.

The imaginary contradiction between monitoring and autonomy exists in private industry as well. In PSOs such as consulting companies, the organization structure of a partnership can overcome this contradiction (Greenwood et al. 2007; Greenwood and Empson 2003). In a partnership, all members are principals and agents simultaneously. Partnerships are organizations with a strong collaborative community in which shared values and norms are more important than formal rules (Adler and Heckscher 2006, 2011). Similarly, in partnerships, numbers are helpful for a common reflection about shared and collaborative decisions toward a mutual goal.

(continued)

These imaginary contradictions describe ‘second-level management’, that is, rectorate/superiors can only supply opportunities for people to conduct research and teaching. The rectorate cannot directly monitor, reward, or punish the production of research or teaching. Both research and teaching must be managed at a second level. Superiors must treat employees as the most valuable asset the organization offers because transformational governance is a vulnerable factor: it is easier to undermine than build up.

References

Adler PS, Heckscher C (2006) Towards collaborative community. In: Heckscher C, Adler PS (eds) The firm as collaborative community. Reconstructing trust in the knowledge economy. Oxford University Press, Oxford, pp 11–105

Adler PS, Heckscher C (2011) From clans to collaboration: collaborative community as the basis of organizational ambidexterity. Working paper. https://t.co/JMYiMI9s. Accessed 1 May 2014

Ahrne G, Brunsson N (2011) Organization outside organizations: the significance of partial organization. Organization 18(1):83–104

Bass BM, Avolio BJ (1993) Transformational leadership and organizational culture. PAQ Spring 1993:112–121

Bernard CI (1938) The functions of the executive. Harvard University Press, Cambridge

Bourdieu P (1988) Homo Academicus. Stanford University Press, Stanford

Carney BM, Getz I (2009) Freedom, Inc. Crown Business, New York

De Boer H, Endres J, Schimank U (2007) On the way towards new public management? The governance of university systems in England, the Netherlands, Austria, and Germany. In: Jansen D (ed) New forms of governance in research organizations. Springer, Dordrecht, pp 137–154

Eisenhardt K (1989) Agency theory: an assessment and review. Acad Manag Rev 14:57–74

Elster J (1989) The cement of society. Cambridge University Press, Cambridge

Fernet C, Senécal C, Guay F, Marsh H, Dowson M (2008) The work tasks motivation scale for teachers (WTMST). J Career Assess 16:256–279

Freidson E (2001) Professionalism. The third logic. University of Chicago, Chicago

Frey BS (1997) Not just for the money. Edward Elgar, Cheltenham

Frey BS, Neckermann S (2008) Academics appreciate awards: a new aspect of incentives in research. Working paper 32. Center for Research in Economics, Management and the Arts, University of Zurich

Frey BS, Osterloh M (2002) Successful management by motivation. Springer, New York

Frey BS, Homberg F, Osterloh M (2013) Organizational control systems and pay-for-performance in the public service. Organ Stud 34(7):949–972

Frost J, Osterloh M, Weibel A (2010) Governing knowledge work: transactional and transformational solutions. Organ Dyn 39(2):126–136

Getz I (2009) Liberating leadership: how the initiative-freeing radical organizational form has been successfully adopted. Calif Manag Rev 51:32–58

Greenwood R, Empson L (2003) The professional partnership: relic of exemplary form of governance? Organ Stud 24:909–934

Greenwood R, Deephouse D, Xiao Li S (2007) Ownership and performance of professional service firms. Organ Stud 28:219–238

Hüther O, Krücken G (2011) Wissenschaftliche Karriere und Beschäftigungsbedingungen (Employment conditions and career perspectives). Soziale Welt 62:305–325

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
Inauen E, Rost K, Osterloh M, Frey BS (2010) Back to the future – a monastic perspective on corporate governance. Manag Rev 21:38–59
Lam A (2011) What motivates academic scientists to engage in research commercialization: 'gold', 'ribbon', or 'puzzle'? Res Policy 40:1354–1368
Lam A (2015) Academic scientists and knowledge commercialization: self-determination and diverse motivations. In: Welpe IM, Wollersheim J, Ringelhan S, Osterloh M (eds) Incentives and performance - governance of research organizations. Springer, Cham
Matiaske W, Weller I (2003) Extra-Rollenverhalten. In: Martin A (ed) Organizational behaviour – Verhalten in Organisationen. Kohlhammer, Stuttgart, pp 95–114
Meyer JP, Allen NJ (1991) A three-component conceptualization of organizational commitment. Hum Resour Manag Rev 1:61–89
Mintzberg H (1989) Mintzberg on management: inside our strange world of organizations. Free Press, New York
Moses I, Ramsden P (1992) Academic values and academic practice in the new universities. High Educ Res Dev 11(2):101–118
Musselin C (2006) Are universities specific organisations? In: Krücken G, Kosmützky A, Torka M (eds) Towards a multiversity? Universities between global trends and national traditions transcript, Bielefeld, pp 34–67
Osterloh M, Frey BS (2013) Motivation governance. In: Grandori A (ed) Handbook of economic organization. Edward Elgar, Cheltenham, pp 26–40
Pelletier LG, Seguin-Lévesqui C, Legault L (2002) Pressure from above and pressure from below as determinants of teachers' motivation and teaching behaviors. J Educ Psychol 94:186–196
Reeve J, Deci EL, Ryan RM (2004) Self-determination theory: a dialectical framework for understanding socio-cultural influences on student motivation. In: Van Etten S, Pressley M (eds) Big theories revisited. Information Age, Greenwich, pp 31–60
Ryan RM, Deci EL (2000a) Intrinsic and extrinsic motivation: classic definitions and new directions. Contemp Educ Psychol 25:54–67
Ryan RM, Deci EL (2000b) Self-determination theory and the facilitation of intrinsic motivation, social development and well-being. Am Psychol 55:68–78
Ryan RM, Deci EL (2006) Self-regulation and the problem of human autonomy: does psychology need choice, self-determination, and will? J Pers 74:1557–1558
Starbuck WH (1992) Learning by knowledge-intensive firms. J Manag Stud 29:713–740
Trigwell K, Prosser M (2004) Development and use of the approaches to teaching inventory. Educ Psychol Rev 16(4):409–424
Vallerand RJ, Pelletier LG, Blais MR, Briere NM, Senecal C, Vallieres EF (1992) The academic motivation scale: a measure of intrinsic, extrinsic, and motivation in education. Educ Psychol Meas 52:1003–1017
Van Thiel S, Leeuw FL (2002) The performance paradox in the public sector. Public Perform Manag Rev 25:267–281
Wilkesmann U (2012) Auf dem Weg vom Gelehrten zum abhängig Beschäftigten? Zwei deutschlandweite Surveys zur Lehrmotivation von Professoren. In: Wilkesmann U, Schmid CJ (eds) Hochschule als organisation. Springer, Wiesbaden, pp 363–381
Wilkesmann U (2013) Effects of transactional and transformational governance on academic teaching – empirical evidence from two types of higher education institutions. Tert Educ Manag 19(4):281–300
Wilkesmann U, Schmid CJ (2012) The impacts of new governance on teaching at German universities. Findings from a national survey in Germany. High Educ 63:33–52
Wilkesmann U, Schmid CJ (2014) Intrinsic and internalized modes of teaching motivation. Evidence Based HRM 2(1):6–27

Wilkesmann U, Wilkesmann M, Virgillito A (2009) The absence of cooperation is not necessarily defection: structural and motivational constraints of knowledge transfer in a social dilemma situation. Organ Stud 30(10):114–164
Wilkesmann U, Blutner D, Müller C (2011) German football: organizing for the European top. In: Gammelsaeter H, Senaux B (eds) The organisation and governance of top football across Europe. Routledge, New York, pp 138–153

Research and Teaching Awards as Elements of Incentive Systems in Academia

René Krempkow

Abstract Research and teaching awards are non-monetary incentives. This contribution asks which role awards may play in order to acknowledge performance in teaching and research (This contribution is a revised version of the lecture given at the conference "Innovation, achievement performance measurement and incentive systems in academia and business—Governance of knowledge-intensive organizations" at the Technische Universität München (Munich University of Technology) on January 14th and 15th 2014. For helpful comments to my presentation and for this book chapter I like to thank some participants and the organizers of this conference. For the translation of this article I like to thank Dorit Rowedder and Susan Harris-Hümmert). It is based on surveys conducted in the context of the project "GOMED" (The project "GOMED—Governance of university medicine: Intended and non-intended effects of decentralised incentive systems using the example of the performance-based funding within the respective faculty in medicine" was funded by the German Federal Ministry of Education and Research). The chapter is divided into three parts: The first part discusses the potential of teaching and research awards in incentive systems in academia. This discussion is based on the available literature and on our own findings. The second part presents investigations on the number as well as the prize money of awards in Germany. This includes teaching and research awards that are relevant for researchers at medical faculties in Germany as well as those that are open to other subjects. The third part summarizes and discusses possibilities of further developments for teaching and research awards, e.g., a higher prize money or a higher number of team awards.

R. Krempkow (✉)
Forschungsinstitut für Bildungs- und Sozialökonomie, Berlin, Germany
e-mail: r.krempkow@fibs.eu

I.M. Welpe et al. (eds.), *Incentives and Performance*,
DOI 10.1007/978-3-319-09785-5_13

1 The Potential of Awards in Incentive Systems in Academia

Previous research on effects of monetary incentives, e.g., on performance-based funding in academia, shows that such fundings reach their objectives only to a limited extent (cf. the contributions in this book and in addition the contributions in Grande et al. (2013), Wilkesmann and Schmid (2012), Winter and Würmann (2012). This even applies under relatively positive conditions of effectiveness as, e.g., in medicine. In medicine experiences with high amounts of performance-based fundings exist for a long time (cf. Krempkow et al. 2013). They show a number of non-intended effects, e.g., the publication of many brief instead of few comprehensive articles ("salami tactics"), a preference for "go-it-alone strategies" (in contrast to cooperation strategies) and "mainstream research" (cf. ibid.).[1] However, intended effects such as perceived greater transparency of the research achievements and fiercer competition among research institutions can also be witnessed. The intensity of such intended effects is related to an increase in the reputation of involved academics.[2] Since "reputation is an essential incentive in the scientific community" (Osterloh and Frey 2008), non-monetary incentives such as research and teaching awards[3] may also serve as an effective (and possibly efficient) reward for scientific achievements.[4] Hornbostel (2002) observes an explosive proliferation of academic awards since the 1970s and states that honor or recognition in the form of awards played a major role also in modern societies. The German Association of University Professors and Lecturers (DHV 2000) stated "Lecturers must be

[1] On founded presumptions on non-intended incentive effects for the research area see also German Council of Science and Humanities (2011). For the teaching area, e.g., Dohmen and Henke (2012) show that intended effects are accompanied by unintended effects.

[2] Furthermore the goal attainment of the performance-based funding is very positively related to the justice perception of performance-based funding and the discussion of the findings of performance-based funding (this applies to all target dimensions: transparency of the research achievements, efficiency and quality). It is negatively related to the reward of "mainstream research" (cf. Krempkow et al. 2013). Additionally, there was an effect of the justice perception of the performance-based funding in publication analyses, i.e.: Those faculties where the justice of the performance-based funding is perceived to be greater are the same that show greater publication performance (cf. Krempkow et al. 2013).

[3] According to Ziegele and Handel (2004) recognition and reputation along with freedom (in the sense of autonomy), time allocation (e.g., research sabbaticals of one semester) and transparency (for all those involved with a view to triggering cost-conscious dealings) are among the non-monetary incentives. However, performance-based funding may also create space and as a result additional autonomy as well as transparency with regard to the achievements if designed correspondingly (cf. Krempkow et al. 2013). This may be the case by creating research possibilities that may otherwise not be available (cf. in greater detail Krempkow 2007). As awards usually aim at reputation and recognition, above all else, they appear to be a particularly interesting example of non-monetary incentives in this context.

[4] The economy, too, provides examples that some companies prefer to reward quality work with personal recognition instead of money "because personal recognition has proved to be an extremely effective motivational tool" (cf. Hochschild 1998; Oelkers and Strittmatter 2004).

motivated more by non-monetary incentives than financial acknowledgement",[5] suggesting the introduction of an annual German science award including a high prize money (as well as for the National Scientific Award). According to the DHV, special awards were to be created for excellent research across subjects and for the promotion of junior academics, outstanding teaching or comprehensible imparting of scientific findings to the public. Finally, a more recent interview study (Becker et al. 2012) emphasized non-monetary incentives as the most important motivator though monetary incentives were offered academics above all else.

Awards may be seen as signals from a more theoretical perspective. Signaling theory helps to explain award bestowals (Frey and Gallus 2014).[6] Awards transmit signals that transform the content and interpretation of information emitted by actors. They are non-material and derive their value from their symbolic nature. The value to the recipient usually exceeds the costs that the giver incurs.

The basic idea reported by Frey and Gallus (2014) is: If the signal is perceived as credible it will influence beliefs and may thus also alter the signal receiver's behavior towards the award giver[7] For example, awards can display which behaviors the award giver values, with no need to exactly define, measure and enumerate the winners single deeds.[8] Thus, the less easily performance criteria and tasks can be defined ex ante and observed ex post, the more prevalent awards are. In contrast, monetary rewards, particularly in their more stringent form of pay-for-performance, require precisely-defined measures of performance. By using awards, the principal circumvents important limitations posed by monetary rewards: Even in situations where the desired tasks are vague and cannot be contracted, the principal maintains the ability to influence the behavior of the recipient and, most importantly, of future candidates and the wider audience. Moreover, the principal reduces the risk of motivation crowding out[9] when using awards instead of monetary rewards. Awards also strengthen intrinsic motivation because the principal signals trust and

[5] Also Witte et al. (2001) call for reductions of the teaching load and sabbaticals to be used as non-monetary incentives to avoid "that monetary incentives replace or even destroy the existing intrinsic motivation" (cf. Minssen and Wilkesmann 2003; Hellemacher et al. 2005; Krempkow 2007). As far as incentives in the context of university teaching are concerned, it is uncertain though whether reductions of the teaching load and similar releases from lecture can point the way to the desired direction in order to strengthen the significance of teaching compared to research as sought by politics.

[6] The theory of awards has not had any major developments in the last 30 years (Frey and Gallus 2014).

[7] Frey and Gallus (2014) used the principal-agent framework to describe the relationship between the award giver and the award winner and (potential) recipients.

[8] By bestowing awards, the principal emits signals about his or her quality that monetary rewards cannot transmit: First, the principal signals a high degree of interpersonal skills (e.g., attentiveness); second, the award can serve as a signal of the principal's authority within the organizational hierarchy.

[9] Crowding out means (in short terms): External interventions reduce intrinsic motivation and replace by extrinsic motivation if the individuals affected perceive them to be controlling (and not supporting) (cf. Frey 1997).

confidence in the latter's future performance by bonding his or her name to the recipient's. This implicit backing can even provoke a crowding-in effect to enhance intrinsic motivation.

The signal strength and effects differ according to the specific circumstances. Frey and Gallus (2014) distinguish between two types of awards because they vastly differ in their role and strength as a signal: *Confirmatory awards* are bestowed at regular intervals with defined performance criteria and the award is always given to whoever was the previous period's best performer. Confirmatory awards are an addendum to regular incentives (e.g., bonuses) that employees compete for. *Discretionary awards* are awards where the principal may decide when and upon whom they are bestowed. They can be given, for instance, for unexpected services of an agent (such as helping colleagues), which would not be honored in the standard incentive scheme. These awards tend to be given ex post for the observed behavior, often to the surprise of the winner. By adding a monetary prize to the award, the principal can signal the seriousness of the award and can establish the award among competing awards.

When the above aspects are taken as a basis, there are a number of arguments that seem to suggest that to date non-monetary incentives have untapped potential to acknowledge achievements in academia. We should therefore take a closer empirical look at the effects of non-monetary incentives.[10]

In studies on the effects of performance-based funding in medicine, the GOMED project[11] considered which role incentive mechanisms could play, which are either apart from or given in addition to financial incentives. As part of a nationwide survey of professors (cf. Krempkow et al. 2011), as well as a second survey of scientific staff at selected medical faculties (cf. Krempkow and Landrock 2013)[12] one of the questions on the performance-based funding aimed at understanding the medical scientists' perspective as to how important further incentive mechanisms were for medical research in universities.[13] The wording of the question was deliberately broad; no concrete alternative incentive mechanisms were laid down. Before we will later address selected aspects of professional motives and reputation, Fig. 1 represents answers to this question.

[10] Until now, there is little evidence on the effect of awards on performance in German academia [which is also observed in the non-academic sector (cf. Neckermann et al. 2012)]. A noteworthy study on research awards is by Chan et al. (2013), which shows that the scientific performance of the award winner is significant higher than in a synthetic control group. For teaching awards Wilkesmann and Schmid (2012) show in a regression analysis a significant higher intrinsic teaching motivation for teaching award winners at universities.

[11] From 2010 to 2013 the author headed the GOMED project organized as part of the iFQ Berlin, Germany.

[12] *N* in each of the surveys was more than 600 respondents. The scientific staff survey includes post-docs, PhDs and senior researchers. Detailed reports on the findings of the surveys including methodological explanations are available on the project website: www.forschungsinfo.de/Projekte/GOMED/projekte_gomed.asp.

[13] Furthermore, we asked questions such as which performance indicators were used in the faculties of medicine, and which indicators should be used from the view of academics (cf. Krempkow et al. 2013).

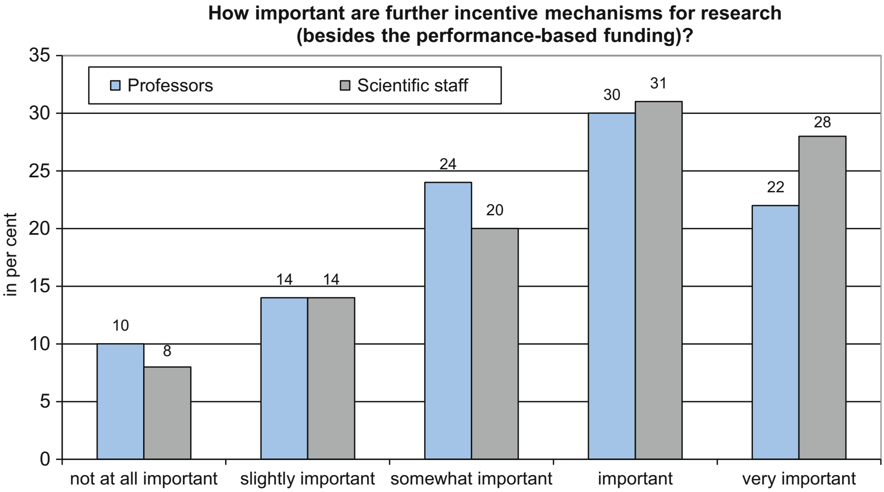


Fig. 1 Importance of further incentive mechanisms [This figure and all following figures were reproduced and translated from Krempkow et al. (2013)]

Figure 1 shows that the majority of the respondents regards further incentive mechanisms as important: More than three quarters (76 % of the professors and 79 % of the scientific staff) regard further incentive mechanisms as at least partly important, more than half of each group even as (very) important.[14] This shows that the participating respondents were very open-minded towards further mechanisms of incentives besides the performance-based funding.

As is well known, the potential for non-monetary incentives are also motive-dependent (e.g., cf. Heckhausen and Heckhausen 2010). As Becker et al. (2012, p. 196) put it, mistakes in the analysis of the motivation may "at best randomly lead to efficient measures and consequences". To better understand the motivation of the participating scientists the figure below provides a closer look at the participants' professional motives.

Figure 2 shows that both groups of respondents rate the possibility of realizing one's own ideas (professional autonomy) and the possibility of working scientifically as most important. This applies to an even greater extent to the professors than to the scientific staff.[15] In addition, professors feel that taking over of coordination

[14] In our survey only those respondents who had previously indicated that they were aware of further mechanisms of incentive were asked this question. Semi-structured interviews (conducted as a part of the GOMED project) permitted us to deduce that those further incentive mechanisms include objective agreements, innovation funds and non-monetary incentives such as awards and sabbatical semesters.

[15] Here we have to take into account that in such survey data the respondents answer not independently from social desirability. For the analysis of the effects of awards it is therefore recommended to include other methods, e.g., publication and citation analyses [for an example of awards for economists see Chan et al. (2013); for bibliometric analyses of publication and citation data in medicine see Krempkow et al. (2013)].

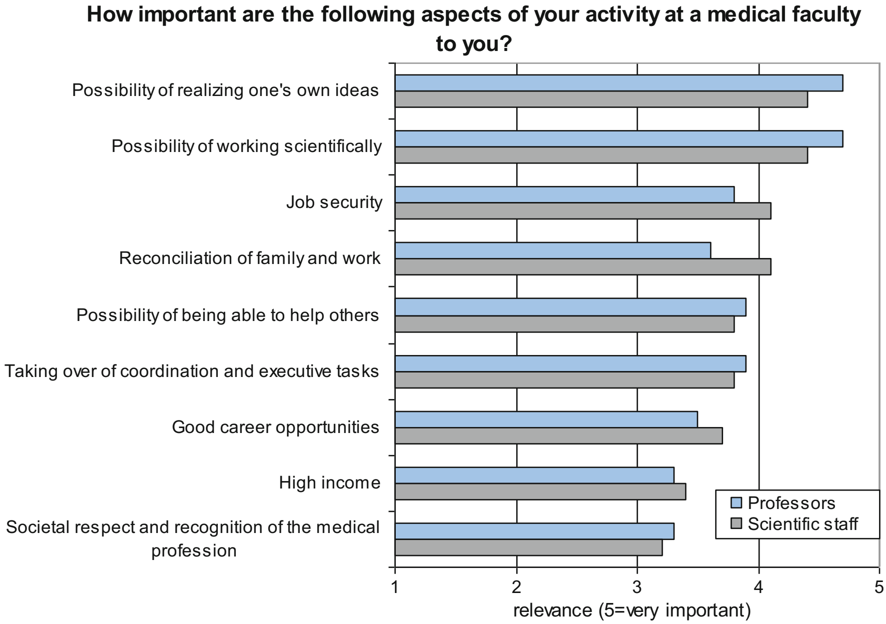


Fig. 2 Academic's motivational aspects at medical faculties

and executive tasks are also relatively important. Scientific staff also regarded the reconciliation of family and work, and furthermore job security as relatively important. Both groups ascribe less importance to high income.[16] Even though the respondents' assessments must be taken with some caution, as they are based on self-evaluation, it can still be stated that monetary incentives do not play the most important role. Since literature, as mentioned above, has repeatedly pointed to the significance of reputation in academia, we show in more detail what that means for professors and the scientific staff.

Figure 3 shows that the prestige of the journals in which one's work is published in is most important to both groups. Next, other publication-related aspects like the number of citations, the impact factor of their own publications and the number of publications are of significance.[17] It underlines that respondents feel that their scientific work and related publications are important. Furthermore this also suggest that it is not the academia-related possibilities of the work alone that—among other things—could be based on the available patient beds, but that also prestige-

[16] For the latter aspect the amount of the income is nevertheless considered to be an important so-called hygiene factor (Herzberg 1966) whose existence alone does not lead to greater achievements on a permanent level, but whose drop below a certain level perceived as appropriate, may have an adverse effect.

[17] For the group of professors the amount of raised external funding is ranked in third position.

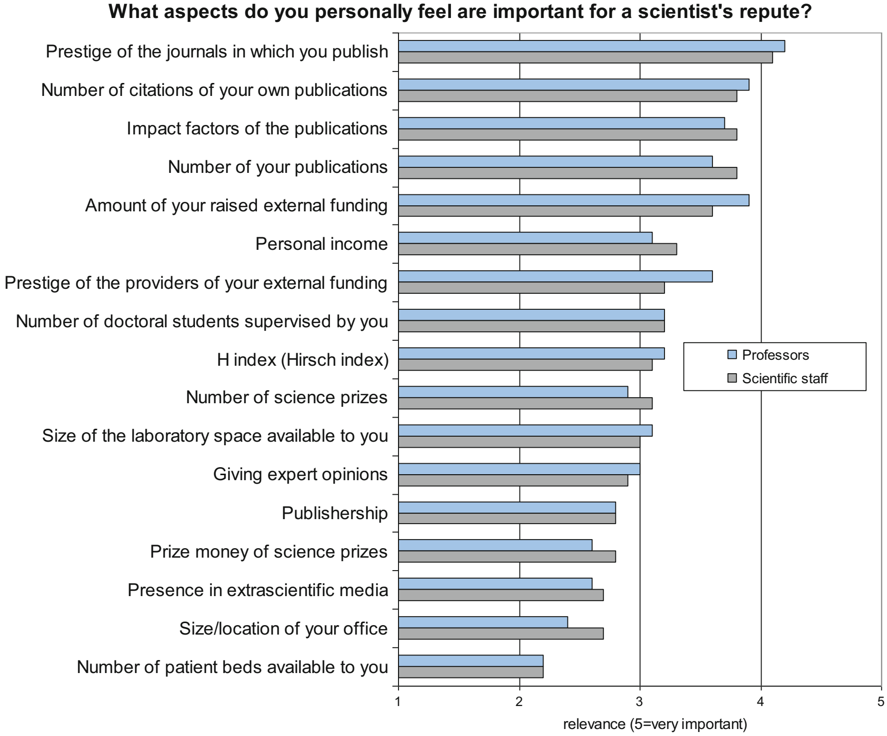


Fig. 3 Reputation-relevant aspects for academics at medical faculties

related aspects play an essential role.[18] In Fig. 3 the significance of awards in the scientists' perception of reputation becomes also apparent. Although awards are clearly more important than office size or the number of patient beds, they are however clearly less important than the publication-related aspects. For scientific staff the number of academic awards is almost more important than average of all aspects (and above the middle of the scale)[19] and somewhat more important than to professors. Here the question arises to which extent the potential of non-monetary incentives described at the beginning of this contribution has been fully tapped by the currently existing awards in academia. It is also relevant to know about the distribution of prize money for existing awards, because prize money can send confounding signals: The amount may be perceived too high, thus overriding the honorific signal of the award, or too low, thereby challenging the seriousness of the award (cf. Frey and Gallus 2014). Therefore, the following section is to take stock of the number and the prize money of research and teaching awards in Germany.

[18] There might also be other reasons. Some scientists might care more for publications in order to advance their career and/or gain power.

[19] The "should-be state" of performance indicators was assessed in a survey of young academics in the field of economics in Germany. Therefore the number of research awards was also revealed as an indicator which should be used to measure research performance (cf. Wollersheim et al. 2014).

2 Taking Stock of Research and Teaching Awards

There is an extensive variety of awards for academics in Germany ranging from awards for research and teaching to awards for knowledge transfer to the public[20] (cf. Jorzik 2010; Krempkow 2010; Krempkow et al. 2013; Wilkesmann and Schmid 2010). In the following we address research and teaching awards in Germany only. Examples of renowned research awards are the Leibniz prize, the Alexander von Humboldt prize or the Sofja Kovalevskaja prize (including prize money of more than one million €). Examples of teaching awards are the Ars Legendi prize of the German University Rectors' Conference (HRK, 50,000 €) as well as the prize for excellence in teaching awarded by the Stifterverband.[21] The latter is an award for academics, teaching teams or institutions which includes a prize money of up to one million € over three years.

The following overview of these awards and their prize money is a starting point for assessing the potential of the use of relevant awards besides and alongside performance-based funding. We assume that an award including high prize money could represent either a monetary incentive or a reputation gain (cf. Krempkow 1999, 2010) that may have a signalling effect (e.g., in application procedures). The greater the prestige of the prize giver is, the greater is this signalling effect.

To date to our knowledge there exists no comprehensive overview of academic awards offered in Germany. The following represents our findings on awards and prize money that were conducted within the framework of the GOMED project (cf. Krempkow et al. 2013).[22] The analyses are based on three basic sources: (1) two databases provided by the Internet portals "academics.de" and "forschen-foerdern.org"[23] (each including a subset of academic awards), (2) two books on teaching awards (Tremp 2010; Cremer-Renz and Jansen-Schulz 2010), as well as (3) personal knowledge and supplementary Internet investigations mainly on the prize

[20] According to a nationwide survey, 9 % of the teaching staff has already received teaching awards. It can be assumed that many of these are faculty- or subject-level awards. In the overview of Jorzik (2010) presumably the letter are not included. In comparison—according to Wilkesmann and Schmid (2010)—only 5 % of the professors received achievement-related income bonuses and 42 % a performance-based funding on chair level.

[21] The Stifterverband is the business community's innovation agency for the German academic system.

[22] The survey was conducted in 2010/2011 and the prize money of individual awards may have changed in the meantime. It can be assumed, however, that the charted fundamental dimensions have not been very liable to change. Furthermore it can be assumed that our survey is still the most recent and most complete survey that is available, at least in this area. Frey and Gallus (2014) also stated that reliable data on award bestowals is not yet available and will require much work in the future.

[23] 'Medicine/healthcare' was entered as the special field in the drop-down menu of the database on academics.de: We gathered all scientific awards that were found. In the database on forschen.foerdern.de 'human medicine', 'diagnoses' were entered as special fields and 'scientists', 'university lecturers' as target groups.

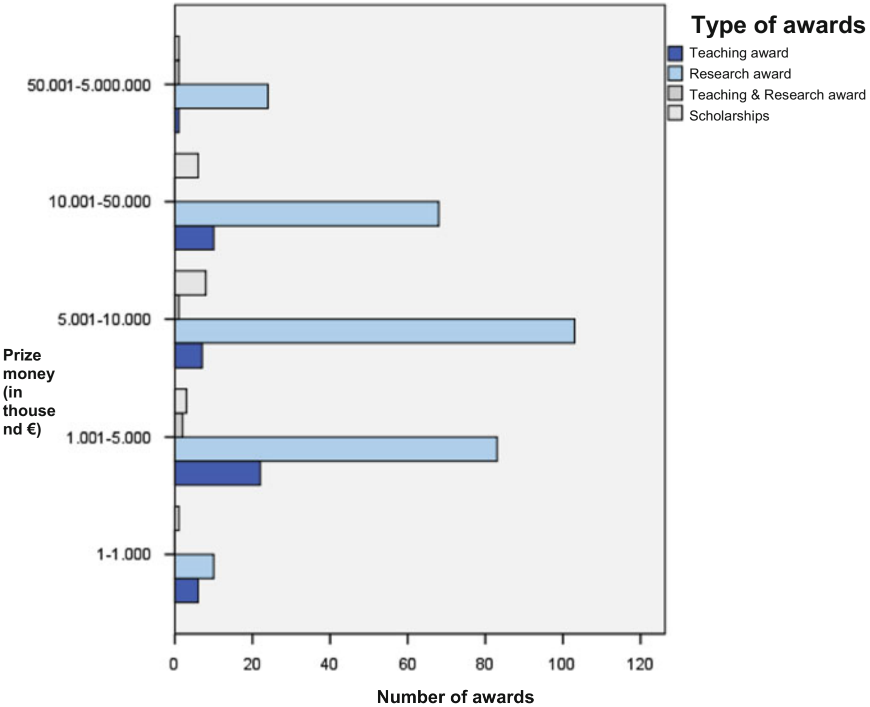


Fig. 4 Number and prize money of academic awards in Germany

money of awards.[24] As the focus of the project was the medical field, we restricted our investigations to awards that are relevant for scientists in medicine. We also considered all those awards that are not exclusively awarded for medicine research but are open to all subjects. According to our investigations there were about 400 academic awards and fellowships that were relevant for scientists in medicine in 2011. Figure 4 shows their distribution grouped by the amount of the prize money and the type of awards.[25]

---

[24] All findings that were not yet available in the database of academics.de were gathered (comparing the titles using the search function). As there were hardly any details available on awards in the database of forschen-foerdern.org, we used google.de to search for the name of the award to receive information on prize money. As a rule, the first two result pages were searched, as this turned out to be sufficient in most cases. When there was no prize money, this was marked in our database in addition to when no information was found. Any variations in prize money that was indicated in the database were noted. When there were additional awards (medals, certificates) this was also documented. 'Scholarships' were specified separately. Any items that did not fall in the categories 'research awards' or 'scholarships' were included in 'Miscellaneous'. I thank our former assistant Verena Walter for this work.

[25] According to Frey and Gallus (2014) two types of awards can be assumed, while most research and/or teaching awards are confirmatory awards. Unfortunately in the databases it is not possible to allocate it exactly.

As can be seen in Fig. 4, the prize money ranges from (less than) thousand up to five million €.[26] Out of the 400 awards considered, 301 are exclusively bestowed for research (prize money between 1,000 € and five million €). Approximately a third of these research awards include prize money of up to 5,000 €, another third includes prize money between 5,001 and 10,000 € and the remaining third includes prize money of more than 10,000 €. In addition, there are a few awards for research *and* teaching as well as some awards including high prize money that (at the same time) are offered as scholarships. For the sake of completeness they are added in the graph above; however, due to their low number we do not take a closer look at them. The average prize money for research awards is about 70,000 €.[27] The prize money of the 47 teaching awards we considered[28] is between 300 € and one million €, but the majority of the awards (two third) include prize money of up to 5,000 €. For approximately a fourth of the teaching awards 10,000 € and more is awarded.[29] The average prize money for teaching awards is rounded 14,000 €.[30]

Conclusion

This chapter—based on research findings on intended and non-intended effects of incentive systems—has shown that financial incentive systems have only partly reached their objectives and that they go hand in hand with significant non-intended effects. This has been put down to the fact that reputation is an essential incentive. Though academics are motivated highly by non-monetary incentives in reality they were offered monetary incentives in the first place. Therefore, we suggest that non-monetary incentives are a still untapped potential to acknowledge scientific achievements. This potential could be used as an alternative or it could supplement financial incentives.

(continued)

[26] In a specific search run for medicine alone, approximately 300 of those awards could be found and thus display the majority.

[27] This is the arithmetic mean. The distributions are skew. The median for research awards is 10,000 €.

[28] Some federal states (such as Baden-Württemberg, Bavaria, Rhineland-Palatinate) and some German universities have been awarding teaching awards for some time; not all of these are included in our analysis (e.g., the Universität Münster, the Technische Universität Dresden, the Technische Universität Bergakademie Freiberg as early as in the 1990s). Many of these awards have been bestowed only recently. This may be the reason why teaching awards are not completely recorded, despite the intensive efforts undertaken by Jorzik (2010). The nine percent of teaching staff that according to Wilkesmann and Schmid (2010) have received an award may also be incompletely recorded because of this reason. However, teaching awards usually attract public attention so that the estimated number of unreported cases might be low.

[29] cf. Hornbostel (2002) on the relation of frequency and value of awards in the former German Democratic Republic (East Germany between 1949 and 1990).

[30] This is the arithmetic mean. The distributions are skew. The median for teaching awards is 5,000 €.

The empirical findings of our surveys show—exemplary for medical scientists in Germany—that in fact more than half of the respondents indicate other than monetary incentives as (very) important. In line with the studies mentioned we show that immaterial aspects are most important. Examples are the possibility of working scientifically and the possibility of realizing one's own ideas. High income is less important—even though the conditions for monetary incentives are regarded as more favorable in medicine compared to many other subjects (cf. e.g., Gläser and von Stuckrad 2013; Krempkow et al. 2013). Nevertheless high income is relevant. All groups feel that the prestige of the journals in which they publish is most important as well as some other aspects referring to publications. However, to scientific staff, the importance of the number of academic awards is almost above average and more important than to professors. The same goes for prize money. This shows that awards are important in the scientists' perception of reputation, but they are less important than publication aspects. In their view, reputation depends on publication.[31] We conclude from our findings that research and teaching awards have a high potential to supplement incentive systems. Therefore it could be worthwhile to take a closer look at the existing range of research and teaching awards as elements of incentive systems in other subject areas or other countries and to conduct further research.

A notable aspect of the stocktaking of research and teaching awards is the considerably low prize money of the majority of teaching awards in relation to research awards. Some of the respondents regard appreciable prize money of teaching awards as necessary to signal recognition with such awards.[32] A first step in this direction was the Ars legendi prize of the HRK for excellent university teaching whose prize money (50,000 €) is considerably higher than that of any other awards previously awarded in Germany in this area. It was first awarded in 2006 within the framework of the HRK annual meeting (for medicine, after that in other subjects as well). Further examples of teaching awards including higher prize money have been found such as the Leuphana teaching award including 25,000 €,[33] the Freiburger Universität teaching

(continued)

[31] Multivariate analyses (that we have not introduced here) show, furthermore, that the goal attainment of incentive systems is positively related to the reputation relevance (cf. Krempkow et al. 2013).

[32] In an unpublishes analysis of case studies of the implemetation of incentives in German universities was found, that the background of the implementation of teaching awards often was to find a balance with the existing research incentives (cf. Dohmen et al. [in preparation] and to other project results cf. Dohmen and Henke 2012).

[33] The total amount could be divided into 10 awards of 2,500 € each.

award[34] of 25,000 € per year and the Baden-Württemberg state teaching award[35] of 50,000 € per award (for each type of university).[36] The prize money of the more recent examples is more or less in line with that of other countries such as the "New Australian Awards for University Teaching" in Australia.[37] However, these teaching awards including higher prize money in Germany are exceptional. Therefore it is to be assumed that the majority of teaching awards is currently unfolding a low incentive and reputation effect (cf. Frey and Gallus 2014; Jorzik 2010) in particular compared to research awards. Therefore, it would stand to reason to see a possible further development in higher prize money of teaching awards.

A greater distribution and higher prize money of teaching awards (comparing to research awards) does not seem to be realistic in the foreseeable future. This seems to be the case in particular for awards due to the available amounts of prize money. There are also conceptual reasons against too high prize money due to expectable non-intended negative effects (cf. Frey and Gallus 2014; Wilkesmann and Schmid 2010; for a more detailed discussion Krempkow 2010). It may therefore seem appropriate to use awards consistently as non-monetary incentives in addition to other monetary or non-monetary incentives. This could be done through an official award-giving ceremony which enhances the reputation of those being awarded (cf. also Frey and Gallus 2014; Webler 2010; Wilkesmann and Schmid 2010), or through increasing media coverage by collaborating with large organizations (whose prestige, if any, could also be used). This does not always require large financial resources. In this context prize money would then be a token of appreciation and would reach its objectives (primarily) through the reputation rather than financial aspects.

(continued)

[34] This can be divided in up to 10 awards; it is awarded for outstanding lectures, long-standing teaching on a high level of content and didactics (proven by evaluation), and innovative teaching concepts.

[35] This state award "for particularly good achievements in teaching" was already established in the middle of the 1990s. According to the Ministry of Science, Research and Arts Baden-Württemberg (MWK Baden-Württemberg) it "not only led to a fundamental discussion of the subject of teaching and the criteria of good teaching, but also encouraged a competition of ideas and implementations of good teaching both within and across universities and thus contributed to an increase in quality altogether". It allows individuals, or working groups of up to five members or entire organizational units in charge of teaching, e.g., faculties, institutions, and seminars, to be put forward (MWK BaWü 2009, the same as MWK BaWü 2012).

[36] In comparison the prize money for the "lecturer of the year" was 5,000 €.

[37] "New Australian Awards for University Teaching": $ 50,000 (Prime Ministers Award "Teacher of the year"), $ 25,000 (40 awards, corresponds to the approx. number of universities) up to $ 10,000 (210 awards: approx. number of faculties).

A broader effect might also be reached by establishing awards for groups of academics (be it for groups of researchers or groups of teaching staff).[38] In this way it would be possible to promote competition between teams and at the same time to promote collaboration within teams to avoid so-called "go-it-alone strategies".

References

Becker FG, Tadsen W, Wild E, Stegmüller R (2012) Zur Professionalität von Hochschulleitungen im Hochschulmanagement: Organisationstheoretische Erklärungsversuche zu einer Interviewserie. In: Wilkesmann U, Schmid C (eds) Hochschule als Organisation. VS – Verlag für Sozialwissenschaften, Münster, pp 191–205

Chan HF, Frey BS, Gallus J, Torgler B (2013) Does The John Bates Clark Medal boost subsequent productivity and citation success? Working paper no. 2013-02, CREMA – Center for Research in Economics, Management and the Arts, Zürich

Cremer-Renz C, Jansen-Schulz B (eds) (2010) Innovative Lehre – Grundsätze, Konzepte, Beispiele. UniversitätsverlagWebler, Bielefeld

Dohmen D, Henke J (2012) Wettbewerb durch leistungsorientierte Mittelzuweisungen? Zur Wirksamkeit von Anreiz- und Steuerungssystemen der Bundesländer auf Leistungsparameter der Hochschulen. die hochschule 2:100–120

Dohmen D (ed) QualitAS-Lehre. Theorie und Praxis von Anreiz- und Steuerungssystemen im Hinblick auf die Verbesserung der Hochschullehre. Endbericht. FiBS - Forschungsinstitut für Bildungs- und Sozialökonomie, Berlin (in preparation)

Frey BS (1997) Not just for the money: an economic theory of personal motivation. Elgar, Cheltenham-Brookfield

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

Gläser J, von Stuckrad T (2013) Reaktionen auf Evaluationen. In: Grande E, Jansen D, Jarren O, Rip A, Schimank U, Weingart P (eds) Neue Governance der Wissenschaft: Reorganisation, Externe Anforderungen, Medialisierung. Transkript, Bielefeld, pp 73–93

Grande E, Jansen D, Jarren O, Rip A, Schimank U, Weingart P (eds) (2013) Neue Governance der Wissenschaft: Reorganisation, Externe Anforderungen, Medialisierung. Transkript, Bielefeld

Heckhausen J, Heckhausen H (eds) (2010) Motivation und Handeln. Springer, Heidelberg

Hellemacher L, Knobloch T, Stelzer-Rothe T (2005) Die Bewertung der Erfolgsfaktoren von Fachhochschulen. Eine empirische Studie des hlbNRW. Die Neue Hochschule 4(5):40–43

Herzberg FI (1966) Work and the nature of man. Crowell, New York

Hochschild AR (1998) Der Arbeitsplatz wird zum Zuhause, das Zuhause zum Arbeitsplatz. Total Quality Management bei der Arbeit, Taylorismus im Familienalltag – kann das auf Dauer gut gehen? Harv Bus Manag 3:29–41

Hornbostel S (2002) Ehre oder Blechsegen? Das Auszeichnungswesen der DDR. In: Best H, Gebauer R (eds) (Dys)funktionale Differenzierung? Rekrutierungsmuster und Karriereverläufe von DDR-Funktionseliten. SFB 580 Mitteilungen (3), Jena/Halle, pp 33–39

[38] By bestowing awards on only selected agents, the principal runs the risk of affronting those who are not awarded. The danger of negative effects is particularly high in small and homogenous groups (Frey and Gallus 2014). Another reason might it be that research (and partly teaching) is in some fields (nowadays) conducted in groups and thus group performance should be honored as such.

Jorzik B (2010) Viel Preis, wenig Ehr'. Lehrpreise in Deutschland. In: Tremp P (ed) Ausgezeichnete Lehre! Lehrpreise an Universitäten. Erörterungen – Konzepte – Vergabepraxis. Waxmann, Münster, pp 117–140

Krempkow R (1999) Ist "gute Lehre" meßbar? - Untersuchungen zur Validität, Zuverlässigkeit und Vergleichbarkeit studentischer Lehrbewertungen. Tectum, Marburg

Krempkow R (2007) Leistungsbewertung, Leistungsanreize und die Qualität der Hochschullehre, Konzepte, Kriterien und ihre Akzeptanz. UniversitätsverlagWebler, Bielefeld

Krempkow R (2010) Lehrpreise im Spannungsfeld materieller und nichtmaterieller Leistungsanreize. In: Cremer-Renz C, Jansen-Schulz B (eds) Innovative Lehre – Grundsätze, Konzepte, Beispiele. UniversitätsVerlagWebler, Bielefeld, pp 51–71

Krempkow R, Landrock U (2013) Wie effektiv ist die Leistungsorientierte Mittelvergabe? Zwischenbilanz ein Jahrzehnt nach ihrer Einführung in der Hochschulmedizin. In: Grande E, Jansen D, Jarren O, Rip A, Schimank U, Weingart P (eds) Neue Governance der Wissenschaft: Reorganisation, Externe Anforderungen, Medialisierung. Transkript, Bielefeld, pp 95–111

Krempkow R, Landrock U, Schulz P, Neufeld J (2011) Die Sicht der Professor/innen auf die Leistungsorientierte Mittelvergabe an Medizinischen Fakultäten in Deutschland. Auswertung der Onlinebefragung des Projektes GOMED – Governance Hochschulmedizin. iFQ, Berlin. Online: www.forschungsinfo.de/Publikationen/Download/LOM_Professorenbefragung.pdf

Krempkow R, Landrock U, Neufeld J, Schulz P (2013) Intendierte und nicht-intendierte Effekte dezentraler Anreizsysteme am Beispiel der fakultätsinternen leistungsorientierten Mittelvergabe in der Medizin. Abschlussbericht des Projektes GOMED – Governance Hochschulmedizin. IFQ, Berlin. Online: www.forschungsinfo.de/Projekte/GOMED/GOMED-Abschlussbericht.pdf

Minssen H, Wilkesmann U (2003) Folgen der indikatorisierten Mittelverteilung in nordrhein-westfälischen Hochschulen. Beiträge zur Hochschulforschung 3:106–129

MWK BaWü (2009) Ausschreibung zum Landeslehrpreis 2009. Brief des Ministers für Wissenschaft und Kunst Baden-Württembergs an die Rektorinnen und Rektoren der Universitäten und Hochschulen vom 8.4.2009

MWK BaWü (2012) Ausschreibung zum Landeslehrpreis 2012. Brief des Ministers für Wissenschaft und Kunst Baden-Württembergs an die Rektorinnen und Rektoren der Universitäten und Hochschulen vom 19.01.2012

Neckermann S, Cueni R, Frey BS (2012) Awards at work. ZEW discussion paper no. 12-004, Zentrum für Europäische Wirtschaftsforschung, Mannheim

Oelkers J, Strittmatter A (2004) Leistungslöhne für Lehrkräfte? Pädagogik 5:48

Osterloh M, Frey BS (2008) Anreize im Wissenschaftssystem. Universität Zürich und CREMA – Center for Research in Economics, Management and the Arts, Switzerland

Tremp P (ed) (2010) Ausgezeichnete Lehre! Lehrpreise an Universitäten. Erörterungen – Konzepte – Vergabepraxis. Waxmann, Münster

Webler WD (2010) Qualitätsfördernde Wirkungen individueller Lehrpreise? Qualität in der Wissenschaft 2:52–57

Wilkesmann U, Schmid C (2010) Wirksamer Anreiz? Forschung und Lehre 7:504–507

Wilkesmann U, Schmid C (eds) (2012) Hochschule als organisation. VS – Verlag für Sozialwissenschaften, Münster

Winter M, Würmann C (eds) (2012) Wettbewerb und Hochschulen. 6. Jahrestagung der Gesellschaft für Hochschulforschung in Wittenberg 2011. die hochschule 2:6–324

Witte J, Schreiterer U, Müller-Böling D (2001) Gestaltungsfragen bei der Umsetzung des Professorenbesoldungsgesetzes. CHE-Arbeitspapier Nr. 37, Gütersloh

Wollersheim J, Welpe IM, Ringelhan S (2014) Sein und Sollen: Bewertung von Forschungsleistungen in den Wirtschaftswissenschaften. Forschung und Lehre 4:276–278

Ziegele F, Handel K (2004) Anreizsysteme im Hochschuleinsatz. In: Benz W, Kohler J, Landfried K (eds) Handbuch Qualität in Studium und Lehre. Raabe, Berlin, pp 1–22

Part IV
Innovation and Creativity: Fostering and Impeding Conditions

Myths, Challenges, Risks and Opportunities in Evaluating and Supporting Scientific Research

Martin Quack

Abstract We summarize and discuss critically the various procedures used for decisions in evaluating and funding scientific research in view of the goal of obtaining the best research with a given funding volume. Merits, difficulties, and problems in some procedures are outlined. We identify a number of myths appearing frequently in the process of research evaluation. We indicate some similarities in deciding on funding research projects, making appointments to academic and leading research positions and selecting prize winners for high prizes for scientists. Challenges, risks and opportunities in research and its funding are identified.

1 Introduction

The present short essay on some challenges, risks and opportunities in evaluating and funding of scientific research is based on the "president's speech" given in Leipzig on the occasion of 111th meeting of the Bunsen Society in Leipzig, of which printed records exist [in German, (Quack 2012a, b)]. Our goal is to summarize and discuss critically the various procedures that are currently being used and identify their strengths and weaknesses. We identify some myths that are circulating in the community involved in evaluating scientific research and the resulting dangers from an uncritical belief in these myths. We conclude with a general discussion of the role of scientific research in terms of its value to the individual and to the society. The topic of our essay should be of obvious importance for all those being actively involved in scientific research, but also to mankind as a whole. Besides the enormous opportunities there are also hidden risks and dangers which we want to discuss.

M. Quack (✉)
Physical Chemistry, ETH Zürich, CH-8093 Zürich, Switzerland
e-mail: Martin@Quack.CH

I.M. Welpe et al. (eds.), *Incentives and Performance*,
DOI 10.1007/978-3-319-09785-5_14

2 Some Basics in Evaluating and Funding Scientific Research

We shall discuss the following main questions.

1. Who decides upon funding research? (Institutions, committees, bureaucracies, individuals as sponsor or as Maecenas)
2. How does one decide, what should be funded? (Procedures and criteria used).
3. What is good practice in evaluating and funding scientific research?
4. What is the goal of research funding? (Discoveries and new knowledge, future income and profit)
5. What is the goal of scientific research in a broader context?

Let us first consider the funding of research by institutions. Well known examples are in the USA the "National Science Foundation" (NSF), in Great Britain the "Engineering and Physical Sciences Research Council" (EPSRC), in Germany the "Deutsche Forschungsgemeinschaft" (DFG), in Switzerland the "Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung" (SNF, also "Swiss National Science Foundation" or "Fonds National Suisse", FNS, and SNFNS as Acronym), in Austria the "Fonds zur Förderung der wissenschaftlichen Forschung" (FWF) and in France with a somewhat different structure the "Centre National de Recherche Scientifique" (CNRS) or in the European Union the relatively new and quite remarkable "European Research Council" (ERC) besides many more worldwide, of course.

My closest personal contact is obviously to the Swiss National Science Foundation, where I like in particular the mentioning of the *goal* "zur Förderung der wissenschaftlichen Forschung" ("for the support of scientific research") in the German name of the institution, because all too often such institutions tend to forget their goals.

It would be of obvious interest to discuss the quality of the various institutions mentioned, but we shall refrain from this in order to avoid unnecessary embarrassments. Rather we shall turn now more generally to the procedures used by such institutions in decisions on funding research projects. Similar procedures are also used in decisions on appointments for higher academic and research positions or in attributing higher prizes for scientific research. Indeed, such decisions can be considered as providing support for scientific research in a broader sense as well. We can distinguish here the following main types of procedures:

A. The decisions are made by a committee ("panel") of competent persons, who are themselves actively involved in scientific research of the general field considered, covering at the same time a rather broad range of expertise. The decisions of the committee are based on the careful study of written research proposals, often obtaining more detailed confidential assessments from external experts in the particular topic of the research proposed.

B. The decisions are made by a group of bureaucrats, which uses various combinations of indices and other measures as well as sometimes expert assessments, or simply "gut feeling" ("Bauchgefühl").
C. The decisions are based on pure chance ("Lottery").

Of course, there exist various mixed and intermediate forms of the three basic types of decision making in research funding. Also, if there is sponsoring by an individual sponsor and not an institution, the sponsor might simply base the decisions on "personal taste" ("gut feeling"), but more frequently some form of decision following the three main types mentioned is chosen also by individual sponsors.

Let us turn first to point C, the "pure chance choice" which may appear to some as a joke. However, there are serious proposals to base decisions on research funding on a random choice. We can cite here Les Allen following Buchstein (2011), where one can also find further discussion of this topic: "I suggest that the Engineering and Physical Science Research Council throw out the panels, throw out the referees and have a lottery for all the available funds. Such a system would be fairer then the present one and would also be better at supporting truly original research. Pure chance must give more hope than the opinions of a subset of my peers" (Les Allen). We might mention also the discussion of a random choice selection in the publication of journal articles by Osterloh and Frey (2012) in this general context (Osterloh and Frey 2012), see also (Osterloh 2013).

An argument which is frequently advanced in favour of the "random choice selection" procedure is its alleged "fairness" (or doing proper "justice").

It is certainly true that with an honest procedure, the random selection is in a certain sense "fair" (nobody is preferred). However, by those being subjected to such a procedure luck is perceived as "fair and good" only if it is "good luck" and rather as bad and unfair if it turns out to be "bad luck". We shall not waste too much time on this procedure, as it appears to us as obvious nonsense in the present context, and even immoral in a certain sense. We can make the following comparison: What would our students say if we simply made a choice of their marks by a lottery instead of basing the marks (including the decision of failure or success) on a careful expert evaluation of examination results? They would certainly complain justly about what would appear as an arbitrary, truly immoral, if very simple and time saving procedure.

Another argument sometimes given in favour of the random choice selection is the well-known bias in the alternative "expert selection", due to experts favouring "mainstream" as opposed to "revolutionary" (Kuhn 1962) projects or results. We think that this argument misses an essential point in the evaluation by a good committee of experts: A "good expert" is fully aware of such a bias and takes it into account giving room in the decisions on funding to what one might call "high risk" projects. Of course, even this does not exclude expert errors, of which there are many examples. Still, a well informed and conscientiously made expert selection appears to be a far better option than the random choice selection. We can give another example for this from daily life. With a serious illness we would consult an expert medical doctor, in fact, the one with the best possible medical education and medical record and rather trust his diagnosis and therapy (perhaps after getting a second opinion from a further expert). We do so in perfect knowledge of the existence of errors of expert medical doctors. We still prefer this approach simply because of the alternative of a "random choice" selection of some arbitrary person (or even a randomly chosen medical doctor) for a diagnosis and therapy is completely unreasonable (and also definitely less hopeful and successful).

Indeed, the random choice selection is rarely used purposefully although we shall discuss below that it plays sometimes involuntarily an important role in mixed forms of "combination procedures". Most frequently the methods A. and B. of evaluation and decision taken by some committees and persons of various origins are used in practice.

Thus now we address the further question of the *criteria* used in making decisions on funding particular scientific research projects or also in selecting persons for tenured academic positions or as prize winners of high academic prizes. Here we can distinguish the following main criteria

a. The quality of the research project proposal, most frequently as assessed by detailed expert evaluations.
b. The personality of the researcher as assessed by the past achievements, research record and publications, and further information such as personal (oral) presentation of the research, interview with a committee of experts etc. The president of the Humboldt foundation Helmut Schwarz has expressed this in a concise fashion as "Fund people, not projects" (Kneißl and Schwarz 2011). Evidently, this approach is mostly used in academic appointment decisions.
c. Bureaucratic indices such as citation indices, total numbers of citations, h-index (Hirsch 2005), past funding record, (such as sum of previously acquired funds and number of previously funded projects), number of participating researchers or research groups in larger collaborations, number of previous publications, perhaps weighted by the impact factor of the journal, the position as "first author" or "last author", sometimes only counting publications in "Science" or the like.

According to my experience, these three main criteria are used with widely varying weights by the various institutions and committees. We state this fact at this point without any further judgement. In the section II (mathematical and physical-chemical sciences) of the Swiss National Science Foundation, where I was active as member of the research council for about 10 years, one uses mainly the criteria a. and b. (perhaps fortunately so), whereas in more recent years arguments resulting from the criteria c. are mentioned in the discussion but without much weight for the final decisions. One might mention here that the members of the Swiss National Research Council are active research scientists, carefully selected from the members of the Swiss research community. They act in their function in the research council only for a limited number of years and contributing only a limited fraction of their time to this task. This is to be seen within the Swiss political culture of the "Milizsystem" ("militia system"), where, for instance, in principle, every Swiss citizen is a “part-time soldier” for a limited time in the year and period of their life, officers being selected from this pool as well. At the same time they remain active in their ordinary life in whatever profession they may have. Similarly, the members of the Swiss National Science Foundation’s Research Council remain regular active scientists in universities, research institutes etc. during their period of service on the council. Such a committee is by its very nature more influenced by scientific arguments and much less so by bureaucratic indices. However, we know also of other institutions and committees where the reverse is true and the criteria under c. can become dominant, indeed. Between these two limits one can find many intermediate situations, in practice.

After having introduced now already some judgement into our discussion, we shall reinforce these by mentioning some serious and increasing problems. We shall do so by referring to what we might call some “myths” propagating in the evaluation and funding of scientific research as well as in selecting researchers for academic and other positions or for prizes.

3 Some Myths in Evaluating and Funding Scientific Research

Myth 1 High rejection rates R (in funding schemes, journal publications etc.) are a measure of high quality of the procedure (“competitiveness”). The nonsense of this rather widespread belief can be easily established by considering the limit $R \to 1$, where everything is rejected, nothing accepted (for funding, publication etc.). This would be then the ideal procedure, where people write proposals and papers, expert reviewers assess them and committees decide in the light of these expert reports to accept in the end . . . nothing. This would be of course in reality the complete idling and waste of time and effort. While the nonsense of such procedures is obvious, also in the cases of very large rejection rates R close to 1 (but $R < 1$), the procedures are of poor quality, by general experience. What happens in such situations is that in the end among a large number of, in principle, worthy projects only a few are selected by pure chance, a random choice selection, because other valid criteria are not available. While this is inefficient (and perhaps even immoral, see above) as one might as well use a lottery instead of an assessment in these cases, it frequently happens involuntarily in procedures with very high rejection rates. Of course, also the other limit of no rejections ($R = 0$) will not in general lead to an adequate and responsible use of research funds.

It is not possible to define a “correct” rejection rate R for an optimal funding scheme. In principle, it would be correct to fund all good projects and reject all poor projects. The corresponding rejection rates would depend on the research field, the research culture and tradition in a given field, but also the political context, the country and so forth. From my experience in physical-chemical research in a European context values of $R = 0.4 \pm 0.2$ would fall in a reasonable range and with $R > 0.8$, where less than 20 % of the projects are funded, the quality of the selection process degrades rapidly and drastically.

Myth 2 Citation numbers are an adequate measure for the importance of a scientific publication. Experienced scientists know from numerous examples that this statement is pure nonsense. We shall illustrate this here with just one prominent example selected by R. N. Zare (2012a, b). The paper "A Model of Leptons" by Steven Weinberg (1967) has contributed most importantly to the current "Standard Model" of Particle Physics (SMPP) and has also contributed to the Nobel Prize awarded to Weinberg. It has even importance in relation to our physical-chemical understanding of molecular chirality (Quack 2011a). However, according to Zare (2012a, b), the paper by Weinberg (1967) was not cited at all in 1967 and 1968 and just once in 1969 and 1970 each (1971 it had 4 citations including one self-citation). This implies that this truly important paper has not contributed to the "impact factor" of the corresponding journal (Phys. Rev. Letters), and if the editor had taken this "measure" as a criterion for selecting papers, he should not have accepted the publication. Also, this paper would not have contributed to funding or tenure decisions for Weinberg, if the relevant people in these decisions had looked for this measure (they fortunately did not). There are numerous similar examples from many different fields, even if not all of these end up in a Nobel Prize. Molinié and Bodenhausen have generated a graphics for some "classics" of NMR spectroscopy (Molinié and Bodenhausen 2011). While today (long after the Nobel prize) Weinberg's paper is highly cited (more than 5,000 citations) this fact is irrelevant with respect to the period for further funding decisions (and in many cases also further tenure decisions) which would have been typically the period of 1967–1970 in the case of the author of this 1967 paper. Today the citation numbers for the paper are irrelevant for further funding of this research. Many fallacies of the citation analysis have been discussed over the years including "folk citations" for non-existing authors such as S. B. Preuss (Straumann 2008; Einstein 1931; Einstein and Preuss 1931). Another aspect of showing the nonsense in just counting citations is the neglect of the "sign" of the citation. A publication can generate high numbers of citations because it is criticized by many as erroneous (perhaps even with forged results). This "lack of sign" problem in the "impact" has been discussed in a humorous way by Petsko (2008), but it is a serious matter and we know numerous "real" examples from experience which leads us now to a closely related myth.

Myth 3 The "impact factor" of a journal as derived from the citation statistics of the first years after publication is a measure of the quality of the journal (*Science*, for example, with its enormous impact factor, would then be an outstandingly good journal). The example cited under myth 2 already demonstrated that the very basis of such a statement as derived from the citations of an individual paper in the journal would be erroneous. However, one can find statements that the cumulative use and assembly from many different individual papers in a journal in the end leads to the "impact factor" being a meaningful measure of its quality. Again, experienced scientists know that this is wrong. In the author's own field physical chemistry and chemical physics, for instance, there are good journals with very modest "impact factors" (J. Chem. Phys., PCCP, J. Phys. Chem, Mol. Phys., etc.) compared to *Science*, which is a journal of very questionable quality. Even if one does not subscribe to the nasty comment that "the condition for a paper to be published in *Science* is that it is either stolen or wrong", there would be many experts, at least in some of our fields of research, who would agree with the milder statement that the relation of Science to the four journals mentioned for this particular field (and related cases in other fields) is similar to the relation of quality of newspapers like "Blick" or "Bildzeitung" to the "Neue Zürcher Zeitung" or "Frankfurter Allgemeine Zeitung" to give examples from the German speaking areas (there are analogues for English, French etc.)

As examples for wrong results published in *Science* we might mention the interesting critical discussion of such an example by Volkmar Trommsdorff (2002) or the wrong paper on ortho- and para-water by Tikhonov and Volkov (2002), commented upon by Albert et al. (2006) and Manca Tanner et al. (2013), besides many more examples. Well known and particularly serious examples are the many wrong papers published by H. Schön et al. in *Science* (a case of "wrong" even by forgery).

Of course, there are also journals of good quality with relatively high impact factors, although not the very highest, such as "Angewandte Chemie". Thus a high impact factor does not necessarily imply low, "Boulevard Journal" type quality. There is in fact no simple relation between quality and the impact factor of a journal. There exist good and bad journals with low and high impact factors. The next myth to be discussed is even more serious in that it deals with individual scientists.

Myth 4 The so-called h-index (Hirsch-index) is a suitable measure for the importance or quality of a scientist. Hirsch (2005) has introduced this bibliometric measure and has made such a claim and has, indeed, seriously proposed to use the h-index in professional appointment and tenure decisions. The nonsense in such a statement is again well known to experienced scientists involved in such decisions and we shall return to the question of academic appointments below. Here, we shall cite the critical and very competent discussions by Molinié and Bodenhausen (2010a, b, 2011), as well as by Richard Ernst (2010a, b), who provide ample evidence rejecting Hirsch's proposal. Thus, without going into more detail of this absurd "quality measure for scientists", we shall turn to a further myth widely promoted in the science bureaucracy.

Myth 5 The amount of research funding acquired by a scientist (or a group of scientists) is a good measure for the corresponding importance (or quality) of the researchers. One might express this "Funding importance" F_I by Eq. 1.

$$F_{\mathrm{I}} = \frac{\text{Sum of aquired research finances}}{\text{Number of participating scientists}} \tag{1}$$

Of course, such a number can be easily derived for every researcher or research group, thus its popularity, and "the higher the F_I the better the research group". However, after giving some thought to this matter, one quickly comes to the conclusion that for an optimal use of research funds one should rather use a measure (if any), where the sum of research funds used would appear in the denominator, and one might call this the *research efficiency* R_E.

$$R_E = \frac{\text{Scientific knowledge generated}}{\text{Sum of finances (acquired and used)}} \quad (2)$$

We may quote here Martin Suhm (2010) (freely translated): "It would not be misleading, if the sum of acquired funds appeared in the denominator, the place, where it should appear in the name of efficiency and sustainability, instead of appearing in numerator of such measures". Of course, the bureaucratic use of Eq. (2) is hampered by the "Scientific knowledge generated" not being measurable by some number, unless one replaces it by "numbers of publications generated" or "numbers of citations generated", which is, indeed, sometimes done, but would be nonsensical as discussed above for the myths, 2, 3, and 4. Thus, if Eq. (2) is to be used in a sensible way, one has to interpret it as a symbolic, not as a numerical equation and thus no simple number can be generated from it for some rankings.

We shall conclude this section on the various myths propagated in decisions on supporting scientific research by some comments on the particular dangers arising from the use or rather abuse of bibliometric data in this context. Indeed, in recent years we are increasingly confronted with this abuse, be it by the science bureaucracy or by scientists themselves believing in bibliometry. I can quote here from personal experience from an expert report in an appointment procedure for a professorship (in anonymized and somewhat altered form for reasons of confidentiality) "... in our country bibliometric counts are most heavily weighted". The "expert" was from a country in Northern Europe and in the end drew his conclusions based on bibliometric data for the candidates to be evaluated. Fortunately, the appointment committee considered this particular expert report as irrelevant (for several good reasons) and did not take it into account in the final decisions. One must be afraid, however, that some poor appointment committees would follow such poor advice. Indeed, uncritical belief in bibliometry can be occasionally found in serious publications supported by major academies [see, for instance, (Gerhards 2013)] and also with some active scientists who contributed important work in their own specialty. From one such scientist, I heard the comment that there is "no objective alternative" to bibliometry. Again this is pure nonsense. An obvious alternative has been formulated by Richard Ernst (2010a, b) in his most relevant article. "And there is, indeed, an alternative: Very simply start reading papers instead of merely rating them by counting citations". Of course, following such an advice requires time and knowledge of the subject, and the bureaucracy lacks the knowledge and does not want to invest time.

We shall discuss now further, related alternatives, in relation to the very important question of procedures in appointments for professorships or other higher academic and research positions.

4 Criteria Used in Academic Appointments

Good appointments of academic positions (professorships or research positions) at universities and research institutions are among the most important ways of supporting scientific research. They have a long term effect and are truly efficient investments for the institution as for science overall, if they are carried out successfully. We can cite here almost literally from Zare (2012a, b) for what we might call a summary of criteria used in appointment procedures with good academic practice (here for the example of the Chemistry Department at Stanford University):

1. First of all they must be good departmental citizens.
2. Second they must become good teachers.
3. The Department wants them to become great researchers (This last criterion is the most difficult). We ask experts, whether the research of the candidate has changed the view of the nature of chemistry in a positive way. ... it is not based on the number of papers, with an algorithm on impact factor, etc. ... do not discuss h-index metrics ... do not count publications or rank them as to who is first author. We just ask: has the candidate really changed significantly how we understand chemistry?

From my experience in presiding appointment committees at ETH in many departments as a delegate of the president (for more than 15 years), I would add that similar considerations prevail at ETH Zürich, even though every once in a while attempts are made to introduce bibliometry and the like into the discussion, but without much effect. Particularly the younger people looking for an appointment tell me, however, that they in fact know universities, where bibliometric and similar data are used importantly, even dominantly, in professional appointments. My reply to this is, sure, there exist also poor universities, and if I am then asked how to tell apart the good from the poor universities, I would answer certainly not by using bibliometry and "rankings", but for instance, by looking at the procedures they use in professional appointments, among other criteria. Of course, even good procedures do not exclude occasional erroneous decisions. Prominent examples include a Nobel prize given for a wrong discovery (see Quack 2013), although the person selected deserved the prize, but obtained it for the wrong reason.

That the criteria 2. and 3. concerning teaching and research are important for appointing professors at universities can appear as self-evident. The first criterion, requesting a "good citizen" might come as a surprise to some, and might support a suspicion that only "well adjusted" candidates are selected. This, however, is not the meaning of this requirement for a "good citizen". It arises from the often painful experience that "bad citizens can damage good science". This problem is frequently covered with silence in the scientific community, or given only minimal weight. It is, however, a very serious problem because the damage caused to scientific research by some "bad citizens" can be huge, directly and indirectly. Bad behavior can appear as a straight-forward, criminal fraud and forgery of research results, and then the damage caused is much larger than any potential advantage hoped for by the criminal (see, for example, Pfaltz et al. 2009). However, cases of fraud and deception towards a partner in cooperations also exist. One such case, the quarrel between O. Piccioni and E. Segré in the discovery of the antiproton, was brought to the courts and reached a wider public, with enormous indirect damage for the reputation of scientific research (Heilbron 1989). That Segré did a severe wrong to his colleague Piccioni can hardly be subject of doubt. This wrong was not punished, however, perhaps it was even rewarded, which puts this area of science in this period in a somewhat dubious light. That the problem reached the public is a rare exception. Most of the time, such events in scientific research are covered with silence. I shall not mention such a case from physical chemistry, as the aim of this section of our essay is not to sing a "dies irae". In principle, a request for "good citizenship" in the republic of science is not specific to research or science, it is rather a generally valid principle in human relations.

ETH Zürich has a motto, which fits in this context "Prima di essere ingegneri, voi siete uomini". It was formulated by one of its first professors, Francesco de Sanctis (1817–1883) in his inaugural lecture. Correct human behavior supports scientific research by preventing damage, among other things. However, also reliability of universities in establishing contracts with their professors and keeping long term promises and escaping the temptations of later breaches of contracts are important elements in supporting research, which in recent times have been increasingly eroded even in the best institutions. Fundamental research, though, needs adequate freedom (Kneißl and Schwarz 2011) provided by the generous and reliable appointment contracts at the top universities. Freedom of teaching and research is the most important pillar of science, innovation and creativity, which are also strengthened by reducing bureaucracy. Indeed, the incessant and unchecked growth of bureaucracy is one of the greatest risks in the current support of scientific research, a risk which can cause great damage (Szilárd 1961; Quack 2013). Again, this phenomenon is not restricted to science, but it is a general phenomenon in modern society. Much has been said and written about this and Fig. 1 shows results from a classic in this field, here with the growth of bureaucracy in the British colonial administration following Parkinson's law (Parkinson 1957). Parkinson's Law has given rise to many joking comments; it is, however, a serious matter. Also the growth of cancer cells follows such a law of growth until it is ended by a catastrophe. An analysis of the growth of the staff in science and university bureaucracy shows close analogies upon which I will not further comment, here. I do not intend to enter here into a general bashing of university administration. Indeed, there exists also the truly "good administration", which serves and supports science. However, the staff in this part of administration does not grow and its relative importance rather decreases. We have elsewhere identified the general growth of bureaucracy as one of the great risks of mankind, besides nuclear war and climate change (Quack 2011d, 2013), and shall conclude this discussion here with this brief remark, turning finally to some aspects of opportunities of scientific research, rather than risks, after a brief summary of what might be considered good practice.

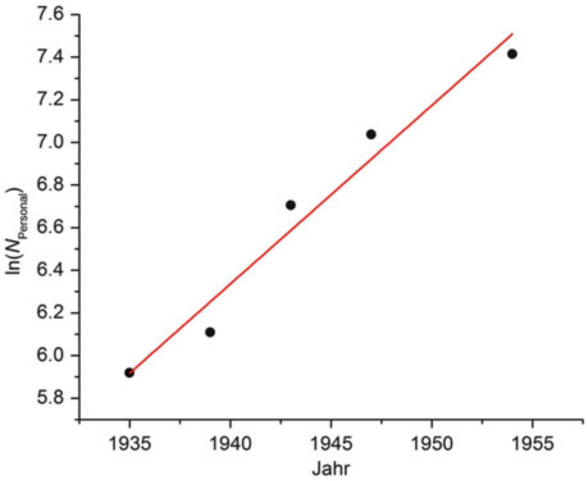


Fig. 1 Logarithm of the number of persons in the staff of the British Colonial Administration $\ln(N_{Personal})$ as a function of time [Following an exercise in the Kinetics course at ETH with data from Parkinson (1957)]

5 A Brief Summary of Good Practice

While implicitly contained in the text of the previous sections, it may be useful to summarize here what may be called the "good practice in evaluating and funding scientific research and in making academic appointments". We mention advantages of this "good practice", but also the associated difficulties (Diederich 2013). A good procedure for distributing available research funds and making academic appointments can be summarized by the following steps:

Bring together a group of competent and trustworthy experts covering a sufficiently broad range of the field under consideration in order to avoid too narrow a view and to neutralize possible conflicts of interest.

Get every research proposal and person looked at in detail by the relevant experts and the whole group, obtaining more detailed specialized outside reviews, if necessary.

Have the group discuss every individual case in depth before coming to a conclusion by the group as a whole, by consensus or by vote, if necessary.

This procedure is not new and it is used by good funding institutions such as the SNF or appointment committees at good universities, academies as well as prize committees for major prizes. It has many advantages and minimizes the chances for severe errors, although even the best procedure cannot exclude occasional errors completely. It has one major disadvantage: It is time consuming and costly (Diederich 2013). It also requires the cooperation of experts, sometimes difficult to obtain. These disadvantages have led some institutions to use one or another shortcut as mentioned under the 5 myths. However, using such shortcuts should be considered foolish or even fraudulent, as it replaces expert knowledge by bureaucratic superstition and forgery. It would be as immoral as the "shortcut of the bad citizen" replacing serious experimentation and carefully analyzed data in scientific research by some invented results, when proposing or testing a hypothesis.

Sometimes it is argued that using statistical data, say from bibliometry, can be justified by some perhaps existing correlation with "real data". For instance, it is claimed that really good scientists (as judged by a careful evaluation as outlined above) statistically have a higher h-index, than scientists of lower quality. While one may have some doubts, whether this correlation really exists, even if so, it would be at most very rough, with numerous easily proven "outliers". On the other hand, the decisions to be taken (in deciding on a proposal or an appointment) are often very important decisions on individual cases. A rough statistical correlation (even if it existed), is of no use in this case. We may use here the analogy of evaluating written examination papers: From long experience we definitely know that there is, indeed, a very rough (but not tight) correlation between the number of pages in a written solution of an examination paper and the final examination result, as carefully assessed by an expert examiner. The more pages, the better the results, statistically. However, we also know that there are many outliers, some brilliant solutions are very short, and sometimes also many written pages of an exam paper contain just erroneous solutions, thus a poor final result. Using the "time saving short cut" of simply counting written pages would correspond to a fraudulent and immoral procedure of the examiner. There are many analogies of this example with the use of bibliometric or other indices (for instance neither a page count nor a citation count or h-index needs an expert, if could be done by any administrative staff). No more needs to be added on such fallacious "short cut" procedures.

6 Why Scientific Research? The Opportunities for Creativity and Innovation

We should address also the general question as to why one should consider supporting scientific research at all. In a commencement speech of 2004 which in the meantime has been printed in several versions (see (Quack 2011b) and references cited therein), I have summarized some important reasons for fundamental scientific research:

1. For the personal satisfaction of discovery and knowledge.
2. As contribution to the edifice of knowledge of mankind, towards understanding the world as well as the human condition.
3. Directly and indirectly to contribute to improving the conditions for human life and of mankind—and for its survival.

The first reason is an important, intrinsic, subjective, personal motif of the researcher. The second and third reasons provide objective grounds, why society should support science financially and otherwise as an investment in the future of mankind and society.

The first, intrinsic motif was formulated already 2,400 years ago by Demokritos of Abdera (ca. 460–370 B.C.) in an inimitable way.

βούλεσθαι μᾶλλον μίαν εὑρεῖν αἰτιολογίαν ἢ τὴν
Περσῶν οἱ βασιλείαν γενέσθαι

He talks about the scientist-philosopher, freely translated here "He would rather make a single fundamental discovery than become the king of the Persians." We can also cite Rose Ausländer (2002) with one of her poems (imperfectly translated here)

You are
irresistible
Truth
I see you
and name you
Bliss

One can also translate these texts in a somewhat extended, completely free way as applied to the scientist and researcher (Quack 2011b, c, 2012a):

He would rather make and teach a single fundamental discovery than:

- become president of the United States.
- obtain the wealth and power of Bill Gates.
- build a large bomb.
- have 10 publications in Science—the magazine.
- to reach the top position in citations of scientists.
- have 100 presentations on TV.

One can also phrase this in the form of a "non-motivation" (Quack 2011b, c, 2012a):

Fundamental research: Why not?

1. Not to damage others.
2. Not to beat somebody in competition.
3. Not to have power.
4. Not to become rich.

These positive as negative points concern the personal motivation. There are, however, also the objective reasons concerning the service to mankind. This aspect is obvious with all forms of applied research, but is frequently forgotten, when fundamental research is considered. Nevertheless, fundamental research, innovation and creativity can be considered to be among the most important driving forces in improving the human condition (Perutz 1982). Indeed, the support of fundamental research can be considered to be the greatest opportunity of all investments of society and mankind in their future.

I shall illustrate this here with the Schrödinger equation, which is one of the fundamental equations for physics and chemistry (Schrödinger 1926a, b, c, d, e)

$$\mathrm{i}\frac{h}{2\pi}\frac{\partial\Psi(q,t)}{\partial t} = \hat{H}\,\Psi(q,t) \tag{3}$$

This equation was introduced by Schrödinger in 1926 in order to provide a deeper formulation and understanding of quantum theory (see also Merkt and Quack 2011). It was "pure research" in theoretical physics in its purest form, far removed from any practical goals or technical applications. Today there are estimates that about 20 % of the gross national product of modern industrial countries are based in a general way on applications of quantum mechanics. Equation 3 prevails in all applications of optical spectroscopy which range from the study of the Earth's atmosphere to combustion in car engines and industrial processes (Merkt and Quack 2011; Quack and Merkt 2011). We can also mention here nuclear magnetic resonance spectroscopy (NMR) with applications, for example, in MRI (Magnetic Resonance Imaging) available today in hospitals all over the world to name just these selected examples (Ernst et al. 1987; Ernst 1992) among many more.

Another example, comparable to the development of quantum mechanics in the twentieth century can be named with the development of electricity and magnetism in the nineteenth century. Technical applications of these are visible everywhere in our daily life today. However, the original developments were made long before these uses were obvious, although such future uses were predicted by Faraday with great foresight. He is reported to have replied to a question concerning the allegedly non-existing "profit" from his research "Lord Gladstone, one day you will tax it" (see Kneißl and Schwarz 2011). We know, how true this prediction was, although this became obvious only many decades later than the government under Lord Gladstone existed. Many further examples could also serve to illustrate the different time scales of political governments, on the order of a decade in modern democracies, and the time delay of often many decades from a fundamental discovery to make it into textbooks and finally some practical, technical use (Quack 2014).

We shall conclude with an anecdote on the Schrödinger equation, which can serve as another illustration of what support to fundamental research implies. If one reads in Moore's Schrödinger biography on the history of the Schrödinger equation (Moore 1989) one finds that the first success in the search for this equation occurred during a stay of Schrödinger over the Christmas and New Year's holidays in Arosa 1925/1926, thus quasi as a private person in his free time without "financial support for this research". Obviously, a corresponding research proposal for "holidays, with the goal of discovering an equation" would hardly be funded by SNF today (it did not exist then), but the research was nevertheless indirectly publicly supported by giving Schrödinger the freedom of a professor with his salary at the University of Zürich to do research when and wherever he wanted to. Just how private this holiday was, can be guessed from some further gossip (Moore 1989), on which I shall not expand (Popper 1989). Moore states that Schrödinger stayed in "Villa Herwig" in Arosa, and provides a correct photograph of the house where holiday guests stayed (in contrast to the tuberculosis patients of Dr. Herwig in Arosa).

Fig. 2 Bergkirchli in Arosa (built 1492, Photograph by R. Quack)

However, if one looks in Arosa one finds that the house is actually called “Frisia”, it still exists today. There is ample evidence that Schrödinger made his breakthrough during this holiday, and the first paper of the famous series was written and submitted immediately after return from this holiday in January 1926. However, we do not know, of course, whether the illumination came in this house, or perhaps some Café in Arosa. My preferred (rather unfounded) historical hypothesis is a visit of Schrödinger’s to the Bergkirchli, a little church built by the Walser in 1492 and located in a most beautiful spot at 1,900 m altitude only a short walk away from the Villa Frisia (Fig. 2).

Such a heavenly inspiration of Schrödinger to find his equation describing the “music of atoms and molecules” can at best claim circumstantial evidence, although it is known that, for instance, Einstein’s attitude towards science as a means of understanding the world had a strongly religious component (Quack 2004), not so different from the inspiration drawn by writers of “real music” such as Bach (Quack 2004, 2012a). This is, however, only one of several attitudes to be found with scientists, motivating their research as a route towards discovering fundamental underlying truths of the world. Another, more modest attitude aims at just providing fruitful models of the world, but in the end these two approaches may be closer relatives than obvious at first sight (Quack 2014). Independent of the particular philosophy of research, it deserves support as a basic part of our culture to understand and shape the world in which we live.

Acknowledgment Our own scientific research receives support from the ETH Zurich, Schweizerischer Nationalfonds and by an Advanced Grant from the European Research Council. I am grateful to the hospitality of the Schmid family in their house in Arosa, where the present manuscript was written during a holiday and I am greatly indebted to Ruth Schüpbach for transferring my handwritten notes into a clean manuscript. The paper is dedicated to Richard Ernst on the occasion of his 80th anniversary.

References

Albert S, Meier BH, Quack M, Seyfang G, Trabesinger A (2006) On the possibility of stable nuclear spin symmetry isomers in H_2O. Chimia 60:476

Ausländer R (2002) Und nenne dich Glück. Gedichte. Fischer, Frankfurt

Buchstein H (2011) Der Zufall in der Forschungsförderungspolitik. Swifts "Akademie von Lagado". Forschung und Lehre 8:596–597 (cites Les Allen after Neil Duxbury, Random Justice, Oxford 1999, p 89)

Diederich F (2013) Are we refereeing ourselves to death? The peer-review system at its limit. Angew Chem Int Ed 52:13828–13829

Einstein A (1931) Zum kosmologischen Problem der allgemeinen Relativitätstheorie. Sitzber Preuss Akad Wiss Phys-Math Kl:235–237

Einstein A, Preuss SB (1931) Akad Wiss:235 (This is the very common false version of the citation)

Ernst RR (1992) Nuclear Magnetic Resonance Fourier Transform Spectroscopy (Nobel Lecture). Angew Chem Int Ed Engl 31:805–823; Angew Chem (1992) 104:817–836

Ernst RR (2010a) The follies of citation indices and academic ranking lists, a brief commentary to 'bibliometrics as weapons of mass citation'. Chimia 64:90–90

Ernst RR (2010b) The follies of citation indices and academic ranking lists, a brief commentary to 'bibliometrics as weapons of mass citation'. Bunsen-Magazin 5:199–200

Ernst RR, Bodenhausen G, Wokaun A (1987) Principles of nuclear magnetic resonance in one and two dimensions. Clarendon, Oxford

Gerhards J (2013) Der Deutsche Sonderweg in der Messung von Forschungsleistungen. In: Wissenschaftspolitik im Dialog, Schriftenreihe der Berlin-Brandenburgischen Akademie der Wissenschaften, Berlin

Heilbron JL (1989) The detection of the antiproton In: De Maria M, Grilli M, Sebastiani F (eds) Proceedings of the international conference on the restructuring of physical sciences in Europe and the United States 1945–1960 (Rome, 1988). World Scientific, Singapore, pp 161–209

Hirsch JE (2005) An index to quantify an individual's scientific research output. Proc Natl Acad Sci USA 102:16569–16572

Kneißl D, and Schwarz H (2011) Editorial: Grundlagenforschung braucht exzellente Wissenschaftler – und Freiräume. Angew Chem 123:12578–12579

Kuhn TS (1962) The structure of scientific revolutions. University of Chicago Press, Chicago

Manca Tanner C, Quack M, Schmidiger D (2013) Nuclear spin symmetry conservation and relaxation in water (${}^{1}H_2{}^{16}O$) studied by Cavity Ring-Down (CRD) spectroscopy of supersonic jets. J Phys Chem A 117:10105–10118

Merkt F, Quack M (2011) Molecular quantum mechanics and molecular spectra, molecular symmetry, and interaction of matter with radiation. In: Quack M, Merkt F (eds) Handbook of high-resolution spectroscopy. Wiley, Chichester; New York, pp 1–55 (see also preface to this Handbook)

Molinié A, Bodenhausen G (2010a) Bibliometrics as weapons of mass citation. Chimia 64:78–89

Molinié A, Bodenhausen G (2010b) Bibliometrics as weapons of mass citation. Bunsen-Magazin 5:188–198

Molinié A, Bodenhausen G (2011) The kinship or k-index as an antidote against the toxic effects of h-indices. Chimia 65(6):433–436

Moore WJ (1989) Schrödinger, life and thought. Cambridge University Press, Cambridge

Osterloh M (2013) Das Paradox der Leistungsmessung und die Nachhaltigkeit der Forschung. Nova Acta Leopoldina NF 117:103–113

Osterloh M, and Frey BS (2012) Input control and random choice improving the selection process for journal articles (Manuscript preprint); Ranking Games, to appear in Evaluation Review (2014)

Parkinson CN (1957) Parkinson's law and other studies in administration. The Riverside Press, Cambridge, MA

Perutz MF (1982) Ging's ohne Forschung besser? - Der Einfluss der Naturwissenschaften auf die Gesellschaft. Wissenschaftliche Verlagsgesellschaft mbH, Stuttgart

Petsko GA (2008) Having an impact (factor). Genome Biol 9, 107

Pfaltz A, Van Gunsteren WF, Quack M, Thiel W, and Wiersma DA (2009) An investigation with respect to the possible fabrication of research data reported in the thesis ETH No 13629 and in the papers Journal of Chemical Physics (2000) 112:2575 and (2000) 113:561, July 2009, ETH. Available: http://www.ethlife.ethz.ch/archive_articles/120123_Expertenbericht_tl/120123_Expertenbericht

Popper K (1989) Popper on Schrödinger. Nature 342:337–337

Quack M (2004) Time and time reversal symmetry in quantum chemical kinetics. In: Brändas EJ, Kryachko ES (eds) Fundamental world of quantum chemistry a tribute to the memory of Per-Olov Löwdin. Kluwer, Dordrecht, pp 423–474

Quack M (2011a) Fundamental symmetries and symmetry violations from high resolution spectroscopy. In: Quack M, Merkt F (eds) Handbook of high resolution spectroscopy. Wiley, Chichester; New York, pp 659–722

Quack M (2011b) Naturwissenschaften! Warum überhaupt? Warum nicht? In: Gegenworte, Heft 26, "Zweckfreie Forschung". Berlin-Brandenburgische Akademie der Wissenschaften, pp 28–35

Quack M (2011c) Naturwissenschaften! Warum überhaupt? Warum nicht? Bulletin der Vereinigung der Schweizerischen Hochschuldozierenden VSH/AEU 37:7–14

Quack M (2011d) Rede des Ersten Vorsitzenden anlässlich der 110. Hauptversammlung der Deutschen Bunsen-Gesellschaft für Physikalische Chemie in Berlin, 2. Juni 2011 (printed version of the speech). Bunsen-Magazin 13:138–143

Quack M (2012a) Rede des Ersten Vorsitzenden anlässlich der 111. Hauptversammlung der Deutschen Bunsen-Gesellschaft für Physikalische Chemie in Leipzig, 17. Mai 2012 (im Druck publizierte Fassung). Bunsen-Magazin 14:181–189

Quack M (2012b) Risiken und Chancen in der Forschungsförderung (Auszug aus der Rede des 1. Vorsitzenden der Deutschen Bunsen-Gesellschaft für Physikalische Chemie, Leipzig Mai 2012). Deutsche Zahnärztliche Zeitung 67:726–730

Quack M (2013) Error and discovery: why repeating can be new. Angew Chem Int Ed 52:9362–9370; Angew Chem (2013) 125:9530–9538

Quack M (2014) The concept of law and models in chemistry. Eur Rev 22:S50–S86

Quack M, Merkt F (eds) (2011) Handbook of high resolution spectroscopy. Wiley, Chichester; New York

Schrödinger E (1926a) Der stetige Übergang von der Mikro- zur Makromechanik. Naturwissenschaften 14:664–666

Schrödinger E (1926b) Quantisierung als Eigenwertproblem I. Ann d Phys 79:361–376

Schrödinger E (1926c) Quantisierung als Eigenwertproblem II. Ann d Phys 79:489–527

Schrödinger E (1926d) Quantisierung als Eigenwertproblem III. Ann d Phys 80:437–490

Schrödinger E (1926e) Quantisierung als Eigenwertproblem IV. Ann d Phys 81:109–139

Straumann N (2008) Materie, Antimaterie, und Dunkle Energie. In: Walde P, and Kraus F (eds) An den Grenzen des Wissens. Vdf Publishers, Zürich, pp 103–126 [Here it is shown that a paper of Einstein A (1931) (see citation) has been transformed by citations (obviously from people who never read or even saw the paper) until it reached its 'canonical' form with a second coauthor Preuss SB (Einstein A and Preuss SB 1931)].

Suhm MA (2010) Fahren mit angezogener Handbremse. Bunsen-Magazin 12:200

Szilárd L (1961) The Mark Gable Foundation. In: The voice of the dolphins, and other stories. Simon & Schuster, New York

Tikhonov VI, Volkov AA (2002) Separation of water into its ortho and para isomers. Science 296:2363–2363

Trommsdorff V (2002). Gesteinslaboratorium Alpen (Abschiedsvorlesung ETH Zürich). Bildarchiv der ETH-Bibliothek [Prod.]

Weinberg S (1967) A model of leptons. Phys Rev Lett 19:1264–1266

Zare RN (2012a) Editorial: assessing academic researchers. Angew Chem Int Ed 51:7338–7339

Zare RN (2012b) Research impact. Curr Sci 102:9

Why Are Some Nations More Successful Than Others in Research Impact? A Comparison Between Denmark and Sweden

Gunnar Öquist and Mats Benner

Abstract Bibliometric impact analyses show that Swedish research has less international visibility than Danish research. When taking a global view on all subject fields and selecting publications cited higher than the 90th percentile, i.e., the Top 10 %—publications, the Swedish Research Council shows that although Sweden ranks 15 % above world average, Denmark, the Netherlands and Switzerland rank 35–40 % above. To explain these different performances, The Royal Swedish Academy of Sciences asked us to compare the national research systems on three levels: priority setting at national level, governance of universities and direction and funding of research. There are of course many similarities between the Danish and Swedish research systems but there are still subtle differences that have developed over time, which may explain the different international visibility. First of all, it does not depend on different levels of public spending on research and development. However, the core funding of universities relative external funding is higher in Denmark than in Sweden. The academic leadership of Danish universities in terms of board, vice-chancellor, faculty dean and department chair is also more coherent and focused on priority setting, recruitment, organization and deployment of resources to establish research environments that operate at the forefront of international research. On all these points we see a weaker leadership in Sweden. Furthermore, over the last 20 years, public funding of research in Sweden has become more and more unpredictable and program oriented with many new actors, while the Danish funding system, although it also has developed over time, shows more consistency with strong actors to fund individuals with novel ideas. The research policy in Sweden has also developed multiple, sometimes even conflicting goals, which have undermined conditions for high-impact research, while in Denmark a policy to support excellence in research has been more coherent.

G. Öquist (✉)
Department of Plant Physiology, Umeå Plant Science Centre, Umeå University, Umeå, Sweden
e-mail: Gunnar.Oquist@umu.se

M. Benner
School of Economics and Management, Lund University, Lund, Sweden
e-mail: Mats.Benner@fpi.lu.se

I.M. Welpe et al. (eds.), *Incentives and Performance*,
DOI 10.1007/978-3-319-09785-5_15

1 Introduction: Nations, Research Governance and Scientific Impact

This paper analyses the long-term (1990–2010) development of scientific publication impact in Denmark and Sweden. It sets out to explore the scientific impact of two countries with similar conditions in the form of relatively small and predominantly publicly funded research systems, but with very different scientific impact patterns. In terms of high impact publications, Denmark by far surpasses Sweden. We assume that differences in patterns of scientific impact—across nations and over time, and as measured in bibliometric accounts of scientific impact—can be explained by institutional mechanisms (cf. Hedström and Swedberg 1998). Among these, two elements are assumed to be the most important: first, the structure of resource allocation to the universities and in particular the share of funding that is at the direct disposal of the university; and second, the form and authority of university governance. Hence, we assume that there is a connection between how national priorities are set and how organizations operate within these systems. We assume that differences at the national level can be explained by the interplay between national policy and institutional conditions at the level of universities.

The connections between institutional set-ups and impact profiles have been a recurrent theme in studies of industrial relations and welfare systems (cf. Esping-Andersen 1990; Hyman 1995, and more recent literature). It has been less prevalent in studies of scientific impact, which studies either have set out to distinguish impact differences (often on the basis of bibliometric methods) or to distinguish policy and governance models and policy templates (in policy analysis). Only seldom have the connection between the two been explored.

2 Bibliometric Comparisons

Our starting-point is bibliometric patterns, as revealed in the Web of Science. The Swedish Research Council has since 2006 repeatedly published bibliometric analyses comparing the international impact, or visibility of Swedish research, with that of some other European countries of roughly similar population size and with national ambitions to build knowledge based economies for the future (Karlsson and Persson 2012). For the analyses, The Council mainly used the publication database of Thomson Reuters defining 251 journal subject fields. This database is somewhat biased towards the natural and life sciences with a weaker coverage of the social sciences, humanities, engineering and mathematics. However, for a comparative analysis this bias should not distort the main findings, since the publication strategies in the countries of comparison should be roughly similar.

When taking a global view on the publications of all subject fields (Karlsson and Persson 2012, 2014), Finland, Norway and Sweden have today lower mean citation rates and lower proportions of highly cited papers (defined higher than the 90th

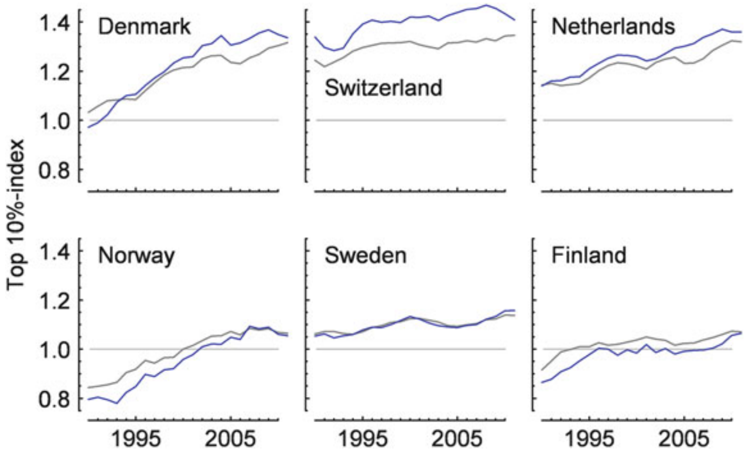


Fig. 1 Development of the top 10 %-index between 1990 and 2011 for Denmark, Switzerland, the Netherlands, Norway, Sweden and Finland. For comparison the national mean citation rate is shown as a grey *curve* and the *grey horizontal line* (with a value of one) shows the world average. A value of 1.2 means 20 % above world average, etc. The *curves* are based on 3-year moving averages [Reproduced from Karlsson and Persson (2012, 2014)]

percentile, i.e., The Top-10 publication index) than found in Denmark, the Netherlands and Switzerland (see Fig. 1). In 2011, the Top-10 publication index for Sweden scores 15 % above world average, while Denmark, the Netherlands and Switzerland score in the range of 35–40 % above world average, approaching the performance of the US. Although both Finland and Norway show a positive development after 1990, the impact has levelled off just above world average.

These results are corroborated (Karlsson and Persson 2012, 2014) when looking at the frequency distribution of national research organizations (mainly universities) as a function of the Top-10 publication index and with an annual production of 50 publications or more. The weakly performing countries Finland, Norway and Sweden centre around world average, while in Denmark, the Netherlands and Switzerland most organizations perform well above world average. The same pattern in publication impact is clearly visible when the highly prestigious journals Nature, Science and the Proceedings of the National Academy of Sciences (PNAS) are compared. The bibliometric analyses furthermore show that the recruitment rate of new generations of scientists establishing themselves among the category of highly cited scientists is weaker in the three low performing countries than in the three high performing countries. These findings are corroborated by bibliometric studies conducted by NordForsk (2011, 2014), an agency under the Nordic Council of Ministers.

Differences in international visibility of publications could be affected by many factors such as fraction of publications never cited, different degree of international cooperation, different degree of interdisciplinarity, different proportions between subject fields or different degrees of self-citations. Although there are some

Table 1 Number of subject fields where the country publishes at least 10 papers per year

Country	No. of fields selected	No. of subject fields where			Contribution to total national output of fields where	
		Top 10 %-index < 0.8	Top 10 %-index >1.5	Top 10 %-index > 2	Top 10 %-index >1.5	Top 10 %-index >2
Denmark	156	16	50	13	27 %	6.7 %
Finland	155	46	16	6	7 %	2.7 %
Netherlands	213	5	62	10	24 %	2.7 %
Norway	157	36	20	3	11 %	1.1 %
Sweden	190	34	30	5	11 %	1.5 %
Switzerland	181	17	74	13	48 %	4.4 %

A Top 10 % value of 1 means world average, <0.8 means more than 20 % below world average, and 1.5 means 50 % above world average, etc. Reproduced from Karlsson and Persson (2012, 2014)

differences noticed in these parameters, they cannot explain the different global impacts of research results produced in the countries of this comparison (Karlsson and Persson 2012, 2014).

In this communication, we will focus on a comparison between Denmark and Sweden. The conclusion from the bibliometric comparison is that the international visibility of Swedish research on average falls well behind that of Danish research. Detailed field analyses, however, clearly show that both Denmark and Sweden have research fields that show top performances in international comparisons (See Table 1), but it is equally clear that these fields in number are fewer in Sweden than in Denmark. Apparently, Swedish universities are more prone than Danish universities to nurture subject fields with limited international visibility. The question is how this difference in visibility correlates with quality of research in terms of advancing the frontier of knowledge. Can we infer that research communication that is highly visible internationally is also more likely to be of breakthrough character? This can be discussed since pioneering results that challenge accepted views and eventually prove to be right may have difficulties to be accepted and therefore initially ignored. However, integrated over time pioneering research breaking new ground for understanding is likely to become highly visible. This does of course not mean that all research results that are highly visible also are of breakthrough character. Our point of view is that when research scores high on international visibility it is actively part of setting the international knowledge platform, and over time the probability of finding real breakthroughs is higher among high impact publications than in low impact publications. Our conclusion is therefore that the bibliometric comparison reveals that on average the quality of Danish research exceeds that of Swedish research although in specific subject fields Swedish research may still compete with Danish research.

We now move on to search for explanations why Swedish research during the last 20 years shows a weaker development than Danish research when it comes to international visibility. These analyses are based on a study that we made in 2012 on behalf of the Royal Swedish Academy of Sciences (Öquist and Benner 2012), and upon request of the Norwegian Research Council later updated to include also Norway (Benner and Öquist 2014). For more detailed analyses we refer to these two studies.

3 Priority Setting at National Levels

3.1 Denmark

During the 1990s, a series of policy reforms were instigating with major investments in basic research, and an overhaul of the institutional structure of the Danish research system (Olesen Larsen 2010). During the last 20 years, the policy has been to strengthen the universities, to merge universities and to integrate sectoral institutes into the university system, to reform postgraduate education to match international standards, to strengthen academic leadership at all levels, to put an emphasis on international recruitments and more transparent career paths, and to reform the funding system to serve clear missions in terms of competitive support for basic research, strategic research and research linked to innovation. The infusion of funding has gradually increased, and with the Globalization strategy in 2006 a total of five billion Euros for research, innovation and education reinforced core funding of universities as well as external funding for different purposes (Aagaard and Mejlgaard 2012). Currently, the proportion between university core funding and external funding is around 3/2, slightly lower than a decade ago when it stood at 70 % (Ibid.). These reforms were made in order to reinvigorate the rather stagnant academic system of the 1980s (Andersen 2011), but also seen as a means to create new ground for economic growth. Through this series of reforms, the national research policy has been to emphasize the need for a stronger international orientation aided by a clear policy focus on academic excellence. As a consequence, the international impact of Danish research has clearly grown to an international top position.

All these reforms can not be seen as the result of a unified master plan but is rather the result of a stepwise development involving various stakeholders transcending the boundaries between government, universities, funding organisations and industry, all governed by a culture to strive of academic excellence (Olesen Larsen 2010). However, the system has not been transformed without conflicts. The University Act of 2003 altering the governance of universities to a more centralized mode at the expense of the decisive role of collegial faculties was met with great criticism among professors.

In addition, the merge of sectoral research institutes and a number of professional educations into the university system during the 2000s seem in part to have worked less smoothly. There is a clear risk that the research mission of universities, similar to the situation in Sweden, will be blurred by conflicting goals when expectations of relevance may weaken the collegial focus on academic excellence. Furthermore, the recently decided merge between the Strategic Research Council and The Danish Advanced Technology Foundation based on the ambition of creating a stronger link between research and innovation reinforces a more top-down governance model of research and may change the balance in favour of relevance at the expense of scientific excellence.

3.2 Sweden

Like in Denmark, the Swedish research policy has gradually been transformed from being a primarily academic issue to be a mean to transform the society into a knowledge-based economy. A number of reforms have been instigated to achieve this. It actually began with the post-war period when the strategy was not to develop an extensive institute sector as in many other countries. Instead, a string of sectorial research agencies were established to support research of societal or industrial needs within the universities, especially so in engineering, medicine and the social sciences. This resulted in a growing number of research groups within the universities supported by 'soft money' with a focus on relevance in research to solve problems rather than on breaking new ground of knowledge in a more general sense. With this development, both basic and applied research environments developed alongside each other with limited articulation. In parallel, all professional education and training were in 1977 integrated with the university system and traditional academic education was restructured to match the need of the labour market. With these reforms, universities became the nodes of an expanding number of regional Higher Education Institutions, some of which later have gained university status. This proliferation of the academic system distinguishes Sweden from Denmark, which has refrained from this kind of numeral expansion (Sweden has 39 universities and university colleges, Denmark only 8). To the broad mandate of Swedish universities should be added that since 1997, collaboration with the society is placed on par with teaching and research as one of the universities' three missions. This mission of collaboration was further emphasized in the latest research bill to the parliament. Parallel with these developments, the autonomy of universities has gradually increased with external board members being in majority, with the chairman of the board being external and with the board proposing the appointment of vice-chancellor.

To this should be added that the public policy to fund research has been to steadily increase external, competitive funding at the expense of university core funding. Today the majority of public research funding of universities comes from external sources—for some universities like Karolinska Institutet and Chalmers University of Technology the figures are close to 70 % external funding. Swedish universities have as a result developed a very broad spectrum of goals within a single organizational set-up where a multitude of project grant recipients—rather than the university themselves—set the strategic direction for their activities. Danish universities have also developed multiple tasks with an extension of the organisational complexity but the leadership is in better control of the development and the whole system remains more committed to employ stringent scientific standards, an issue that we return to below.

4 Governance of Universities

4.1 Denmark

Before 1990, the governance of Danish universities was loose with recourses and power widely dispersed in the system. Leadership played a minor role in matters of recruitment and resource allocation which were instead managed by various committees with broad representation (Andersen 2011). A modest reorientation began already in the 1980s when some universities and other research units began raising their expectation of publication and international orientation. With the 1993 reform, Danish universities took the first steps away from decentralized decision-making by empowering in particularly department heads but also vice-chancellors. Universities gained more organizational autonomy but signed contracts with the state regarding performance (Olesen Larsen 2010). Another important reform in 1993 was a reformed postgraduate education with 3-year streamlined research programmes, some of which were organized in graduate schools.

The University Act of 2003 instigated major changes in university governance. Responsibilities were now centralized to university boards composed by a majority of external members. The board appointed the vice-chancellor, who appointed deans who in turn appointed department heads. This reform streamlined decision-making, it became more centralized and it weakened the role of the collegial board of faculty. Today, deans have a strong role in recruitment, organization of departments and allocation of internal resources. The active role of the dean for international recruitment is often emphasized as a driver for increased quality and visibility of Danish research. Clearly, the Danish leadership of universities have gone from being largely reactive a couple of decades ago to now being proactive and taking responsibility for fostering excellence with the often expressed ambition to make their university be ranked among the internationally leading one.

In Danish universities, the professorial positions remain relatively few in comparisons with the situation in Swedish universities. This is because the leadership has opposed the use of a promotion system and kept control of recruitment as a mean to establish strong research environments. The Danish system has with the infusion of new resources also seen an increase in the number of postdoctoral positions on temporary contracts but one has refrained from appointing faculty positions on external funding. The leadership of Danish universities also more and more emphasize international competitive recruitment. With this strict university control of recruitment it is somewhat surprising that a clear, and internationally competitive tenure track system for recruiting professors has not emerged.

With the 2006 reform to reduce the number of universities and to align the sectoral research institutions and professional educations with the university, the size of the remaining universities has not only grown but the complexity of governance of these growing conglomerates has also increased. This is a challenge, and while the merge of universities generally is considered reasonable, the aligning the more sectoral institutes seems to have been less successful to date. Today, four major universities remain—Copenhagen, Aarhus, Southern Denmark University (SDU) and the Danish Technical University (DTU). These universities now account for two-third of Danish research and higher education and one can also notice a trend towards specialization among the universities, all governed by strategic decisions and implementation.

The centralisation of university governance was met by strong resistance in many quarters, since it weakened the professional role of collegial faculties in decision making reducing their role to an advisory capacity. At the same time, it seems reasonable to argue that reforms—in conjunction with changes in the funding of Danish research—has had considerable impact on the Danish research system. Nonetheless, fears of managerial overload is still noticeable as there is always a risk that central decision-making will impede creativity and innovation in research. Tendencies in these directions should be offset first by a continuous reliance on the professional academic staff to plan and execute their research within broad profiles decided by the board and second, by boards that appoint vice-chancellors with strong academic, professional merits and an understanding of how to foster academic excellence through appointments and prioritization of resources. This view is corroborated by the findings by Amanda Goodall (2006, 2009; Goodall et al. 2014) that presidents or vice-chancellors of universities, but also department chairs, who have strong scholarly merits and high academic legitimacy are more successful in leading their universities towards better performance than are leaders recruited based on primarily managerial merits. Hence, the strong development of Danish research will be conditioned on the balance between steering, integrity and professional judgement.

4.2 Sweden

Until the late 1970s, Swedish universities were governed by academic representatives. Faculty leaders comprised university boards selected in a collegial process and departments were headed by chaired professors. However, after the 1970s universities gradually expanded into conglomerates of activities with broadened missions in research, development, education and cooperation with society. From now on many stakeholders created gradually new power structures in which non-tenured teaching staff, university bureaucracy, university teachers' associations and student representatives became increasingly important. In 1992, the university boards were further transformed and external members gained majority. In 1997, vice-chancellors were replaced as board chairman by an external representative and from now on the university boards took over the responsibility to propose the appointment of vice-chancellors. During this development, the autonomy of universities grew gradually, which is a development that has continued. Today boards and the vice-chancellors have the governing power to organize the operational structures and functions in a way that they find most appropriate to reach the goals set by the government.

Swedish universities have never been as dependent on external stakeholders as they are today. This development has come gradually and after the latest research bill the core funding has dropped to a level just under 50 % of funding of research (Swedish Government 2012/2013). This development has definitely increased competition for external funding and increased productivity but the quality control is very much in the hands of the granting agencies of which some emphasize relevance at the expense of scientific quality. This development has definitely weakened the quality control at the university level, and with the heavy dependence on external funding for research the core funding of universities is to a large extent used to match external grants. This has made academic leadership reactive in setting and implementing priorities. Altogether this means that university departments in Sweden today tend to be a loose collection of externally funded projects rather than an environment with a concerted collegial action to foster research along a defined long-term plan (Schwaag Serger et al. 2014).

The career structure and recruitment mode for faculties is in Sweden not under control of a visionary leadership that systematically use these appointments to create research environments with the potential to excel in research. During the financial crises in the 1990s, the universities faced some quite severe budget cuts that affected the career opportunities negatively. For a couple of decades, young people pursuing a career in academia mostly have been forced to rely on a range of externally funded opportunities provided by research councils, foundations etc. The universities have responded to these changes by increasingly filling faculty positions at different levels on external funding. This reactive mode was further manifested when in the late 1990s a reform was launched, which more or less abandoned the professorial department chairs with responsibility for a discipline or a subject field, and gave the right to lecturers to be promoted to professor irrespective of mode of funding. To be promoted to professor meant that you received the professorial title but it was not linked to a defined position. There was some resistance among universities to abandon the chair system but after a few years the old structure was more or less eroded. Between 1998 and 2002 the number of professors grew as much as 45 % without being backed by an increase in core funding to universities—the funding of the increase was instead borne by a combination of hikes in external funding and in student intake (Benner 2008).

The erosion of the career system in the 1990s, the relaxed attitude to documented merits for promotion to professor and the acceptance of funding faculty positions on "soft money" provided by different external stakeholders without long term commitments are probably key factors behind the relatively weak recruitment of newcomers performing at the high impact level (Karlsson and Persson 2012, 2014). Today, universities try to restore the career system but an internationally competitive tenure track system is still lacking and universities still have a long way to go before they can offer international competitive conditions for recruitment of faculties at all levels. In reality, recruitment is only partially in the hands of the leadership. Instead it is predominantly grant holders and their research environments that provide career opportunities and de facto recruit future faculties.

The fact that the career and recruitment processes are not, or only marginally, under the control of a leadership troika of vice-chancellor—dean—department chair, as we find in leading universities around the world, has undermined the role of the academic leadership in Sweden. Another reason is that university boards have had a tendency to select vice-chancellors after their presumed ability to manage the complex conglomerate of activities of Swedish universities rather than academically more experienced scholars (Engwall 2014; Engwall and Lindwall Eriksson 2012). This has weakened the legitimacy of the leadership in the eyes of scientifically, leading faculties. This is a major concern, since as previously mentioned, Goodall has in a series of studies (2006, 2009; Goodall et al. 2014) shown a positive correlation between academic standing of vice-chancellors or department chairs and international recognition of academic institutions. To enhance the international standing of Swedish universities there seems to be a pressing need to develop an internationally competitive career and recruitment system that are in the hands of a strong and visionary academic leadership at all levels.

5 Direction and Funding of Research

5.1 Denmark

In the 1980s, new resources were infused in the Danish research system promoting industrial clusters and technological development. The research council system to support basic research was traditionally run in a rather conservative form with small grants and insular selection criteria. However, a major adjustment of focus occurred in 1992 when the Danish National Research Foundation (DNRF) was established with the mission of funding basic research without any other criterion than scientific excellence (Olesen Larsen 2010). DNRF introduced rigorous international peer review in the selection of proposals to grant. The focus was on individuals with challenging ideas and successful applicants were generously funded for up to 10 years with a far reaching autonomy with respect of research agenda and the use of the funds. Through these Centres of Excellence, complemented with a top international recruitment programme at the professorial level, DNRF has not only fostered new generations of scientific leaders but it has also provided outstanding conditions for PhD students and postdoctoral training in nurturing environments.

DNRF has played an important role in the internationalization of Danish research and it has definitely been a strong driver for scientific excellence. A resent review confirms this as it notices that the funding strategy has "enabled researchers to venture into novel and often risky projects which may eventually lead to ground-breaking results" (Evaluation of The Danish National Research Foundation 2013). Bibliometric analyses show that the 66 centres funded so far have a publication impact comparable with the best research institutions in Europe. DNRF is a successes story in Danish research policy and it has substantially contributed to rise the profile of Danish research to an international top position.

In the 2000s, the research council system was reformed merging the previously six independent councils into what is now the Danish Council for Independent Research. The Danish Council for Strategic research was also established to channel resources into centres and networks in areas primarily identified by the political system. The Danish National Technology Foundation was also set up to offer private companies and universities support for the innovation of new technologies. With these reforms, the Danish funding system consisted of independent, public agencies with clear missions for basic research, strategic research and research/innovation. The system was in 2006, as mentioned, boosted through the Globalization strategy with a total of five billion Euros for research, education, innovation and other forms of professional development. About half of the investment has been for research and the university core funding has increased with approximately DKK 1billion annually. Through the UNIK programme, the universities have been granted for major undertakings in strategic areas. The ratio university core funding/external funding is today on average 3/2.

During these more than 20 years of growing support for research, there has been a relatively good national support to promote research careers. This has been through the research council supporting Young Elite Researchers, DNRF supporting mid-career research leaders and recently the research council established a national career program Sapere Aude ("Dare to Know" in Latin) in three steps from postdoctoral to professorial level. However, Danish universities have, as mentioned, not yet established an internationally competitive tenure track system for the recruitment of professors.

5.2 Sweden

During the 1980s the research council and sectoral agencies became more and more important for the funding of research, although the Government also infused fresh resources into the universities. With the economical crises in the early 1990s, and the wishes to lay a new foundation for the development of a future knowledge based economy, a new layer of research funding instruments was established based on money from the dismantled Earner Wagers Funds. These new foundations were set up for various strategic purposes, and they infused considerable financial resources into research and innovation. The largest foundation, the Swedish Foundation for

Strategic Research funded centres and networks within areas considered to be strategic for Swedish industry. The support of postgraduate studies was an important element of the funding. With this development, funding of Swedish research was geared so support expanding PhD training and research networks within areas motivated by utilitarian needs rather than propelling research on novel ideas along the concept of the Danish National Research Foundation established at the same time. Looking back, this new layer of funding most likely curtailed risk-taking in Swedish research since it focused more on problem solving in different sectors rather than on challenging accepted views and perspectives in the pursuit of new ideas and understanding.

As the economic problems continued, more radical measures were taken in the mid-1990s in order to curb the budget deficit. The budgets for research councils, mission-oriented agencies and universities were cut by around 20 % affecting primarily the career opportunities for young researchers. Some of these cuts were offset by the new foundations, which had been set up a few years earlier, but the core funding of universities and the individual project support from research councils were never recouped. This severely weakened the ability of universities and research councils in their support of research. Making matters even worse, university core funding has since then only marginally been adjusted for inflation (Sundqvist 2010).

With the gradual recovery of the nation's economy, the resource cuts in the 1990s have been superseded by resource hikes and a reinforced policy on conditions for basic and applied research. A unified research council was established in 2001 to offset the fragmentation of funding for basic research (Benner 2001). Research bills have since then infused considerable new resources into research to support strong environments, networks and defined strategic areas. Recently, research council resources have also been directed to improve career opportunities for young scholars, to emphasize excellence in granting the most promising projects more long term, and to stimulate international appointments. However, these kind of reforms have come later than in Denmark and the positive effects are still to be seen. The funding policy has been such that emphasis has been on external funding with the effect that on average the ratio university core funding/external funding has now fallen just below one. Swedish research policy after 2000 has empowered established scientists, particularly those who have worked in areas considered to be of strategic importance for the country. To make a career, young researchers often have had to follow in the steps of the leading generation of scientists. This development has most likely undermined the kind of renewal of research and innovation that builds on young, incoming generations of scientists been given the trust, freedom and resources to develop and pursue their own ideas.

6 Discussion: Why Is Danish Research More Visible on the International Scene?

The starting point of our analysis is the interplay between national policy and institutional governance and how that interplay determines scientific impact as measured and defined in bibliometric terms. We have asserted that Denmark and Sweden have similar preconditions but that they show different outcomes. Similar levels of public investments have been allocated in different ways, with different effects on university governance, in terms of academic authority and recruitment patterns. This, we argue, is what explains the different patterns of Danish and Swedish research, and in particular the strongly positive development of Danish research.

If we first look at priority settings at national level, it is clear that the Swedish research policy so far has been less consistent than what we find for Danish research policy. In both countries, research policy operates with multiple goals but Sweden to a much larger extent. Swedish universities have evolved into organizational conglomerates operating with a broad range of missions in education, research, development and collaboration with various, sometimes conflicting, goals. All these missions are of course important in a knowledge based society but merging them all in one organizational set up has, we argue, eroded the ability to stay at the international forefront of knowledge renewal in terms of new discoveries and understanding. This is reflected in the bibliometric impact analyses showing that Sweden has a surprisingly large number of underperforming subject fields and only a limited renewal of its research cadres performing at the highest level of recognition and visibility. Our evidences also suggest that even individual scientists within subject fields that score relatively high in international comparisons feel increasingly alienated in their role as scientists: 'we suffer under the burden of an excel tyranny despite suboptimal working conditions' reported one professor in engineering. These developments have not gone as far in Denmark as in Sweden. In contrast, Denmark has during the last 10 years merged and concentrated rather than proliferated their research efforts, and have—on the basis of DNRF's programmes and parallel support from government agencies—identified and empowered a number of high-performing scholars, who also receive ample support from their respective universities. A key to the future visibility of Danish research will be to retain this model of a shared responsibility and a careful process of recruiting, selecting and empowering talent. However, with the merge of sectorial research institutes and professional educations into universities, Denmark embarks on the same path towards multiple missions and goals, and there is a growing risk that also Danish universities may lose their focus on excellence in research as defined by international benchmarking, and instead venture into a model of multiple goals and stakeholder balance instead of a sharp focus on identifying and nurturing talent. In both countries we now have mass universities. The challenge for the future is to conduct academic education and research with long-term perspectives and of the highest distinction of excellence, and at the same time satisfy the need of professional educations, research, development and collaboration to satisfy national needs in the short- to medium-term. This is a leadership issue and the Nordic countries, which all share the same challenge, may work together to find new organizational solutions for universities to better satisfy different missions and goals.

In the comparison of governance models, Denmark stands out by developing a model of academic leadership which is being proactive by not only deciding on strategies but also implementing them through prioritizing of recourses, organizational changes, recruitments, etc., and in particular in responding positively to signals of high quality recruitment and clear leadership. One important factor behind this ability is of course that in relative terms the Danish universities have better core funding than Swedish universities, which are more dependent on external funding. However, the leadership hierarchy, vice-chancellor—deans—department chairs, seems in Denmark to be particularly important for recruitment, often international, and for the management of research environments where the predominant goal is research excellence measured by international comparisons. Such environments become very successful in the competition for external grants, of which some granting agencies and foundations fund long-term with substantial amounts of money. In this way, there is a rather informal cooperation between universities and granting organisations in fostering excellence in research.

The Danish leadership rests on the conditions that the universities are in full control of not only recruiting faculties but also providing salaries and start up packages. This role of the Swedish university leaderships has been more or less eroded by the growing dependence of external granting for providing salaries for appointed faculties and by the so far excessive use of promoting lecturers to professor without linkage to financial recourses. A remarkable weakness of the Swedish research organization is also that in most cases the dean does not appoint department chairs based on scientific merits to orchestrate the scientific development of the environment. A department is therefore the sum of externally funded projects granted on a fluctuating market, and the appointed head, the prefect, has primarily an administrative role. In such a system, it is in reality the successful grant recipient that recruits and provide career opportunities while the leadership formally appoint faculties. This extreme bottom-up mechanism reduces mobility and over time scientific inbreeding is a threatening development. Clearly, the present recruitment policy of Swedish universities has severely undermined mobility threatening over time the creativity and renewal of research (Törnqvist 2009).

When it comes to the direction of funding, Danish policy has been more consistent in terms of striking a balance between direct core funding of universities and indirectly through external granting agencies—currently the figure is 3/2, whereas Sweden has an opposite relation with less than half of research income in the form of block grants to universities. However, in both countries, there has been a shift from institutional core funding to competitive, external project or program funding. This policy has gone very far in the Swedish case, weakening the universities ability to actively set their research agendas by providing the necessary core funding. With this development, university leadership shows signs of opportunism in relation to the financial opportunities provided rather than developing the institutions own agenda, which we see more of in Denmark with a better core funding. Another effect is that successful scientists attracting substantial external funding more or less use universities as "research hotels" without engaging themselves in governing issues.

Another important element is the format of external support. The Danish model has been much more clear-cut and focussed, whereas Sweden's is fuzzier. Both in Denmark and Sweden the recourses for research have grown substantially and in both countries the 1 % goal of GDP has roughly been achieved. We however notice one important deviation in the early 1990s, which has had considerable effect. While Denmark established the National Danish Research Foundation to support major undertakings in basic research with the focus on individuals with leadership ability and creative new ideas as the single criteria for funding across disciplines, Sweden's research politicians decided to channel substantial new resources through a new set of foundations supporting various presumed strategic needs for different sectors of the society. In Denmark, a recent international review reveals that DNFR through its more than 20 years of operation, with only a small share of public funding, has substantially increased the international standing and visibility of research. There is also evidence that the international focus on scientific quality established by DNRF through the use of an internationally rigorous peer review process has been contagious and emphasized the international quality focus among other funding organizations (Aksnes et al. 2013). We see no parallel development in Sweden. In Sweden, new recourses were directed mainly through external channels towards strong environments, networks and defined strategic areas undermining the proportion of university core funding. In contrast, the Danish infusion of new revenues was directed both directly to universities and indirectly through competitive schemes using available channels for basic research, strategic research and research linked to innovation. As a consequence, the Danish infusion of resources has been more consistent over time, while in the Swedish case the infusion has been more opportunistic and top-down oriented. Furthermore, the targeting of new recourses in Sweden have primarily benefitted established scientists, while in Denmark a substantial share has also gone to support individual research careers.

To sum up, Denmark rests on a model of relative coordination between funding streams, where the DNRF has had an immense 'cultural' impact on the notions and ambitions, and where this has been followed up by universities and by other government agencies. In Sweden, the picture is more complex and sometimes even contradictory, and universities have been weakened as collective agents and instead function as containers of multiple activity flows. In lieu of a change agent like DNRF, and without accompanying broad-based university reforms, there has been a fuzzy set of incentives and support measures that do not form a uniform and coherent pattern, and it has not triggered a changing conception of governance within universities but rather propelled a passive, financially motivated model of recruitment and promotion. While Swedish research holds a higher than international standard overall and a sizable presence in high-impact research, Denmark does far better. This, we argue, reflects a Swedish research governance model that is relatively weak in fostering a mind-set, modes of operation and institutional culture where bold ambitions prevail.

For a national research policy with the ambition to foster excellence in research at the highest level it is the creative, often young, individual researcher in a nurturing environment with complementary skills, individuals with novel ideas that should be targeted. It is particularly important that the research career is made attractive to young, talented, visionary people. This was well articulated by the 2012 EU Conference in Excellence in Research (The Aarhus Declaration 2012); emphasizing the following drivers for excellence in research: recognizing and nurturing talents, trust and freedom, long-term perspectives, creative and dynamic research environments, and beyond and across disciplines. These words are for universities (research institutions) and granting agencies to follow if one wants to foster quality of the highest distinction in research and innovation meeting the requirements of excellence at the global level. In a successful knowledge based society and in a more globalized world, national research policy, but also education policy, needs to match international criteria in terms of visionary strategies and implementation.

References

Aagaard K, Mejlgaard N (2012) Dansk forskningspolitikk efter årtusindskiftet. Aarhus universitetsforlag, Aarhus

Aksnes D, Benner M, Brorstad Borlaug S, Foss Hansen H, Kallerud E, Kristiansen E, Langfeldt L, Pelkonen A, Sivertsen G (2013) Centres of excellence in the Nordic countries. A comparative study of research excellence policy and excellence centre schemes in Denmark, Finland, Norway and Sweden. NIFU, Oslo

Andersen JE (2011) Academic staff in Denmark: The consequences of massification in a small country. In: Enders J (ed) Academic staff in Europe. Changing contexts and conditions. Greenwood Press, London, pp 49–66

Benner M (2001) Kontrovers och konsensus. Nya Doxa, Nora

Benner M (2008) Kunskapsnation i kris? Nya Doxa, Nora

Benner M, Öquist G (2014) Room for increased ambitions? Governing breakthrough research in Norway 1990-2013. The Research Council of Norway. http://www.forskningsradet.no/publikasjoner. Accessed 5 Mar 2014

Engwall L (2014) The recruitment of university top leaders: politics, communities and markets in interaction. Scand J Manag 30:332–343

Engwall L, Lindwall Eriksson K (2012) Leaders of modern universities: Primi inter pares of chief executive officers? In: Tengblad S (ed) The work of managers. Oxford University Press, Oxford, pp 206–225

Esping-Andersen G (1990) The three worlds of welfare capitalism. Polity, Cambridge

Evaluation of the Danish National Research Foundation (2013) Danish agency for science, technology and innovation. http://www.fivu.dk. Accessed 5 Mar 2014

Goodall A (2006) Should top universities be led by top researchers, and are they? A citations analysis. J Doc 62:388–411

Goodall A (2009) Highly cited leaders and the performance of research universities. Res Policy 28 (7):1078–1092

Goodall AH, McDowell JM, Singell LD (2014) Leadership and the research productivity of university departments. IZA Working Paper No. 7903
Hedström P, Swedberg R (eds) (1998) Social mechanisms. An analytical approach to social theory. Cambridge University Press, Cambridge
Hyman R (1995) Industrial relations in Europe: theory and practice'. Eur J Indus Relat 1(1):17–46
Karlsson S, Persson O (2012) The Swedish production of highly cited papers. Vetenskapsrådets lilla rapportserie 5:2012. The Swedish Research Council
Karlsson S, Persson O (2014) The production of highly cited papers. Vetenskpsrådets lilla rapportserie 5:2012 updated to include Norway. In: Benner M, Öquist G (2014) Room for increased ambitions? Governing breakthrough research in Norway 1990-2013. The Research Council of Norway. http://www.forskningsradet.no/publikasjoner. Accessed 5 Mar 2014
NordForsk (2011) Bibliometric research performance indicators for the Nordic countries. Nordforsk, Oslo. http://nordforsk.org. Accessed 5 Mar 2014
NordForsk (2014) Comparing research at Nordic universities using bibliometric indicators. Nordforsk, Oslo. http://nordforsk.org. Accessed 5 Mar 2014
Olesen Larsen P (2010) Stages on the research road: Danish research policy, past and present. Syddansk universitetsforlag, Odense
Öquist G, Benner M (2012) Fostering breakthrough research: a comparative study. Akademirapport, Royal Swedish Academy of Sciences
Schwaag Serger S, Bienestock A, Benner M, Lidgard A (2014) Combining excellence in education, research and impact: Inspiration from Stanford and Berkeley and implications for Swedish universities. SNS, Stockholm
Sundqvist B (2010) Svenska universitet: lärdomsborgar eller politiska instrument? Gidlund, Hedemora
Swedish Government (2012/2013) Forskning och innovation. http://www.regeringen.se/sb/d/15650/a/201368. Accessed 5 Mar 2014
The Aarhus Declaration (2012) http://www.excellence2012.dk/the-aarhus-declaration/. Accessed 5 Mar 2014
Törnqvist G (2009) Kreativitet i tid och rum: processer, personer och platser. SNS, Stockholm

Part V
Innovative Approaches to Research Evaluation

Social Media and Altmetrics: An Overview of Current Alternative Approaches to Measuring Scholarly Impact

Katrin Weller

Abstract This chapter describes the current state of the art in altmetrics research and practice. Altmetrics—evaluation methods of scholarly activities that serve as alternatives to citation-based metrics—are a relatively new but quickly growing area of research. For example, researchers are expecting that altmetrics that are based on social media data will reflect a broader public's perception of science and will provide timely reactions to new scientific findings. This chapter explains how altmetrics have emerged and how they are related to the academic use of social media. It also provides an overview of current altmetric tools and potential data sources for computing alternative metrics, such as blogs, Twitter, social bookmarking services, and Wikipedia.

1 Introduction

Scientific communication is, to a large degree, built upon the process of reading existing literature and situating one's own work within this broader research context. This context is primarily established by citing other scholars' publications (Weller and Peters 2012), and over time, these citations have become indicators of scholarly impact (Cronin 1984). Therefore, authors, journals, and/or papers that are most frequently cited are considered to be the most influential. Up to now, the various measures used to evaluate academic performance and impact are largely based upon the counting of citations. The disciplines of bibliometrics and scientometrics (Leydesdorff 1995) deal with the challenges of providing useful indicators and balancing the shortcomings of individual metrics (see also Haustein and Larivière 2015). Classical bibliometric indicators for measuring scholarly activity, such as publication numbers and citations, have a tradition that dates back to the creation of the Science Citation Index in the 1960s. Publication output and citations can thus have very practical implications for scholars, as they are used

K. Weller (✉)
GESIS Leibniz-Institute for the Social Sciences, Cologne, Germany
e-mail: katrin.weller@gesis.org

I.M. Welpe et al. (eds.), *Incentives and Performance*,
DOI 10.1007/978-3-319-09785-5_16

to evaluate individuals as well as institutions and can even influence job appointments, tenure, and project funding.

In addition, citation counts are used to retrieve information from publication databases. In fact, citation metrics were introduced at the Institute for Scientific Information's (ISI) Web of Knowledge to support new search functions in the Web of Science databases. Citations help in the identification of popular or highly recognized articles and can also be used to explore research topics by recommending similar publications; e.g., based on co-citations or bibliographic coupling. Thus, citation metrics help with both performance measurement and navigation within the enormous amount of scholarly literature.

With the growing importance of the Internet, additional approaches to calculating metrics that are based on Web links or download numbers—so-called "webometrics" (Thelwall 2008)—have also been explored. This chapter shows how webometrics have evolved into alternative metrics, or "altmetrics". Altmetrics are evaluation methods based on various user activities in social media environments. The next section thus describes the role of social media services in academia and the types of user activities that can possibly be measured through altmetrics. We then present the current state of altmetrics, both in practical application (Sect. 3) and research (Sect. 4).

2 The Rise of Social Media and Alternative Metrics

2.1 Web 2.0 and Science 2.0

For the last few decades, the Internet has been changing the way that scholars all over the world and across disciplines carry out their work. The Internet and related technologies influence the way that scholars gather research data, retrieve information and find relevant literature, present and distribute their research results, communicate and collaborate with colleagues, and teach and interact with their students (Tokar et al. 2012). These changes are closely connected to current initiatives to foster "open science" (Bartling and Friesike 2014; see also Friesike and Schildhauer 2015), i.e., opening up scholarly processes to enhance transparency and accessibility, including open access (Sitek and Bertelmann 2014) to research publications and to data (Pampel and Dallmeier-Tiessen 2014).

Today, many popular online spaces[1] are based on user-generated content, user networks, and user interactions. This was different 10 years ago, when Tim O'Reilly and Dale Dougherty of O'Reilly Media coined the term "Web 2.0" in a conference panel about the new Internet phenomena (O'Reilly 2005). They described a new era in which the Internet was no longer comprised of static

[1] According to the Alexa.com ranking as of June 2014, the most accessed Web sites include Facebook, YouTube, Wikipedia, Twitter, and LinkedIn.

Websites and content provided by just a few individuals or institutions. Instead, the "consumers" of Websites were increasingly contributing to the production of Web content [a phenomenon that has been labeled "prosumerism" (Buzzetto-More 2013) or "produsage" (Bruns 2008)] by sharing short texts, photos, videos, bookmarks, and other resources online through platforms that enabled user contributions with little effort and no programming skills. The infrastructural backbone of Web 2.0 has been a series of new platforms that enable content sharing and networking, which is comprised of interactions and connections among groups of users. Web 2.0 soon became known as the "Social Web," and Web 2.0 platforms were called "social media" (e.g., Kaplan and Haenlein 2010) or "social networking sites" (Nentwich and König 2014). This terminology is still used today, although the characteristics of the initial vision of Web 2.0 have been widely adopted across numerous Web sites and users may no longer be able to see a difference between the Web in general and social media in particular. At the same time, social media is being increasingly studied across scholarly disciplines, regarding, for example, the role that it plays in political communication and online activism (e.g., Faris 2013), journalism (e.g., Papacharissi 2009), disaster relief (e.g., Silverman 2014), and health care (e.g., McNab 2009).

The addition of "2.0" was quickly combined with many other concepts: integrating social media in teaching and learning became "e-learning 2.0" (Downes 2005), knowledge management using wikis and other social networking platforms that facilitate knowledge sharing (e.g., reference management communities) became "knowledge management 2.0" and libraries that added user-generated keywords to their catalogues were described as "library 2.0" (Maness 2006). Comparable names were coined for other concepts—both in everyday life and in academia. Similarly, the term "science 2.0" was quickly coined as an umbrella term for approaches to use social media to support knowledge exchange and workflows in academic environments (Waldrop 2008; Weller et al. 2007). Science 2.0 is closely related to eScience (Hey and Trefethen 2003), cyberinfrastructures (Hey 2005), and cyberscience (Nentwich 2003; Nentwich and König 2012). Tochtermann (2014, in press) provides an up-to-date definition of Science 2.0 as well as a summary of recent developments.

Many researchers feel that social media has the potential to help scholars collaborate and communicate more easily. Nentwich and König (2014) describe the potential which social networking sites offer to academics in more detail. They highlight functionalities for knowledge production, processing and distribution, as well as applications in institutional settings. But the uptake of social media usage among scientists has been rather slow. However, single studies seem to indicate that scholars from some disciplines are acting as 'early adopters' while others appear more skeptical in using social media. For example, Mahrt et al. (2014) have summarized several studies which show that Twitter is only used by a small percentage of researchers. Other studies show a higher uptake of social media tools in specific academic communities and disciplines (e.g., Haustein et al. 2013). It appears that Twitter was first popular in disciplines that are themselves related to the Web or computer-mediated communication (for example, the semantic Web research community; see Letierce et al. 2010). In a survey with researchers in the UK Procter et al. (2010, p. 4044) also observed that "computer science researchers are more likely to be frequent users and those in medicine and veterinary sciences less likely". The same study, however, did not find any significant differences in social media usage based on the scholars' age. Currently, such studies mainly consider researchers in specific countries and more research is needed to fully understand practices for using social media usage in academia and to explore interdisciplinary differences. Gerber (2012) has noted that many social media platforms are unfamiliar to most scholars (he found, for example, that 35.8 % of scholars are familiar with the social networking platform ResearchGate, 16.1 % with SlideShare, a site to share and browse presentation slides) and that the "best-known tools, however, were also those that were rejected by the majority of researchers, e.g., Twitter, which was rejected by 80.5 % of the respondents" (Gerber 2012, p. 16), i.e., for example, four out of five scholars had a decisively negative opinion of Twitter.

However, social media use should not only be determined by counting how many people have user accounts on specific platforms; who consumes content from the social media sites without actively contributing to them should also be taken into consideration. Passive use of social media is much more common; e.g., reading blogs or Wikipedia (Weller et al. 2010). Allen et al. (2013) have pointed out that the principle of social media is to push information (tailored to users' interests) to the public instead of waiting for users to pull the information from databases by searching. In this way, the dissemination of scholarly publications has increased in the health and medical domains (Allen et al. 2013). Furthermore, other studies show that social bookmarking platforms (i.e. services that allow sharing lists of bookmarked websites or scholarly references with a community, see e.g., Peters 2009) do indeed provide a notable coverage of the literature in specific domains, with Mendeley including more than 80 % (and up to 97 %) of research papers in given samples from different domains (Bar-Ilan 2012; Haustein et al. 2013; Li et al. 2012; Priem et al. 2012).

These findings reinforce the belief that content and user behavior of social media can be used as a type of indicator for measuring scholarly activity and impact.

2.2 The Search for Alternative Metrics

There are several shortcomings and challenges inherent in classic bibliometrics that have long been acknowledged by the research community (MacRoberts and MacRoberts 1989; see also Haustein and Larivière 2015). Notable shortcomings include the following: (1) Citations do not measure readership and do not account for the impact of scholarly papers on teaching, professional practice, technology development, and nonacademic audiences; (2) publication processes are slow and it can take a long time until a publication is cited; (3) publication practices and publication channels vary across disciplines and the coverage of citation databases, such as the Web of Science and Scopus, may favor specific fields; and (4) citation behavior may not always be exact and scholars may forget to acknowledge certain publications through citations or may tend to quote those papers that are already more visible due to a high number of citations.

Therefore, for many years, taking a variety of indicators into consideration in addition to citation counts has been strongly recommended, and the search for alternative metrics is older than the name “altmetrics.” This term became well-known through the Altmetrics Manifesto, written in 2010 by Priem et al. (2010) (after Priem first suggested the name altmetrics in a tweet). The manifesto opens with the statement that “no one can read everything” (Priem et al. 2010); the authors then go on to describe a landscape of social bookmarking systems (such as Mendeley, Delicious, or Zotero) and other social networking sites where fellow researchers can act as pointers to interesting literature (e.g., via Twitter or blogs). Just as with classic citation counts, the support of information retrieval is a major motivation for establishing altmetrics, but also just like classic bibliometrics, altmetrics have already been considered as novel indicators for identifying influential research and for evaluation. However, to date, altmetrics are not being used to, for example, make funding decisions, although some in the altmetrics community are advocating for this approach (Galligan and Dyas-Correia 2013; Lapinski et al. 2013; Piwowar 2013b).

Other authors have explicitly highlighted the different nature of altmetrics as obtained through social media and “that the most common altmetrics are not measuring impact, insofar as impact relates to the effect of research on [...] practice or thinking” (Allen et al. 2013, p. 1). However, altmetrics offer their own unique type of insight into research practices; for example, “many online tools and environments surface evidence of impact relatively early in the research cycle, exposing essential but traditionally invisible precursors like reading, bookmarking, saving, annotating, discussing, and recommending articles” (Haustein et al. 2013, p. 2). They can also be obtained faster than citation metrics because “social media mentions being available immediately after publication—and even before publication in the case of preprints—offer a more rapid assessment of impact” (Thelwall et al. 2013). Piwowar (2013a, p. 9) has stated the following four advantages of altmetrics: They provide “a more nuanced understanding of impact”, they provide “more timely data”, they include the consideration of alternative and “web-native scholarly products like datasets, software, blog posts, videos and more”, and they serve as “indications of impacts on diverse audiences”. Similar to the first argument, Lapinski et al. (2013, pp. 292–293) presented the idea of different “impact flavors” as “a product featured in mainstream media stories, blogged about, and downloaded by the public, for instance, has a very different flavor of impact than one heavily saved and discussed by scholars”. Furthermore, different activities within one platform may represent different levels (or flavors) of commitment with scholarly content. For example, it requires less commitment to tweet a link to a scholarly article than to write a critical blog post about that article. Fenner (2014) lists a couple of activities that indicate engagement and can be used for altmetrics such as discussing, recommending, viewing, citing, saving. An established classification of different impact flavors is not yet available and the impact of different tools and metrics still needs to be studied. However, it may be useful to distinguish the perspectives described in the next section.

2.3 Social Media Services as Candidates for Altmetrics

Altmetrics research is still in a very early phase, and since the publication of the manifesto by Priem et al. (2010), it has focused on exploring potential data sources. Here, we have organized the different dimensions of alternative metrics into four categories to provide an overview of potential sources for altmetrics data.

(1) Metrics on a new level. The first category refers to alternative sources for measuring the reach of traditional publications on an article-level basis. Often altmetrics initiatives are also described as “article-level metrics”, which highlights their difference from previous journal-level metrics, such as the impact factor. The collection and combination of various metrics for a single article (or for other types of scholarly publications, such as books and conference papers) is one of the core aims of altmetrics. Such metrics can be collected from various sources. Social bookmarking systems (Henning and Reichelt 2008) typically permit the counting of how often a publication has been bookmarked by users; bookmarking systems that are frequently used for altmetrics research are Mendeley (mendeley.com) and CiteULike (citeulike.org; Haustein et al. 2013). Both cover a large amount of scholarly literature. Other bookmarking services (e.g., Zotero and Bibsonomy) are used less often. Social media platforms based on social networking and status updates can be mined for mentions of research papers. Platforms to be considered are Facebook, Twitter, Google+, reddit, Pinterest, and LinkedIn. Adie and Roe (2013) have provided a comparison of how frequently scholarly papers are mentioned across different social media platforms, with Twitter being the richest resource. Besides the popular universal social media services, there are other social networking sites that are particularly dedicated to the academic community, including Academia.edu and ResearchGate. On most of these social media platforms, it is possible to count mentions of scholarly works in users’ posts and, moving to another level, count other users’ interactions with those posts, such as comments, likes, favorites, or retweets. Blogs and Wikipedia are other resources that include explicit citations to scholarly publications and may be used for evaluation purposes. Some blog platforms, such as Nature.com Blogs, Research Blogging, or ScienceSeeker, focus specifically on scientific content and are thus most promising. Other sources from the age of webometrics that have already been studied include presentation slides (Thelwall and Kousha 2008), online syllabi (Kousha and Thelwall 2008), research highlights as identified by Nature Publishing (Thelwall et al. 2013), and article downloads (Pinkowitz 2002). Mainstream media and news outlets can also be monitored as well as discussion forums and Amazon comments. Lin and Fenner (2013) provide an overview of article-level metrics used in the

PLoS database, including social shares on Twitter and Facebook, academic bookmarks from Mendeley and CiteULike, and citations from blogs and Wikipedia.

(2) Metrics for new output formats. The second dimension of using altmetrics is in establishing ways to measure alternative types of research output: not only classic publication formats such as journal articles or books may be measured but also other products such as blog posts, teaching material or software products. Buschman and Michalek (2013) pointed out that negative research results are increasingly not being published in peer-reviewed journals; rather, they are showing up in alternative formats such as blogs; they therefore argue that those information sources need to be considered as well for measuring scholars' academic performance. Blogs are probably the main type of alternative textual research output, but nontextual output is also of interest. Popular forms of this type of information are YouTube videos with scientific content or other video material like recorded lectures, uploaded presentation slides on SlideShare, research data repositories, or source code on Github.

(3) Aggregated metrics for researchers. Finally, altmetrics may also refer to alternative ways to measure the popularity of single researchers. Buschmann and Michalek (2013, p. 38) have argued that the "greatest opportunity for applying these new metrics is when we move beyond just tracking article-level metrics for a particular artifact and on to associating all research outputs with the person that created them. We can then underlay the metrics with the social graph of who is influencing whom and in what ways". Social media offers the opportunity to not only aggregate the output of individual scholars across different channels, but also to monitor the mentions and the level of attention that individual receives throughout these channels (e.g., followers on Twitter or Wikipedia articles mentioning a researcher). Bar-Ilan et al. (2012) have investigated the "footprints" that scholars leave through their activity in different online environments and then related these author-level metrics to citations from Google Scholar and the Web of Science.

(4) Metrics based on alternative forms of citations. Another objective of the altmetrics community is to take into account that there are different types of audiences and that scholarly activities do not only have an impact within the active research community but also within a general public. But metrics based on citations will only capture impact on those researchers that actively contribute to the process of writing and citing academic publications. Altmetrics look at citation-like activities by academics (e.g., linking to a journal publication in a researcher's blog post) as well as activities by students, science journalists or non-specified general audiences (e.g., writing a Wikipedia article about a notable researcher). Altmetrics are not necessarily based on citation-like processes, but can comprise other measures that reflect "readership", e.g., bookmarks and downloads of scholarly articles. Haustein et al. (2014b) show how publications with the most mentions on Twitter or with high readership-metrics on Mendeley are often related to topics that are of interest to a general public (e.g., the Chernobyl accident). Of course, this last category is not isolated from the other three, because online activities create their various interaction networks: a scholar may upload an academic YouTube video

which then is being discussed by other scholars on Twitter or by some journalist in a blog post or commented on by the general public in a Facebook group.

Altogether these four categories reflect the different motivations for establishing new metrics as research performance indicators. Altmetrics should respect different sources of academic output besides peer reviewed journals and they should do so on different levels of granularity: down to single articles and up to aggregated metrics for researchers and their entire portfolio. And in order to do this comprehensively, they should also reflect the impact scholarly work can have for different communities, both academic and non-academic. While the exact role and significance of single altmetric indicators is still being studied (see Sect. 4), some of them have recently been transferred into practical applications that may already affect scholars and the assessment of academic performance, as we will see in the next section.

3 Applied Altmetrics in Accessing Research Performance

3.1 Stakeholders and Objectives

Altmetrics affect scholars in at least two ways—researchers can use them to retrieve interesting literature and to promote their own work (similar to the press offices of universities). Piwowar and Priem (2013, p. 10) discussed the use of placing altmetrics on one's CV and listed several advantages for doing so, including to "uncover the impact of just-published work" and to "legitimize all types of scholarly products". If this becomes standard practice, funding agencies and tenure committees might one day need to consider altmetric indicators for assessments, but as previously mentioned, this is not yet happening. In addition, as the quality of altmetrics—especially regarding their use for assessment—is being questioned by some in the scientific community (Cheung 2013), their practical impact in the area of assessment might be small. And Fenner (2014) also points out that it might be easy to manipulate some altmetric indicators by self-promotion and gaming. On the other hand, Buschman and Michalek (2013) argue that funders will also profit from altmetrics if they want to evaluate the immediate impact of a project right after or during the funding period. And some research councils are considering the use of altmetrics for evaluations (Viney 2013). Liu and Adie (2013a, p. 31) emphasize that all stakeholders may have different visions for new impact metrics and that "impact is a multi-faceted concept [. . .] and different audiences have their own views of what kind of impact matters and the context in which it should be presented: researchers may care about whether they are influencing their peers, funders may care about re-use or public engagement and universities may wish to compare their performance with competing institutions".

Meanwhile, other stakeholders are interested in measuring scholarly communication through various online channels. For example, librarians are encouraged to keep up to date with new metrics to better help their customers find specific publications (Baynes 2012; Lapinski et al. 2013). In addition, publishers, as well as certain social media companies, are also interested in altmetrics. In 2013, the publisher Elsevier bought the social bookmarking platform Mendeley. As one of the largest providers of citation metrics (via their database Scopus), Elsevier appears to be particularly interested in ongoing altmetrics developments. Open access platform providers are also contributing to the discussion of altmetrics, like the open access publisher PLoS that uses data from Mendeley, CiteULike, Facebook, and Twitter (Lin and Fenner 2013). Lastly, a new type of stakeholder has entered the scene: companies that are creating new types of products and services, especially centered on altmetrics. We will take a closer look at these new products in the next section.

3.2 First Products Based on Altmetrics

Even though the exact value and the informative value of altmetrics remain vague, products have been released that are based on altmetrics approaches and that mainly combine a variety of new measures into one platform. The first product in this area is Altmetric.com, which is supported by Digital Science (owned by the same company as *Nature*). The purpose of Altmetric.com is to aggregate all article-level metrics that refer to the same publication. It thus monitors tweets, Facebook and blog posts, bookmarks, news, and other sources for mentions of digital object identifiers (DOIs) or other standard identifiers that relate to publications and research data, such as PubMedID or Handle (Liu and Adie 2013a). From these mentions, an "altmetrics score" is computed for each publication. The company's Website states that "articles for which we have no mentions are scored 0. Though the rate at which scientists are using social media in a professional context is growing rapidly, most articles will score 0; the exact proportion varies from journal to journal, but a mid-tier publication might expect 30–40 % of the papers that it publishes to be mentioned at least once, with the rate dropping rapidly for smaller, niche publications" (Altmetric.com n.d.). Altmetric.com is selling access to its data to researchers and institutions who want to monitor who is mentioning them across different online channels. It also provides the so-called "altmetrics badge", an icon that illustrates the altmetric score of a publication. This badge is also used by Elsevier's Scopus. Finally, Altmetric.com also provides information about the users who interact with scholarly publications through social media. As social media platforms like Twitter and Mendeley include information about their users' social networks, some of these information can be mined for gathering information about the audiences of scholarly publications. One may see, for example, where most people who tweet about a scholarly publication live (based on the location Twitter users mention in their profile) or whether people who bookmarked a specific publication are students or senior researchers (based on their Mendeley profiles). Of course these demographics provided by Altmetric.com have some limitations as they are entirely based on the information users include in their social media profiles, which might not always be complete or correct. See Haustein et al. (2014b) for an example how such information can be obtained from Altmetric.com and used in altmetrics research.

The second product, ImpactStory (impactstory.org), is a not-for-profit organization founded by Jason Priem and Heather Piwowar and supported by the Alfred P. Sloan Foundation (Lapinski et al. 2013; Piwowar 2013b). ImpactStory builds metrics around individual researchers rather than single papers. Researchers who log into the platform can import their publications and other identifiable work, such as presentations on SlideShare and source code published at Github, and track mentions across a variety of sources and classic citations. ImpactStory used to be free to use but started charging a subscription fee in autumn 2014. Both data and source code are currently openly available. A third player on the altmetrics market is Plum Analytics, which also collects article-level metrics.

Liu and Adie (2013b, p. 153) have summed up the current state of the art as follows: “However, all of the tools are in their early stages of growth. Altmetrics measures are not standardized and have not been systematically validated; there has been no clear consensus on which data sources are most important to measure; and technical limitations currently prevent the tracking of certain sources, such as multimedia files”. Applications that work with altmetrics have to face several technical challenges (Liu and Adie 2013a, b). Just like the automatic detection of classic citations, the collection of social mentions is not a trivial process. Collecting data from social media platforms requires data access, which is sometimes enabled via application programming interfaces (APIs). However, APIs sometimes have specific limitations, such as the Twitter API, which does not allow tracing back to older tweets (Gaffney and Puschmann 2014). DOIs or other unique identifiers make tracing citations easier, but they are not always available. Currently, many technical pitfalls remain, and certain information may get lost during the data collection process. Not only providers of altmetrics tools are facing these challenges. The scientific community studying the use of alternative metrics has to deal with them as well. Haustein et al. (2014c, p. 659) sum up technical challenges for working with Twitter-based altmetrics as follows: “A general problem of social media-based analyses is that of data reliability. Although most social media services provide application programming interfaces (API) to make usage data accessible, we still do not know if it is possible to collect every tweet, if there are missing data, or what effects download or time restrictions have on available data. In addition, the Altmetric coverage of Twitter may be incomplete because of technical issues, such as server or network downtime. Moreover, an article may be tweeted in a way that is not easily automatically identified (e.g., “See Jeevan’s great paper in the current Nature!”)”.

4 Altmetrics Research

The current research on altmetrics is still at a very early stage, but it is developing quickly (thus, this section can only highlight some remarkable examples). The history of altmetrics research is also summarized by Fenner (2014). Scholars with different backgrounds (but quite a few of them from library and information science or related fields and with prior experience in bibliometrics, among them Mike Thelwall and Judit Bar-Ilan) are acting as pioneers in the development of new metrics and new research approaches. Specialized workshop series such as the Altmetrics Workshops hosted at ACM Web Science conferences are being initialized and sessions organized at established scientometrics conferences such as the International Society of Scientometrics and Informetrics conference (ISSI). Most of this current research is dedicated to outlining the quality and scope of altmetrics indicators. Case studies can be broadly described as those that compare metrics across platforms (either alternative or classic) or those that study scholar performance or participation in social media across specific disciplines.

One question that is of high interest for the entire altmetrics community is whether social media mentions predict subsequent citation rates—or at least correlate to some degree with classic metrics. The most comprehensive comparison of altmetrics and citations to date was conducted by Thelwall et al. (2013). These authors looked at 11 different social media resources. However, metrics assessed in this study could not predict subsequent citations. which once more indicates that most altmetrics measure some different form of impact then citations.

Based on the current literature, Mendeley appears to provide the most article-level metrics: Zahedi et al. (2014) found Mendeley readership metrics for 62.6 % of all publications in their test sample, and Priem et al. (2012) found that close to 80 % of their publication set was included on Mendeley. In contrast, Priem et al. (2012) found only around 5 % of their sample papers cited on Wikipedia and Shuai et al. 2013 even less than this—which shows that different social media channels will provide very different metrics. Haustein et al. (2014c) could also show that the field is still developing: they identified around 20 % of biomedical papers published in 2012 being mentioned in at least one tweet on Twitter, twice as many as in 2011.

Mendeley readership and citation counts were found to show a moderate and significant correlation by several studies (Bar-Ilan 2012; Haustein et al. 2013; Li et al. 2012; Priem et al. 2012; Zahedi et al. 2014). Shuai et al. (2013) also report correlations for citations and mentions in Wikipedia, while Samoilenko and Yasseri (2013) did not find significant correlations of Wikipedia metrics with academic notability for researchers from four disciplines.

Li and Gillet (2013) and Thelwall and Kousha (2014) have collected user data from social media platforms (from Mendeley and Academia.edu, respectively) to determine how user activities relate to the professional levels of academia. Li and Gillet (2013) found that Mendeley users with higher impact were senior scholars with many co-authorships. Thelwall and Kousha (2014) showed that more senior philosophy scholars on Academia.edu get more page views for their profile pages, but they did not observe any positive correlation of page views and citations received.

Disciplinary differences appear to underlay most current altmetrics. Zahedi et al. (2014) conducted a study on altmetrics across disciplines and found that journals from different disciplines are represented to a very different extent on Mendeley (with arts and humanities being the least represented), Wikipedia, Twitter, and Delicious. Several studies have taken a close look at single disciplines and their depiction through altmetrics. Haustein et al. (2014a) focused on astrophysicists, and Haustein et al. (2014b, c) analyzed biomedicine scholars on Twitter. Holmberg and Thelwall (2014) then compare Twitter usage across researchers from 10 disciplines.

All these examples show that much still needs to be studied before we fully understand what alternative metrics derived from social media platforms or other online data can tell us about research activities and scholarly impact.

Conclusion

The call for alternative approaches to measure scholarly performance and impact has been heard and experts from academia, publishing business and research councils are engaged in the present discussion of new indicators. The label "altmetrics" has allowed this community to connect and establish venues for discussing their ideas about how to make sense of user interactions with scholarly content in various online environments. Expected outcomes of this discussion vary, and while some are skeptical about the practical relevance of altmetrics, others believe that altmetrics will become seamlessly integrated to other performance measurements, like Piwowar (2013a, p. 9) puts it: "Of course, these indicators may not be "alternative" for long. At that point, hopefully we'll all just call them *metrics*".

Before this can happen, much more work is needed in order to better understand the nature of user behavior in social media environments and the value of individual metrics obtained through measuring this user behavior. Current research focuses on this task and is step-by-step creating a map of the altmetrics landscape. Meanwhile single publishers have started to provide aggregated altmetrics for publications; other stakeholders enter the altmetrics market. These are still rather niche services, but if altmetrics become popular with a broader community and gain influence, it is quite likely that peopcheckle people might change their behavior in order to achieve better scores or even try to game. This of course is not new and happens as well with traditional bibliometrics and citation counts (e.g., Frey and Osterloh 2011), for example, through self-citations or cartels. But bibliometricians have found ways to adjust their indicators in order to respect such behavior. Consequently, independent research should monitor the use of altmetrics equally carefully and keep on studying how tools, users and metrics interact. For now,

(continued)

both social media platforms as well as publication databases offer a useful environment for browsing scholarly references and discovering interesting information based on peer recommendations so that many academics can already benefit from altmetrics for their everyday work.

References

Adie E, Roe W (2013) Altmetric: enriching scholarly content with article-level discussion and metrics. Learn Publish 26(1):11–17

Allen HG, Stanton TR, Di Pietro F, Moseley GL, Sampson M (2013) Social media release increases dissemination of original articles in the clinical pain sciences. PLoS ONE 8(7): e68914

Altmetrics.com (no date) What does Altmetric do? http://www.altmetric.com/whatwedo.php#score. Accessed 14 June 2014

Bar-Ilan J (2012) JASIST 2001-2010. Bull Am Soc Inf Sci Technol 38(6):24–28

Bar-Ilan J, Haustein S, Peters I, Priem J, Shema H, Terliesner J (2012) Beyond citations. Scholars, visibility on the social Web. In: Proceedings of the 17th international conference on science and technology indicators (STI conference), Montreal, pp 98–109

Bartling S, Friesike S (eds) (2014) Opening science: the evolving guide on how the internet is changing research, collaboration and scholarly publishing. Springer, Heidelberg; New York

Baynes G (2012) Scientometrics, bibliometrics, altmetrics: some introductory advice for the lost and bemused. Insights 25(3):311–315

Bruns A (2008) Blogs, Wikipedia, second life, and beyond. From production to produsage. Peter Lang, New York

Buschman M, Michalek A (2013) Are alternative metrics still alternative? Bull Am Soc Inf Sci Technol 39(4):35–39

Buzzetto-More NA (2013) Social media and prosumerism. Issues Inf Sci Inf Technol 10:81–93

Cheung M (2013) Altmetrics: too soon for use in assessment. Nature 494(7436):176

Cronin B (1984) The citation process. The role and significance of citations in scientific communication. Taylor Graham, London

Downes S (2005) E-learning 2.0. eLearn Magazine. http://elearnmag.acm.org/featured.cfm?aid=1104968. Accessed 15 June 2014

Faris DM (2013) Dissent and revolution in a digital age. Social media, blogging and activism in Egypt. Tauris, London

Fenner M (2014) Altmetrics and other novel measures for scientific impact. In: Bartling S, Friesike S (eds) Opening science. The evolving guide on how the internet is changing research, collaboration and scholarly publishing. Springer, Heidelberg; New York, pp 179–190

Frey B, Osterloh M (2011) Ranking games. University of Zurich Department of Economics Working Paper No. 39. doi:10.2139/ssrn.1957162. Accessed 22 June 2014

Friesike S, Schildhauer T (2015) Open science: many good resolutions, very few incentives, yet. In: Welpe IM, Wollersheim J, Ringelhan S, Osterloh M (eds) Incentives and performance – governance of research organizations. Springer, Cham

Gaffney D, Puschmann C (2014) Data collection on Twitter. In: Weller K, Bruns A, Burgess J, Mahrt M, Puschmann C (eds) Twitter and society. Peter Lang, New York, pp 55–68

Galligan F, Dyas-Correia S (2013) Altmetrics: rethinking the way we measure. Ser Rev 39(1):56–61

Gerber A (2012) Online trends from the first German trend study on science communication. In: Tokar A, Beurskens M, Keuneke S, Mahrt M, Peters I, Puschmann C, van Treeck T, Weller K (eds) Science and the internet. Düsseldorf University Press, Düsseldorf, pp 13–18

Haustein S, Larivière V (2015) The use of bibliometrics for assessing research: possibilities, limitations and adverse effects. In: Welpe IM, Wollersheim J, Ringelhan S, Osterloh M (eds) Incentives and performance – governance of research organizations. Springer, Cham

Haustein S, Peters I, Bar-Ilan J, Priem J, Shema H, Terliesner J (2013) Coverage and adoption of altmetrics sources in the bibliometric community. In: Proceedings of the 14th international society of scientometrics and informatics conference, Vienna, Austria, 15–19th July 2013, pp 1–12

Haustein S, Bowman TD, Holmberg K, Peters I, Larivière V (2014a) Astrophysicists on Twitter: an in-depth analysis of tweeting and scientific publication behavior. Aslib J Inf Manag 66 (3):279–296

Haustein S, Larivière V, Thelwall M, Amyot D, Peters I (2014b) Tweets vs. Mendeley readers: how do these two social media metrics differ? Inf Technol 56(5) (in press)

Haustein S, Peters I, Sugimoto CR, Thelwall M, Larivière V (2014c) Tweeting biomedicine: an analysis of tweets and citations in the biomedical literature. J Am Soc Inf Sci Technol 65 (4):656–669

Henning V, Reichelt J (2008) Mendeley. A Last.fm for research? In: Proceedings of the 4th IEEE international conference on eScience, Indianapolis, pp 327–328

Hey T (2005) Cyberinfrastructure for e-Science. Science 308(5723):817–821

Hey T, Trefethen A (2003) The data deluge: an e-Science perspective. In: Berman F, Fox CG (eds) Grid computing. Making the global infrastructure a reality. Wiley, Chichester, pp 809–824

Holmberg K, Thelwall M (2014) Disciplinary differences in Twitter scholarly communication. Scientometrics. doi:10.1007/s11192-014-1229-3

Kaplan AM, Haenlein M (2010) Users of the world, unite! The challenges and opportunities of social media. Bus Horizons 53(1):59–68

Kousha K, Thelwall M (2008) Assessing the impact of disciplinary research on teaching. An automatic analysis of online syllabuses. J Am Soc Inf Sci Technol 59:2060–2069

Lapinski S, Piwowar H, Priem J (2013) Riding the crest of the altmetrics wave: how librarians can help prepare faculty for the next generation of research impact metrics. Coll Res Libr News 74 (6):292–294 + 300

Letierce J, Passant A, Breslin J, Decker S (2010) Understanding how Twitter is used to spread scientific messages. In: Proceedings of the WebSci10: extending the frontiers of society on-line, Raleigh

Leydesdorff L (1995) The challenge of scientometrics. The development, measurement and self-organization of scientific communication. DSWO, Leiden

Li N, Gillet D (2013) Identifying influential scholars in academic social media platforms. In: Proceedings of the 2013 IEEEACM international conference on advances in social networks analysis and mining. ACM, New York, pp 608–614

Li X, Thelwall M, Giustini D (2012) Validating online reference managers for scholarly impact measurement. Scientometrics 91(2):461–471

Lin J, Fenner M (2013) The many faces of article-level metrics. Bull Am Soc Inf Sci Technol 39 (4):27–30

Liu J, Adie E (2013a) Five challenges in altmetrics: a toolmaker's perspective. Bull Am Soc Inf Sci Technol 39(4):31–34

Liu J, Adie E (2013b) New perspectives on article-level metrics: developing ways to assess research uptake and impact online. Insights UKSG J 26(2):153–158

MacRoberts M, MacRoberts BR (1989) Problems of citation analysis. A critical review. J Am Soc Inf Sci 40(5):342–349

Mahrt M, Weller K, Peters I (2014) Twitter in scholarly communication. In: Weller K, Bruns A, Burgess J, Mahrt M, Puschmann C (eds) Twitter and society. Peter Lang, New York, pp 399–410

Maness J (2006) Library 2.0 theory. Web 2.0 and its implications for libraries. Webology 3(2): Article 25. http://www.webology.ir/2006/v3n2/a25.html. Accessed 14 June 2014

McNab C (2009) What social media offers to health professionals and citizens. Bull World Health Organ 89(8):566–567
Nentwich M (2003) Cyberscience. Research in the age of the internet. Austrian Academy of Sciences Press, Vienna
Nentwich M, König R (2012) Cyberscience 2.0. Research in the age of digital social networks. Campus Verlag, Frankfurt; New York
Nentwich M, König R (2014) Academia goes facebook? The potential of social network sites in the scholarly realm. In: Bartling S, Friesike S (eds) Opening science. The evolving guide on how the internet is changing research, collaboration and scholarly publishing. Springer, Heidelberg; New York, pp 107–124
O'Reilly T (2005) What is Web 2.0? Design patterns and business models for the next generation of software. http://oreilly.com/web2/archive/what-is-web-20.html. Accessed 14 June 2014
Pampel H, Dallmeier-Tiessen S (2014) Open research data. From vision to practice. In: Bartling S, Friesike S (eds) Opening science. The evolving guide on how the internet is changing research, collaboration and scholarly publishing. Springer, Heidelberg; New York, pp 213–224
Papacharissi Z (ed) (2009) Journalism and citizenship. New agendas in communication. Routledge, New York
Peters I (2009) Folksonomies. Indexing and retrieval in Web 2.0. De Gruyter/Saur, Berlin
Pinkowitz I (2002) Research dissemination and impact. Evidence from web site downloads. J Financ 57:485–499
Piwowar H (2013a) Altmetrics. What, why and where? Bull Am Soc Inf Sci Technol 39(4):8–9
Piwowar H (2013b) Altmetrics: value all research products. Nature 493(7431):159
Piwowar H, Priem J (2013) The power of altmetrics on a CV. Bull Am Soc Inf Sci Technol 39 (4):10–13
Priem J, Taborelli D, Groth P, Nylon C (2010) Alt-metrics. A manifesto. http://altmetrics.org/manifesto/. Accessed 15 June 2014
Priem J, Piwowar H, Hemminger B (2012) Altmetrics in the wild. Using social media to explore scholarly impact. arXiv:1203.4745v1. Accessed 15 June 2014
Procter R, Williams R, Stewart JK, Poschen M, Snee H, Voss A, Asgari-Targhi M (2010) Adoption and use of Web 2.0 in scholarly communications. Philos Trans R Soc A Math Phys Eng Sci 368(1926):4039–4056
Samoilenko A, Yasseri T (2013) The distorted mirror of Wikipedia. A quantitative analysis of Wikipedia coverage of academics. http://arxiv.org/abs/1310.8508. Accessed 15 June 2014
Shuai X, Jiang Z, Liu X, Bollen J (2013) A comparative study of academic and Wikipedia ranking. In: JCDL'13 Proceedings of the 13th ACM/IEEE-CS joint conference on digital libraries. ACM, Indianapolis, pp 25–28
Silverman C (ed) (2014) Verification handbook. A definitive guide to verifying digital content for emergency coverage. European Journalism Centre, Maastricht
Sitek D, Bertelmann R (2014) Open access. A state of the art. In: Bartling S, Friesike S (eds) Opening science. The evolving guide on how the internet is changing research, collaboration and scholarly publishing. Springer, Heidelberg; New York, pp 139–145
Thelwall M (2008) Bibliometrics to webometrics. J Inf Sci 34(4):605–621
Thelwall M, Kousha K (2008) Online presentations as a source of scientific impact? An analysis of PowerPoint files citing academic journals. J Am Soc Inf Sci Technol 59:805–815
Thelwall M, Kousha K (2014) Academia.edu: social network or academic network? J Am Soc Inf Sci Technol 65(4):721–731
Thelwall M, Haustein S, Larivière V, Sugimoto C (2013) Do altmetrics work? Twitter and ten other social web services. PLoS ONE 8(5)
Tochtermann K (2014) How science 2.0 will impact on scientific libraries. Inf Technol 56(5) (in press)
Tokar A, Beurskens M, Keuneke S, Mahrt M, Peters I, Puschmann C, van Treeck T, Weller K (eds) (2012) Science and the internet. Düsseldorf University Press, Düsseldorf
Viney I (2013) Altmetrics: research council responds. Nature 494(7436):176

Waldrop MM (2008) Science 2.0. Sci Am 298(5):68–73

Weller K, Peters I (2012) Twitter for scientific communication. How can citations/references be identified and measured? In: Tokar A, Beurskens M, Keuneke S, Mahrt M, Peters I, Puschmann C, van Treeck T, Weller K (eds) Science and the internet. Düsseldorf University Press, Düsseldorf, pp 211–224

Weller K, Mainz D, Mainz I, Paulsen I (2007) Wissenschaft 2.0? Social Software im Einsatz für die Wissenschaft. In: Ockenfeld M (ed) Information in Wissenschaft, Bildung und Wirtschaft. Proceedings der 29. Online-Tagung der DGI, Frankfurt am Main, pp 121–136

Weller K, Dornstädter R, Freimanis R, Klein RN, Perez M (2010) Social software in academia: three studies on users' acceptance of Web 2.0 services. In: Proceedings of the 2nd web science conference (WebSci10), Raleigh

Zahedi Z, Costas R, Wouters P (2014) How well developed are altmetrics? A cross-disciplinary analysis of the presence of'alternative metrics' in scientific publications. http://arxiv.org/abs/1404.1301. Accessed 15 June 2014

Open Science: Many Good Resolutions, Very Few Incentives, Yet

Sascha Friesike and Thomas Schildhauer

Abstract In recent years, a movement has emerged, which assembles itself under the umbrella term "Open Science". Its intent is to make academic research more transparent, collaborative, accessible, and efficient. In the present article, we examine the origins, various forms, and understandings of this movement. Furthermore, we put the aims of individual groups associated with Open Science and the academic realities of their concepts into context. We discuss that much of what is known as Open Science can be viewed through the prism of a social dilemma. From this perspective, we explain why the concept of Open Science finds a lot of support in theory, yet struggles in practice. We conclude the article with suggestions on how to foster more Open Science in practice and how to overcome the obstacles it is currently facing.

1 Science: An Inherently Open System

The concept of Open Science can be somewhat misleading. Ever since the invention of the journal system in the late seventeenth century, the idea of academic research has been to make results open or public (the word "to publish" comes from Latin *publicatio* "making public"). Before the invention of the journal system, science was indeed closed. Individual researchers encountered obstacles claiming ownership of concepts or findings. They would send out anagrams to "file" proof of evidence. Once a finding was known to others, the anagram could prove who found it first. However, without the finding itself, the anagram was of little value. Naturally, this was a highly ineffective system, which led researchers to make the same discoveries over and over again. To counter this ineffectiveness, to avoid

S. Friesike (✉)
Alexander von Humboldt Institute for Internet and Society, Berlin, Germany
e-mail: friesike@hiig.de

T. Schildhauer
Institute of Electronic Business, Alexander von Humboldt Institute for Internet and Society, Berlin, Germany
e-mail: schildhauer@ieb.net

I.M. Welpe et al. (eds.), *Incentives and Performance*,
DOI 10.1007/978-3-319-09785-5_17

repetitive studies, and to make sure that original inventors would receive credit, the journal system was brought to life. Researchers would publish findings under their name in order to claim authorship. The first journal, *Philosophical Transactions of the Royal Society* was founded in 1665. The next 200 years brought upon an enormous extension of the scientific community with a few polymaths during the renaissance and a million scientists in 1850. Today, we estimate that there are 100 million people involved in science worldwide (Bartling and Friesike 2014). As manpower increased, so did the complexity of the scientific community, affecting its disciplines and sub-disciplines. This growing complexity is also reflected in the development of the journal system. While early journals would cover vast amounts of topics (like *Science* or *Physical Review Letters*), their more recent counterparts focus on increasingly specialized subject-matters (like the *Journal of Services Marketing* or the *Journal of Mathematics Teacher Education,* for example). Our scientific system as it is today is the outcome of a slow evolution. It began with the introduction of the journal system and reflects the many developments a growing number of researchers continuously produce. Today, the group of people who form the Open Science movement (Woelfle et al. 2011; Nielsen 2012) challenge what has been conventionally perceived as the most efficient way to publish and diffuse knowledge, questioning whether this approach—maintained over the last 200 years—is really in the best interest of both, the research community and society as a whole.

2 The Origins of Open Science

The idea of Open Science is tightly intertwined with the Internet. Web technologies allow an unprecedented exchange of research data, designs, concepts and more. However, most incentives in the academic system can be traced back to the paper-based publication system with their roots predating the creation of the World Wide Web. Examples are monetary awards for publications in selected journals (WU Wien 2014) or the grant of a tenured position based on publications. The paper-based publication system developed at a time when text printed on paper was the most efficient means of disseminating academic knowledge. Before anything went off to a printing press, it had to be in the correct and finalized form. Any mistake, regardless of its size, could only be corrected in the next edition—if there ever was one. That led to an academic publishing system that focuses strongly on the pre-publication process. For many publication formats, the period between the moment when a manuscript is handed in to an editor and the moment when it is published can last several years. This can become especially bitter when an article is turned down after several rounds of revisions, and is then sent to another journal to go through the same process all over again. Some articles stay in the pre-publication process for over half a decade. In business and economics for instance, the *average* pre-publication time is about 18 months (Björk and Solomon 2013). In the case that an article is rejected after 18 months, and is then sent to another journal, a great deal of time would have accumulated in the process. Many articles never actually get published because no author is willing to carry them through the entire process. To illustrate this common occurrence, let us assume the case of a PhD student who finishes his or her Ph.D. and then pursues a career outside of academia. Reviewer requests are ignored—as the PhD graduate now has other things on their mind—and the work, regardless of how good it is, never gets published. Therefore, potentially relevant research remains unseen by fellow researchers. This has drastic consequences for the research system. For example, considering the case that someone else comes up with the same research question in the future, he or she may pursue research that has already been carried out. Instead of improving existing knowledge, the researcher will likely replicate something that was already done yet got lost in the pre-publication process.

Today, some disciplines use pre-publication services to "publish" what has not yet been published (physics for instance e.g., arXiv.org). Researchers use a (pre-) publication outlet to show their latest results and to be able to read current findings from other researchers. Only reading what is published in traditional, paper-based journals would make current readings lag at least 1 year behind. Researchers from other disciplines must often endure long waiting periods, hoping that they survive the pre-publication process before other publications with similar findings surface.

Aside from the process, the *content* of a publication is also strongly related to our paper-based heritage. This means that a publication is basically a written description of a research process, including its findings and interpretation. Over time, the form of publications has become standardized. Today, journals explicitly instruct researchers on what to include and how their articles should be structured. However, what is *not* included is also what interests the Open Science movement. Paper-based articles do not enable the presentation of vast amounts of research data, nor can they present interactive elements, moving pictures, or executable code. It is this cultural heritage that limits the presentation of research findings to the confines of printed paper (Boulton et al. 2011)—a medium hardly any researchers use today, considering the wide usage of online resources to find and read articles within the scientific community.

The personal computer has long substituted the traditional library in regards to journal-based publications. The fact that what gets published continues to resemble what can be printed on paper is as if one would design cars to best be pulled by horses.

With the Internet playing a central role in more and more aspects in our daily lives, many researchers feel that academic knowledge creation has yet to fully take advantage of the possibilities the web has to offer (Meyer and Schroeder 2013). This can seem somewhat ironic considering the intent behind the creation of the World Wide Web 25 years ago. It did not necessarily seek to transform the many aspects in our daily lives to the extent that it has, but rather to foster the exchange of information among scientists (Berners-Lee 1989).

3 Forms of Open Science

What is today labeled as Open Science is less a concept that can be easily defined, as it is a diverse movement with various interests. When looking for a definition of what Open Science actually means, Nielsen (2011) is most frequently cited: "*Open science is the idea that scientific knowledge of all kinds should be openly shared as early as is practical in the discovery process*." Yet, it remains unclear or at least debatable what "open" in the context of academic research truly means. When we compare the idea of Open Science to Open Source Software, this ambiguity becomes more evident. In Open Source Software "open" means that anyone can see and edit the source code of a program. In contrast, in Open Science there is no consensus on what "open" means. Furthermore, different scholars, activists, and interest groups advocate for various forms and understandings of openness in science. Here are the most prevailing ones:

3.1 Open Means Transparent

This form of Open Science probably comes closest to the commonly held notion of "open" within the Open Source Software community. When open means transparent, Open Science is understood as a form of research that publishes anything that may increase understanding or aid the reproduction of presented findings. This understanding of Open Science can have a multitude of implications, given the differences in methods and data between research projects and disciplines. Advocates of this understanding support the concepts of open research data (making the primary research data available and reusable) (McCullough 2009; Piwowar 2011; Fecher et al. 2014), open code (e.g., publishing the code used to run statistical tests), and open notebook science (sharing lab reports online, allowing fellow researchers to understand every single step in the research process including negative results). The foundation of this argument maintains that science, as a whole, would witness an increase in quality once researchers would ensure that their findings are reproducible in order to enable other researchers to understand how the entire process was previously carried out (Stodden 2009). An open peer review system would allow readers to add thoughts to an article. A group larger than today's reviewers would be able to assess the quality of an article (Nature 2006).

3.2 Open Means Collaborative

In this form of understanding, "open" in Open Science can be understood as "anyone can join". Under this definition, the research process is made public as early as possible. Research questions are posted and lab reports are submitted online, enabling other researchers to see what others are currently working on in order to allow for comments and to invite collaborators. The main idea behind this approach is to get as much external feedback as early as possible. This collaborative process helps avoid repeated mistakes and duplicated research efforts, thereby speeding up the research process overall. A famous example of this form of Open Science is the Polymath Project (Gowers and Nielsen 2009; Cranshaw and Kittur 2011), in which mathematician Timothy Gowers posted a research question online (seeking to find a new combinatorial proof to the density version of the Hales–Jewett theorem). Fellow mathematicians contributed to his post leading to a solution within 7 weeks. While the first problem was solved, another problem arose, which developed into a spin-off of sorts. The results from the Polymath project were published under the pseudonym D.H.J. Polymath (Polymath 2010, 2012). Today, the Polymath Project is frequently cited as a prime example of Open Science in action. Openly publishing data and code allow other researchers to reuse information that may be highly beneficial in unraveling answers to novel research questions. Securing the right to reuse published data would speed up a gridlocked research process, making the entire system more efficient.

3.3 Open Means a Broader Understanding of Impact

This understanding of Open Science is concerned primarily with the impact measurement in academia. Most scholars argue that current impact measurements such as the Impact Factor and the H-index have considerable downsides. There is some consensus surrounding the need to update standards in order to better incentivize researchers to produce quality work instead of quantity in research papers. Most of the concepts presented in this realm are labeled “altmetrics”. They measure a variety of key figures to better describe the actual impact research has. Many of the figures included in altmetrics come from social media. For a more in-depth explanation of this concept, please refer to the chapter on social media and altmetrics written by Weller in this book (Weller 2015).

3.4 Open Means Open to the Public

Most of what is discussed among supporters of Open Science stays within the realm of professional research. However, there are efforts in the Open Science community to engage the general public. This engagement comes in two common forms: *First*, attempts have been taken to write academic research in a manner that is understandable to a non-expert audience. Supporters of this idea claim that scientific research is often phrased in a way that makes it almost impossible for outsiders to understand what the research is about (Cribb and Sari 2010). Consequently, the general public often relies on science journalism. Advocates explain that it would be better to understand research directly and not by proxy. *Second*, there are attempts to include the general public in the research process itself. This concept is labeled "citizen science" (Hand 2010) or "crowd science" (Franzoni and Sauermann 2014). This form of inclusion essentially engages a wide audience in the data gathering process. In this approach, the process can be divided into small chunks and worked on remotely from around the world. Famous examples of this understanding of Open Science are FoldIt and Galaxy Zoo.

3.5 Open Means Accessible to Anyone

This concluding understanding of Open Science is closely linked to the open access movement (Rufai et al. 2012). Advocates of this understanding explain that publicly funded research is often published in copyrighted journals. The very institutions that carry out research end up having to pay for the publications they or their colleagues wrote. Anyone not associated with a research institution is de-facto excluded from reading research findings as individual articles come with hefty price tags. To provide an example, one of the leading journals in management called Management Science charges US$30 to access a single article for 2 days. Even if interested in the subject of a research article, hardly anyone outside of academia will buy access to this original article. Furthermore, most research institutions do not have the means to subscribe to all relevant research outlets. Libraries must make budgetary decisions on where to allocate their funds. Given limited resources, it is common for many researchers to not have access to every journal that publishes relevant findings in their own fields.

There is no dominant school of thought in Open Science. All of the aforementioned understandings are frequently referenced in the literature—none of which could be considered particularly uncommon (Fecher and Friesike 2014). However, these different forms of Open Science (see Table 1 for a brief overview) demonstrate how multi-variant the field is and how many individual movements fall under the same umbrella term. This lack of uniformity further explains why the concept of Open Science can be so difficult to grasp.

Overall, one can argue that Open Science constitutes a cultural shift in how knowledge is created and disseminated. In the grand scheme of things, the Open Science movement has many aims including: increasing research quality, boosting collaboration, speeding up the research process, making the assessment of research more transparent, promoting public access to scientific results, as well as introducing more people to academic research. While all of these intentions sound noble and worthy to support, Open Science is only gaining traction very slowly. The fact that we discuss all of these ideas as a movement and do so a quarter century after the invention of the World Wide Web, provides us with an impression on how slow the development of this realization is.

Table 1 Overview of the different forms of Open Science

Form of Open Science	Main argument	Keywords	Readings
Open means transparent	Science, as practiced today, is not reproducible	• Open data • Open code • Open notebook science • Open peer review	McCullough (2009) Piwowar (2011) Fecher et al. (2014) Stodden (2009)
Open means collaborative	Science would be more efficient if more researchers would work more closely together and merge knowledge pools	• Polymath project Collaborative writing	Gowers and Nielsen (2009) Cranshaw and Kittur (2011) Nielsen (2012)
Open means a broader understanding of impact	The method we currently use to determine the impact of research is insufficient and does not take advantage of web tools	• Altmetrics	Priem et al. (2010) Weller and Puschmann (2011) Weller (2015)
Open means open to the public	Science can benefit from including the general public into its workflows	• Science communication • Crowdfunding • Citizen science • Crowd science	Cribb and Sari (2010) Franzoni and Sauermann (2014) Hand (2010)
Open means accessible to anyone	Publicly funded research results should be publicly available (online)	• Open access	Rufai et al. (2012)

4 Understanding Open Science as a Social Dilemma

As a whole, the considerations underlying the Open Science movement can be regarded as a social dilemma. A social dilemma is a "*situation in which individual rationality leads to collective irrationality*" (Kollock 1998, p. 183). It is a situation in which the collective best interest of a group is not in the best interest of the individuals in the group and therefore the individuals make defecting choices. Or in the context of science: What is in the best interest of the scientific system is not what incentivizes the individual researcher. For an individual researcher this means that he or she is incentivized to share knowledge (information, data, research results etc.) in a matter that best drivers his or her career further.

The ideas, which embody the Open Science movement, show little concern for the individual researcher. Instead, they address overarching problems within the academic knowledge creation system as a whole. Data sharing will be used as an example to illustrate this phenomenon. The expressed idea behind open research data is that the scientific system as a whole would improve, the replication of research results would be more easily avoided, and the application of old data in new contexts would be possible. Yet, the individual researcher receives much—if not most of their academic recognition—through published research. When applying for a new position for instance, the applicant must submit a publication list. Today, data is not considered a publication. Thus, publishing data does not add value to a publication list. While there are a few data journals, most data is considered to be supplementary, which may be of interest to some readers, but does not necessary need be included in most publications. Furthermore, data is messy. In order to make it understandable and reusable, data sets need to be cleaned, annotated and equipped with metadata. This entails a great deal of additional effort, which a scientist must invest before sharing. But this also means that the resources necessary to share data are taken away from elsewhere.

Moreover, there are also potentially harmful consequences that could arise from data sharing. In the case that someone uses data to prove a published finding wrong, the author would be forced to "retract" the article and remove it from their personal publication list in line with the paper-based publication system. Correcting the article after it has already gone through the pre-publication process cannot be reduced to a simple update. Holistically speaking, the process of someone verifying research results does in fact improve the academic system. In an Open Science system, findings that turn out to be wrong are no longer published. On an individual level, however, this requires extra effort to be invested, which in turn may generate a negative payoff. Understanding this micro perspective of Open Science is essential in understanding why ideas that would improve the overall research system fail to gain momentum on the individual level. The same holds true for other Open Science-related concepts as well. While most researchers would second the motion that open access is desirable, they also explain that publishing in traditional journals is more important to them individually. Publishing in highly ranked journals is important for receiving grants, scholarships, and positions in academia (Binswanger 2014; Neill 2008). Many fields do not have highly ranked open access journals (e.g., business and management). An individual researcher in these fields must consequently decide whether to publish in a highly ranked journal or to publish in a lesser ranked open access outlet.

In conclusion, Open Science requires a deviation from current best practices along with a change in thinking within the academic world. It is unreasonable to expect that large numbers of researchers will adopt the principles of Open Science in order to improve the system at the expense of their own career ambitions. Clear incentives are needed to overcome the current social dilemma of Open Science.

5 Incentives to Foster Open Science

Today, most of the incentives in science stem from a time when publications were paper-based. While the Internet and with it the World Wide Web has dramatically changed the way we collect information (e.g., online surveys) and analyze data (e.g., the programming language R) as well as the way we write (e.g., collaborative writing environments, real-time collaborative editing), performance indicators used to evaluate "good" research have not been updated accordingly. If we agree upon the notion that more openness in research has a positive impact on scientific progress, we need to ask what can be done to implement such a system. As is true for all complex problems or social dilemmas, there is no simple solution. However, there are already a few incentive structures in place, which promote Open Science and are paving the way for future developments. These incentives come in three major forms.

5.1 Institutional Incentives

These incentives come directly from the institution(s) a researcher is affiliated with and are not determined by disciplinary practices or general considerations in academic research. Each research institution shapes the working culture of its scientists. For instance, this can develop implicitly under an institute's leadership or explicitly due to a system of performance management (WU Wien 2014). Institutional blogs, data archives, archiving services, open access appointees, and funding options for open access publishing are some examples of institutional incentives. Furthermore, advisors play a significant role in shaping the understanding and orientation of young researchers. In an organizational environment where Open Science is understood as something *others* should start with, new practices will only take up a slow adoption rate. The same holds true for organizations that solely pay lip service to the idea of Open Science, but shy away from implementing it into their research processes. It is important to emphasize the weight of institutional incentives. Luckily, most of us work under conditions, where research institutions are granted a large degree of freedom when it comes to their scientific practices. Yet, with this freedom comes the responsibility to shape the system in a way that is most beneficial for the science as a whole.

5.2 Disciplinary Incentives

Disciplinary incentives address an entire group of researchers who work in the same domain, regardless of the institution they are associated with. The opportunity that lies within disciplinary incentives is that they offer a more level playing field for all researchers in one field itself. If an incentive applies to all researchers in one field implementing Open Science practices, it does not come with individual penalties. As discussed before, Open Science in many cases does not catch on as individual researchers feel that implementing Open Science would be to their (temporal) disadvantage. What is in the best interest of the research community is not necessarily in the best interest of the individual researcher. Disciplinary incentives bypass this issue as they apply to all researchers that work in one field. Examples of such incentives are disciplinary databases like GenBank—a sequence database of all publicly available nucleotide sequences (Benson et al. 2007). Participants of the entire research discipline have agreed upon the fact that it is in everybody's best interest to publish gene sequences. Thus, the community compels its researchers to do so. Further examples are academic journals with data sharing policies. Such journal policies mandate scientists to publish original research data alongside their findings. Examples of journals with such policies are Atmospheric Chemistry and Physics, F1000Research, Nature, and PLoS ONE.

5.3 Systemic Incentives

Incentives considered to be systemic are cross-disciplinary and affect researchers regardless of their field or institution. These incentives reflect the research culture as a whole. Good examples for such incentives are the requirements that come with project funding. In the U.S. for example, the National Institute of Health urges investigators seeking US$500,000 or more (per year) to develop a data sharing plan: "*Timely release and sharing of final data by NIH-supported studies is expected.*" (NIH 2003, p. 25). In the European Union, under the research funding program Horizon 2020, open access publications are considered a "general principle". "*As of 2014, all articles produced with funding from Horizon 2020 will have to be accessible: Articles will either immediately be made accessible online by the publisher [. . .], or researchers will make their articles available through an open access repository no later than six months (12 months for articles in the fields of social sciences and humanities) after publication [. . .]*" (European Commission 2012). The European Union has taken a stride in promoting academic data sharing by way the Horizon 2020 program's systemic incentive structure.

These three forms of incentives reach an increasing number of researchers. It is evident that an institutional incentive affects less people than a disciplinary incentive; and a disciplinary incentive, in turn, affects less people than a systemic incentive (see Fig. 1). With an increase in outreach, the potential damage to the scientific system as a whole increases as well. What does that mean? Open Science is—to a large extent—a deviation from well-known practices. With policy changes that address this deviation always comes the risk of side effects. Side effects could be the loss of power from established players, regulations that are not in line with privacy or copyright laws, or an increase in goal uncertainty for young researchers. To avoid such ripple effects from occurring, systemic incentives should be implemented on a step-by-step basis rather than in one giant leap. While disciplinary incentives such as journal policies can demand more from researchers, there are many different journals researchers can choose from. In contrast, systemic incentives have the ability to change the entire research landscape. With this comes a certain degree of responsibility, reflected in the cautious nature of policy-making and implementation thus far. The open access policy in the Horizon 2020 program for instance, was preceded by a far more liberal policy in the FP7 program

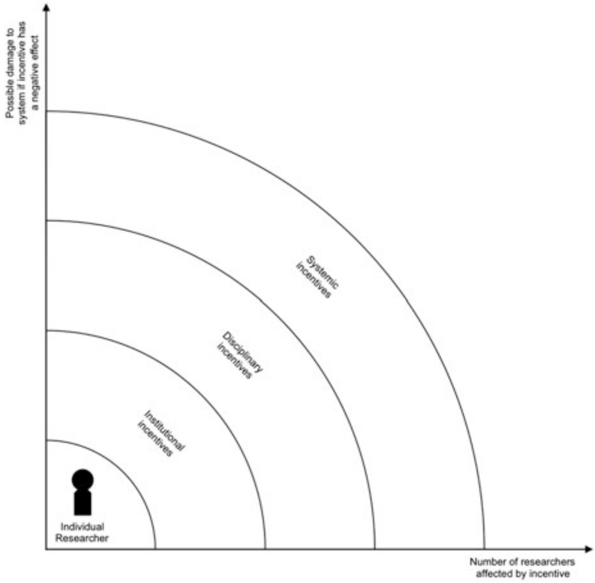


Fig. 1 Three forms of incentives to promote Open Science

(European Commission 2008). While the Horizon 2020 program mandates *all* research results to be published open access, the FP7 program only required 20 % of the research funded to do so. One could argue to test the waters. The current Horizon 2020 program takes this experimental approach for open research data.

6 Future Directions

As seen in the previous section, there are already a few incentives in place seeking to foster more engagement with Open Science. They are, however, exceptions—and not the rule—in academic research. The Open Science movement comes with the promise of a better, faster, and more just scientific system.

This article argues that the current debate on Open Science covers a multitude of understandings. And these varying perceptions are one reason for the slow realization of Open Science. Academic research as a whole has no agenda, and no priorities on what topics to tackle first. The academic system has no centralized governing body with an agenda-setting function. Therefore, the future of Open Science will be determined by the three levels of incentives (institutional, disciplinary, and systemic) discussed above.

It is also important to understand that science—as an established system—is imprinted by customs and culture of its heritage. Although changes can bring potential benefits for the entire system, these cultural traces must be addressed. To provide an example of such an imprint: In many research fields, quotations are referenced with a source as well as a page number. Once a publication becomes digitally retrievable online, a page number is no longer necessary. Today, a reader could simply search for quoted text and would probably do so regardless of a supplied page number. However, methods stemming from a time when publications were printed on paper—and not easily searchable by a personal computer—have remained common practice. To address this imprinted practice, online journals like PLoS ONE insert page numbers on their online publications. What may look like an anachronism to some helps others provide page numbers in reference system that has yet to catch up with today's research reality.

Open Science is an attempt to improve a vast, international and multidisciplinary research system. Given the complexity of the research apparatus, the improvements Open Science can harness will take far longer to actualize than many are willing to accept. Yet, incentivizing individual researchers to rethink their own actions issue by issue remains the only viable solution to the social dilemma Open Science is.

References

Bartling S, Friesike S (2014) Towards another scientific revolution. In: Bartling S, Friesike S (eds) Opening science: the evolving guide on how the internet is changing research, collaboration and scholarly publishing. Springer, Cham, pp 3–15

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J (2007) GenBank. Nucleic Acids Res 36(suppl 1):D25–D30

Berners-Lee T (1989) Information management: a proposal. http://www.w3.org/History/1989/proposal.html. Accessed 20 June 2014

Binswanger M (2014) Excellence by nonsense: the competition for publications. In: Bartling S, Friesike S (eds) Opening science: the evolving guide on how the internet is changing research, collaboration and scholarly publishing. Springer, Cham, pp 49–72

Björk B-C, Solomon D (2013) The publishing delay in scholarly peer-reviewed journals. J Inform 7(4):914–923

Boulton G, Rawlins M, Vallance P, Walport M (2011) Science as a public enterprise: the case for open data. Lancet 377(9778):1633–1635

Cranshaw J, Kittur A (2011) The polymath project: lessons from a successful online collaboration in mathematics. In: CHI'11 proceedings of the SIGCHI conference on human factors in computing systems, Vancouver, pp 1865–1874

Cribb J, Sari T (2010) Open science: sharing knowledge in the global century. Csiro, Collingwood

European Commission (2008) Open access in FP7. http://ec.europa.eu/research/science-society/index.cfm?fuseaction=public.topic&id=1300. Accessed 20 June 2014

European Commission (2012) Scientific data: open access to research results will boost Europe's innovation capacity. http://europa.eu/rapid/press-release_IP-12-790_en.htm. Accessed 20 June 2014

Fecher B, Friesike S (2014) Open Science: one term, five schools of thought. In: Bartling S, Friesike S (eds) Opening science: the evolving guide on how the internet is changing research, collaboration and scholarly publishing. Springer, Heidelberg, pp 17–47

Fecher B, Friesike S, Hebing M (2014) What drives academic data sharing? SOEPpapers on Multidisciplinary Panel Data Research, No. 655, Berlin

Franzoni C, Sauermann H (2014) Crowd science: the organization of scientific research in open collaborative projects. Res Policy 43(1):1–20

Gowers T, Nielsen M (2009) Massively collaborative mathematics. Nature 461(7266):879–881

Hand E (2010) Citizen science: people power. Nature 466(7307):685–687

Kollock P (1998) Social dilemmas. The anatomy of cooperation. Annu Rev Sociol 24(1):183–214
McCullough BD (2009) Open access economics journals and the market for reproducible economic research. Econ Anal Poli 39:118–126
Meyer ET, Schroeder R (2013) Digital transformation of scholarship and knowledge. In: Dutton WH (ed) The Oxford handbook of internet studies. Oxford University Press, Oxford, pp 307–327
Nature (2006) Overview: nature's peer review trial. http://www.nature.com/nature/peerreview/debate/nature05535.html. Accessed 20 June 2014
Neill US (2008) Publish or perish, but at what cost? J Clin Investig 118(7):2368
Nielsen M (2011) Definitions of open science? Okfn mailing list. https://lists.okfn.org/pipermail/open-science/2011-July/000907.html. 28 July 2011
Nielsen MA (2012) Reinventing discovery: the new era of networked science. Princeton University Press, Princeton
NIH (2003) NIH posts research data sharing policy. Anthropol News 44:25
Piwowar HA (2011) Who shares? Who doesn't? Factors associated with openly archiving raw research data. PLoS ONE 6:e18657
Polymath DHJ (2010) Desity Hales-Jewett and Moser numbers. In: Bárány I, Solymosi J, Sági G (eds) An irregular mind, vol 21, Bolyai society mathematical studies. Springer, Cham
Polymath DHJ (2012) A new proof of the density Hales-Jewett theorem. Ann Math 175 (3):1283–1327
Priem J, Taraborelli D, Groth P, Neylon C (2010) Altmetrics: a manifesto. http://altmetrics.org/manifesto/. Accessed 20 June 2014
Rufai R, Gul S, Shah TA (2012) Open access journals in library and information science: the story so far. Trends Inform Manag 7(2)
Stodden V (2009) The legal framework for reproducible scientific research: licensing and copyright. Comput Sci Eng 11(1):35–40
Weller K (2015) Social media and altmetrics: an overview of current alternative approaches to measuring scholarly impact. In: Welpe IM, Wollersheim J, Ringelhan S, Osterloh M (eds) Incentives and performance – governance of research organizations. Springer, Cham
Weller K, Puschmann C (2011) Twitter for scientific communication: how can citations/references be identified and measured? In: Proceedings of the ACM WebSci'11. ACM, Koblenz, pp 1–4
WU Wien (2014) Prämien für Top-Journal-Artikel. http://www.wu.ac.at/academicstaff/awards/publication. Accessed 20 June 2014
Woelfle M, Olliaro P, Todd MH (2011) Open science is a research accelerator. Nat Chem 3:745–748

Success Measurement of Scientific Communication: The Contribution of New Media to the Governance of Universities

Christian Pieter Hoffmann

Abstract Success measurement is a critical component of any organization's governance. It fulfills functions as important as the monitoring, coordination and evaluation of organizational activities. We argue that knowledge-intensive organizations are also communication-intensive organizations: communication is essential for the creation and sharing of knowledge. Universities, as knowledge-intensive organizations, focus significant efforts and resources on scientific communication. We conclude that success measurement of universities can largely be understood as the success measurement of scientific communication. This chapter will argue that new media provide new opportunities for the success measurement of scientific communication. Online media render activities as diverse as citations, bookmarks, views or downloads accessible to analysis—and thereby facilitate a more varied evaluation of effects caused within the scientific community ("impact"). Social media, especially, provide insights into conversations and personal networks. By observing and analyzing new media, universities can generate a richer, more differentiated understanding of their communication success. Thereby, new media have the potential to contribute to the governance of universities.

1 Introduction

Universities tend to be heterogeneous organizations, comprised of departments, institutes, centers, chairs. Individual university units commonly enjoy a significant degree of autonomy which, in turn, decreases the coherence and impedes the coordination of the overall organization (Coleman 1990; Giddens 2004). At the same time, universities—public or private—are increasingly exposed to competitive forces. International competition for scholars, students and funding is forcing universities to rethink the necessary level of organizational coherence and coordination. In reaction to these changing requirements, the governance of universities is

C.P. Hoffmann (✉)
Institute for Media and Communications Management, University of St. Gallen, St. Gallen, Switzerland
e-mail: christian.hoffmann@unisg.ch

I.M. Welpe et al. (eds.), *Incentives and Performance*,
DOI 10.1007/978-3-319-09785-5_18

notably beginning to adopt insights and practices from the strategic management of corporations.

Today, university governance—just like the strategic management of corporations—is faced with the challenge of aligning the activities of its members with overall organizational goals and ambitions (March and Simon 1959). This includes translating organizational goals into individual objectives and providing adequate incentives (Coase 1937; Demsetz 1988). Yet, the alignment of organizational and individual goals, the coordination and co-orientation of a complex organization require not only a sufficiently clear definition of its goals but also an evaluation of its goal attainment.

This chapter will focus on the success measurement of universities' output, more specifically scientific impact assessment (Bollen et al. 2009). After a brief introduction to the importance of success measurement to organizational coordination and coherence, we suggest that universities, as knowledge-intensive organizations, are largely defined by and primarily produce various forms of knowledge which are disseminated based on communication (cf. Blumer 1969). We therefore argue that impact assessment of universities has to be informed by success measurement in (organizational) communication. Next, we will provide a brief overview of the approaches and opportunities available to success measurement in communication as well as current approaches applied to the success measurement of scientific communication. Finally, we will point out how new media alter the opportunities of success measurement in communication and in scientific communication, in particular. We suggest that new media allow for a richer analysis of communication effects, and thereby a more comprehensive evaluation of the success of universities as well as their individual members.

2 The Importance of Success Measurement

Evaluating the impact or success of its activities is a core function of any organization's governance. Organizations are purpose-driven social institutions, they are largely defined by their goals and ambitions—common goals give meaning and substance to an organization, they are a precondition for the coordinated action of its members (Coleman 1990; Cyert and March 1963; Blumer 1969). Many theories have been proposed on how organizations come up with and define shared goals, how strategies are derived. For the purpose of this contribution, we will ignore these (undoubtedly crucial) matters, though, and focus on the opposing end of the governance process: the evaluation of the successful attainment of previously defined goal(s).

Given the importance of shared goals to the identity, the specific function, even the mere existence of an organization (Weick 1995), it is relatively easy to understand why success measurement plays a critical role in the governance of organizations. For without success measurement, an organization can be likened to a sprinter who knows where she wants to run, but doesn't ever know if she has actually arrived—or where she stands relative to other sprinters. Planning, the process of defining goals and choosing adequate measures, therefore is never complete without an evaluation of the achieved success (Carroll and Buchholtz 2006; Freeman 1984).

In effect, success measurement fulfills a number of functions in the governance of organizations—not all of which are always consciously pursued (Coase 1937; Demsetz 1988). All of these functions complement each other and can become more or less salient depending on the current state and activities of the organization:

Information: Success measurement constitutes an analysis that provides important informational input to organizational decision makers.

Monitoring: A key focus of success measurement is on the organizational environment, including the organization's stakeholders. It allows for a monitoring of changes in the organizational environment. In other words, success measurement contributes to the boundary spanning efforts of organizations.

Coordination: Success measurement focusses the attention of organization members on specific goals. It thereby facilitates the coordination of the organization and increases the coherence of organizational behavior.

Surveillance: Through success measurement, the organization's leadership controls the activities of the organizational members in order to ensure that they actually contribute to the common goals. Dangers to the cohesion of the organization are identified and can be addressed.

Steering: Success measurement discloses the current state of the organization's goal realization and thereby allows for timely adjustments and necessary corrections in organizational behavior.

Evaluation: Of course, success measurement is the basis for the evaluation and rewarding of the organizational members' or units' contribution.

In short, there are many reasons for any organization—knowledge-intensive or not—to engage in success measurement and the evaluation of its impact. Yet, knowledge-intensive organizations, like universities, feature some particularities that impact the form their success measurement tends to take.

3 Measuring the Success of Universities

In most cases, knowledge-intensive organizations are also communication-intensive organizations: the creation and dissemination of knowledge is based on communication (Blumer 1991; Rose 1962). Through communication, organizations collect information as raw material for the creation of knowledge. Organizational members communicate in order to share information and coordinate their activities and collaborate in the production of knowledge. The result of that collaboration, in turn, is then communicated to internal and external stakeholders.

Sociologists and communication scholars have pointed out that communication largely constitutes organizations—any kind of organization. Without communication, individuals wouldn't be able to achieve a sufficient level of coordination to be considered an organization (Katz and Kahn 1966; March and Simon 1959; Mead 1965). This observation is especially true for knowledge-intensive organizations as these organizations focus on immaterial objects (i.e., information and knowledge) rather than physical products. In other words, communication is not only a substantive characteristic, but also a key output of these organizations.

Applied to research institutions, we posit: The scientific endeavor is driven by communication. Scientific impact, therefore, is driven by successful communication. A scientist with a strong impact is one who is being heard, whose contribution is taken note of, respected, built upon. In order for these effects to be achieved, the scientist's output, his/her communication, needs to be clear and understandable, interesting and insightful, useful and inspiring. The better his/her communication meets these standards, the more likely his/her output is to be given attention, to be favorably received and employed in other researchers' work.

A key conclusion to take away from these propositions is: The evaluation of success or impact in academia is largely an exercise in success measurement of (scientific) communication. So what do we know about success measurement in communication?

3.1 Success Measurement in Communication

In short, communication success has been achieved whenever communication effects are in line with the predefined communication goals. Frequently, three forms of communication effects are differentiated: cognitive, affective and conative effects (Fishbein and Ajzen 1975). Grunig (1993) points out that there is another, a fourth effect, which can be considered a precursor to cognitive, affective and conative effects: a perceptual effect. In other words, target audiences must be reached, some level of attention must be attained, in order for any further effects to materialize. Therefore, perceptual effects constitute something like the foundation for higher order communication effects.

Cognitive and affective effects constitute a second level of effects. On this level, we find the opinions, attitudes, beliefs, or convictions of the target audiences (Fiske and Taylor 1991; Rosenberg and Hovland 1960). Only when cognitive and/or affective effects have been achieved, can any effects on the behavior of target audiences be expected. Of course, these behavioral effects are usually most valuable to the communicator—only they have tangible or economic benefits. Yet, they are also the most difficult to achieve with any level of certainty.

Research has uncovered a myriad of variables that impact the effects of communication. Among these variables are features of the source or sender (e.g., a scholar), the medium (e.g., an academic journal), the message (e.g., a scientific finding), the recipient (e.g., colleagues within the scientific community), and the social setting (e.g., competition between various research teams) (Hovland

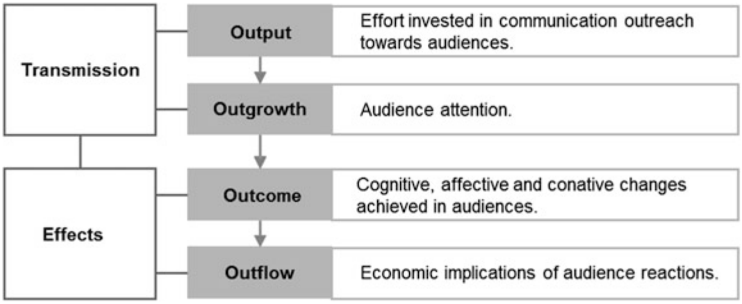


Fig. 1 Levels of success measurement in communication

et al. 1953; Katz and Lazarsfeld 1955; Petty and Cacioppo 1986). While we cannot go into much detail on the vast variety of findings in this field, we can begin to grasp the complexity of any act of communication and the variety of factors that affect whether a communicator is actually able to attain his/her communication goal(s). Again, higher order communication effects are increasingly difficult to realize: In many instances, it is quite difficult to even gain the attention of a target audience, let alone affect their knowledge or attitudes in a specific way—which in turn would be the precondition for specific behavioral effects.

Given the complexity of communication effects, success measurement in communication is frequently differentiated by the level of effect achieved. In organizational communication or public relations research, a framework describing four levels of success measurement has found widespread acceptance (Lindenmann 2003—see Fig. 1). The first two levels of the framework concentrate on the successful transmission of communication. On the topmost, in effect the most basic level, organizations measure the "output" of their communication: the instruments they have chosen to apply in order to reach their target audiences. Typical measures on this level would include the number of events organized, the number of press releases published, etc. While this "output" is most easy to measure, assessing the communication output doesn't really provide any information on whether and how target audiences have been affected.

The second level of the framework focusses on the "outgrowth" of communication, meaning the attention gained among target audiences. Typical measures include the number of attendants at an event or number of downloads of an online document. At this level, success measurement addresses the question of whether any of the output produced by the communicator actually has been seen/heard/considered by a recipient. If yes, success has been achieved on the outgrowth-level.

While the first two levels of the framework focus on the successful "transmission" of communication, the following two levels focus on higher-order communication effects: If the communicator is interested in whether the communication perceived by the target audiences has, in fact, left any mark on their knowledge, understanding, attitude or even behavior, it is necessary to conduct success measurement on the "outcome" level. The final level, the "outflow" level, addresses the question of whether the communication outcome has generated any (economic) value for the communicator.

As we have seen, communication success becomes more and more difficult to attain the higher the level of effect under consideration—it also becomes more difficult to measure. We will illustrate this challenge by applying the framework to the specific domain of scientific communication.

3.2 Success Measurement in Scientific Communication

Today, what is the most common and established measure of scientific impact? Despite widespread criticism, it is most probably still the number of articles published in top-tier journals, with the quality of the outlet being assessed based on the journal's impact factor (Bollen et al. 2009). Considered in terms of communication effects, this measure actually assesses the success of communication on the lowest possible level, the output level. The number of published articles provides information on the effort invested in communication, but it really doesn't provide any information on the effects of this output.

This finding needs to be qualified insofar as the impact factor does in fact contribute to the validity and sophistication of success measurement: Given that the impact factor indicates the average number of citations of articles published in a specific outlet, it also indicates that, on average, articles published in a high-impact journal have enjoyed a higher level of attention and consideration than those published in low-impact journals (Garfield 2003). Complementing the number of published articles by the impact factor of the respective outlet therefore allows for inferences regarding the success achieved on the outgrowth level. Publishing in high-impact journals makes it more likely that a sizeable audience has been reached. Of course, based on the journal impact factor alone, we cannot say for certain if this was in fact the case for a specific article or not.

It may be somewhat sobering to realize how basic success measurement in scientific communication has been, and largely still is. This finding may also explain why there is so much skepticism directed at success measurement in academia—ranging from criticism aimed at particular measures to an overall resistance against the "quantification" and ranking of scientific achievements (Adler et al. 2012; Seglen 1997). At the same time, we need to acknowledge the difficulty of valid and reliable success measurement in communication—in general as well as specifically in the context of scientific communication. As any social scientist can attest, it's not easy to find out about the opinions, attitudes, beliefs, or convictions held by individuals. Really, success measurement in communication is limited by the variety of methods available to the social sciences (but often hobbled by a lack of resources as the application of some of these methods can be quite costly—more costly than would seem warranted in the context of communication success measurement).

Running the risk of oversimplifying a complex matter, we could say that in success measurement of communication, we find three instruments at our disposal: surveys, observations and content analyses. Obviously, there are many different forms of surveys (qualitative as well as quantitative) and observations (directed at more or less controlled environments and featuring various levels of involvement on the part of the observer). As a rule of thumb, though, surveys and observations tend to be invasive and resource-intensive and therefore only sparingly applied to success measurement in communication.

Any academic reading this contribution may ask him- or herself: How frequently would I be willing to participate in a survey analyzing my reception of my colleagues' work? Would I want someone to observe me while scouting for input into my next research endeavor? To date, common—and understandable—apprehension on the part of the target audiences in combination with a lack of resources leads to strong preference for the application of content analyses in communication success measurement.

In public relations, professionals analyze press articles in order to gauge the public's reception of their communication efforts. In financial communication, analyst reports are analyzed in order to anticipate the reactions of investors. In scientific communication, we find a significant advantage in the fact that the target audiences tend to produce artefacts that lend themselves to content analyses, most specifically publications such as journal articles, research reports or books. By analyzing these artefacts, it is possible to estimate the success of scientific communication both on the outgrowth and the outcome level:

As we have already noted, content analyses focused on citations provide insights as to attention gained among the communicator's peers (outgrowth). Aside from attention and consideration (i.e., cognitive effects), citations also signal a level of respect—even in the case of a critical reception—(i.e., affective effects), and they indicate activities on the part of the recipient, such as reading, analysis, or promotion (i.e., conative effects). All of these effects, uncovered by mere content analyses, indicate esteem within the scientific community, which may translate into collaboration, promotion, funding, etc.

The relative richness of the information provided by citation-focused content analyses can be seen as the reason that output-level metrics (in some cases enriched by the insights provided by journal impact factors) have increasingly been complemented by citation counts and the calculation of metrics such as the h-index (Hirsch 2005). The h-index directs our attention away from the overall esteem for an outlet (i.e., a journal), and focuses it on the reception of individual contributions—aggregated by individual scientific communicator. The average number of citations attracted by the publications of a researcher is a better indicator of his or her esteem within the scientific community than the mere number of publications (even if weighted by journal impact factor) because it takes the attention, evaluations and activities of the target audiences caused by specific publications (as opposed to overall outlets) into consideration.

Of course, the emergence of the h-index as somewhat of a new gold standard in individual-level metrics of scientific impact also indicates a very important insight:

	Survey	Observation	Content Analysis
Analog Media	–	–	+
Web 1.0	+	++	++
Web 2.0	+	+++	+++

Fig. 2 New media effects on the applicability of success measurement methods

The options available to success measurement in (scientific) communication are dependent on the media employed (see Fig. 2).

4 The Contribution of New Media to Success Measurement

In the pre-digital era, counting citations by peers, while possible, was still quite cumbersome and too resource intensive to be employed on a broad scale. Online media have significantly altered that situation. The digitization of scientific communication and the resulting artefacts were the precondition for the establishment of searchable databases. These databases, in turn, make it relatively easy to identify and count citations—and to compute metrics such as the h-index.

In fact, online media introduce an entire new range of metrics to the success measurement of scientific communication, called "webometrics". Webometrics include metrics such as page views, bookmarks, hyperlinks or downloads—indicating various levels of engagement with a scholarly publication. They can be applied on the level of a journal (Thelwall 2012; Vaughan and Hysen 2002), a single article (Kousha et al. 2010; Vaughan and Shaw 2003) or an individual researcher (Aguinis et al. 2012).

Similar to citations, but more varied in nature—and, in some cases, more subtle—webometrics also indicate the attention and consideration given to individual contributions. Views indicate overall interest. Bookmarks are an even more valid measure of interest, as it constitutes an activity indicating future use. A bookmark can also be interpreted as an indicator of interest and appreciation, since the publication has been deemed worthy of further attention. Similarly downloads, which indicate that the publication has been judged of sufficient quality to warrant further analysis. All of these metrics can be considered leading indicators to eventual citations. Hyperlinks are strong indicators of esteem, as they direct others' attention to the publication.

In short, we can already state that the establishment of online media has significantly increased the scope and sophistication of the evaluation of scientific impact. It has expanded communication success measurement beyond the output-level to include outgrowth- and even outcome-measures. While all of these measures are still derived from some form of observation or content analysis, the widespread adoption of online media in academia has rendered a larger number of *activities* accessible to success measurement efforts—and thereby levels of engagement, attention, interest and appreciation.

We will argue that the adoption of social media in scientific communication will further expand the scope of scientific impact assessment and add even more depth and richness to success measurement in scientific communication.

4.1 Social Media in Scientific Communication

Why should social media be able to contribute to the evaluation of scientific impact, and thereby to the governance of academic institutions? In order to better understand this contribution, it is necessary to understand the specific nature of social media. In a conventional definition, social networking sites have been defined as: "web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system" (Boyd and Ellison 2007).

As opposed to earlier online applications, like message boards or chat rooms (Bechar-Israeli 1995), social media tend to be "nonymous" in nature. Their purpose is to allow self-expression (Zhao et al. 2008), interpersonal connections (Puckett and Hargittai 2012) and the generation of social capital based on these connections (Ellison et al. 2007; Kane et al. 2014). Therefore, the benefits of social media can best be reaped if users maintain an identifiable—as opposed to an anonymous or pseudonymous—online-presence (Kane et al. 2014).

Above all, social media make it easier than ever before for users to publish content online—be it text, photos, videos, or other data. Personal profiles are the basis for these publication activities; they are also the nucleus of the articulation and management of personal networks. Because the publication of content is as easy as a mouse click, users tend to publish more content than they would have without these new media affordances. In economic terms, social media lower the transaction costs of online publication, thereby increasing the number of such publications. The more cumbersome and costly an online publication, the more users have to critically evaluate its necessity.

By lowering the transaction costs of online publications, social media don't necessarily increase the quality of each individual publication. Quite on the contrary, social media allow their users to publish content that would have previously been deemed unworthy of publication. On the other hand, the apparent banality of some of the data published in social media also increases its richness. Social media make available data on who does what when where and with whom. Observations, opinions, whims, ideas—anything can quickly be shared with a digital audience. Analyzing these tremendous amounts of data ("big data") may require sophisticated algorithms and sizeable IT capacity, but it *is* possible to access and analyze more data than ever before.

In other words: Due to social media, more communication by and within the scientific community than ever before has become accessible to observation and content analyses, and thereby to success measurement. Some social media are expressly geared towards scientific communication, such as academic social networking sites like Academia.edu and ResearchGate. These services include functions such as uploading and sharing articles, endorsing colleagues, or finding literature (Jeng et al. 2014; Thelwall and Kousha 2014—see, for example, Fig. 3). At the same time, mainstream applications have also found avid use within the scientific community. For example, twitter has become a popular tool for purposes as diverse as networking, information gathering, knowledge dissemination, and conference chatter (Mahrt et al. 2013; Nentwich and König 2012).

Gruzd and Goertzen (2013) conducted a survey among the members of three professional social science organizations. They found that, currently, non-academic SNS and blogs are the most popular applications for frequent use, followed by

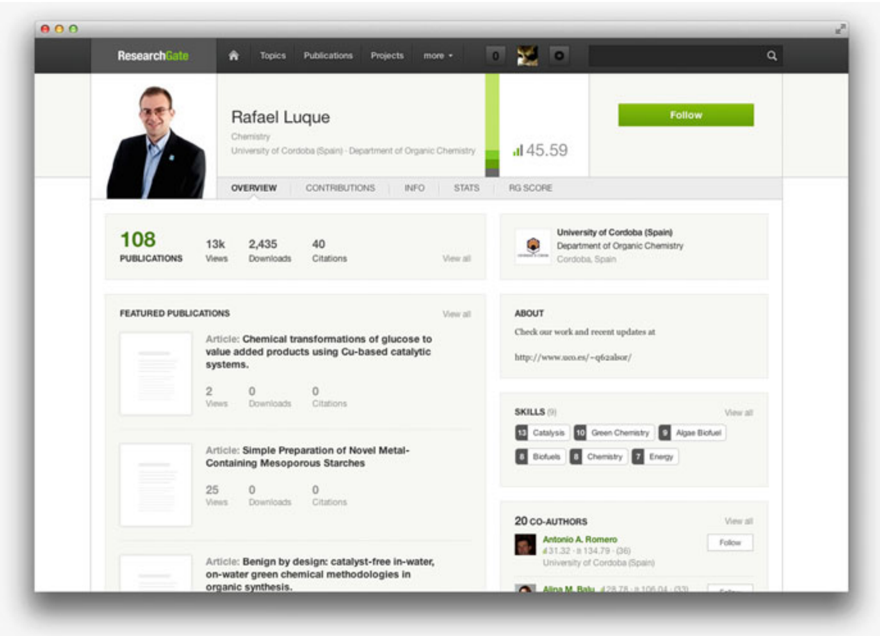


Fig. 3 New media in scientific communication (here: ResearchGate)

online document management services. As to use intentions, academics report to be most prone to adopt presentation sharing sites, bibliographic management sites (e.g., Mendeley), and academic social networking sites (e.g., Academia.edu or ResearchGate).

A recent large-scale survey conducted among German academics of various disciplines revealed that content platforms (such as Flickr or Youtube) are the most frequently used, followed by general social networking sites (such as Facebook), academic SNS and microblogging services (Pscheida et al. 2014). Procter et al. (2010) found that social media adoption among scientists is influenced by demographic characteristics, such as age, gender, but also by position, and discipline. Gruzd and Goertzen (2013) found that social media are primarily used for informational purposes, such as staying up-to-date on colleagues' activities and discovering new ideas or publications.

In summary, we find evidence that social media are in fact rapidly being adopted in scientific communication. These media are applied to uses as diverse as networking, dissemination and promotion of content, information, knowledge management, discussion, and simple small talk. Thereby, we do find the expected increase in the amount and richness of published data as facilitated by social media's low transaction costs of content publication. These observations lead us to the question of what the contribution of social media to scientific impact assessment will look like. What kinds of metrics can social media add to the success measurement of scientific communication?

4.2 New Metrics of Scientific Impact

The information derived from the analysis of social media can be summed up in two terms: "conversation" and "community". Conversations in social media tend to be asynchronous, in written form and to a large degree accessible to the public, thereby ideally lending themselves to content analysis. These are important distinctions between social media conversations and, for example, telephone or e-mail conversations. Of course, the traditional academic conversation, as documented in peer-reviewed journals, is also asynchronous, in written form and accessible to the public. That is why impact assessment has traditionally relied on the content analyses of these publications. Yet this traditional academic conversation is also very slow and heavily filtered—as opposed to social media conversations, which are lively, instantaneous, diverse, more personal, informal and emotional in nature.

Since conversations tend to evolve within specific subsets of network members, another important contribution of social media to success measurement revolves around communities. Members of a community more frequently converse with each other as opposed to non-members. Joining a conversation frequently implies joining a community, too. As "nonymous" media, social media render personal relationships, and thereby the membership and structure of communities, explicit and observable. Through "friendships" or follower-relationships, through the membership in groups or by becoming a fan or follower of a site, it is becoming increasingly easy to identify and observe communities. Structural analyses of social relationships, such as social network analyses, are well established in social sciences (Freeman 2004). Yet, social media provide a new foundation for these analyses and significantly increase the richness of available data (Priem and Hemminger 2010; Priem 2013).

It has long been established that conversation and collaboration within communities is a driving force in scientific impact. The term "invisible colleges" has been used to describe closed networks of researchers employed to refine and promote ideas and publications (Crane 1972). Within invisible colleges, core members—or influentials—collect social capital, also called "academic capital", which is later translated into more tangible assets such as research grants or promotions (Bourdieu 1990).

Since social media allow for unprecedented access to data on community structures, roles within communities (i.e., centrality, prominence, influence), and conversations within communities, they facilitate the application of observation and content analyses in scientific impact assessment. In social media, it is possible to observe and analyze the reaction of peers and target audiences to the communication of individual researchers as well as research institutions. New metrics such as "buzz", "sentiment" or "influence" derived from social media analytics in effect indicate the attention given to a communicator, the cognitive and affective reaction of audiences, the respect and esteem awarded to the communicator. They also indicate conative reactions, for example, in the form of likes, replies, comments, friendship requests, etc. (Thelwall and Kousha 2014).

The access to information on communities and conversations provided by social media goes beyond what webometrics already contribute to the success measurement of scientific communication. Thanks to social media, universities will be able to learn more about the effect of their communication efforts within and beyond the scientific community than ever before. Thanks to new media, success measurement in academia will finally move beyond mere output- or outgrowth-measures and actually consider the outcome achieved through successful communication. Thereby, social media will significantly contribute to the reliability and validity of success measurement as a key element of university governance.

5 A Note on Public Universities

Without going into too much detail, we should note that the new opportunities provided by social media to the governance of universities should be especially attractive to public universities. These institutions are increasingly faced with competitive pressure, due to globalizing markets for students, scholars and funding, but also due to private competition and restrictions in public funding. As von Hayek (1945) pointed out, public institutions suffer from severe disadvantages when faced with competitive pressure due to their disconnection from market information. Since public institutions cannot benefit from the information provided by freely fluctuating prices (driven by supply and demand) about the interests, preferences and intentions of their interaction partners, they are more apt to misallocate resources than private institutions (von Mises 1996).

Public institutions should, therefore, be especially motivated to access information on the judgments, attitudes, tastes and reactions of their target audiences. They should strive to compensate for informational disadvantages unavoidably associated with their legal status. Again, the observation and analysis of communities and conversations in social media may provide more insights into the cognitive, affective and conative reactions of target audience than ever before available to public institutions. This is not only true for the scientific community, but also further stakeholders such as students, the general public, local communities, political interest groups, etc.

Of course, social media analytics cannot fully compensate for the lack of tangible information as provided by the market process (i.e., feedback on the outflow-level). Yet there is reason to be optimistic about the governance of public institutions when considering the affordances of new media.

Conclusion

It should be noted that the perspective presented in this chapter has adopted a distinct functional perspective, presupposing the necessity and benefits of success measurement in knowledge-intensive organization. Of course, such a perspective may be criticized due to its normative implications. There may be reasons to question the necessity of success measurement in knowledge-intensive organization, or universities in particular—such objections would go beyond the scope of a contribution focusing on the role of new media in success measurement, though.

As to the contributions of new media, these contributions are by no means a comprehensive or ideal solution to the challenges associated with success measurement in (scientific) communication. Although more diverse and fluid in nature, more fine-grained and extensive, data published in online media still present only a filtered and selective view into the dynamics of scientific work, the personal interactions and relationships necessary for successful research (Crane 1972). Many conversations in social media are not accessible to success measurement due to privacy restrictions. Therefore, online publications considered in webometrics analyses still tend to be a relatively slow, filtered representation of the dynamics involved in generating captivating research (PLoS Editors 2006). Finally, since many online metrics are based on relatively simple user activities, they are also susceptible to deception or gaming efforts—on the other hand, there is not much reason to expect that they would be more so than traditional metrics of scientific impact.

This chapter has focused on one specific element of university governance: the task of success measurement, which is a necessary precondition for any

(continued)

kind of strategic management of and within organizations. We have pointed out that knowledge-intensive organizations, such as universities, are primarily engaged in communication activities. Success measurement, such as scientific impact assessment, is therefore largely comprised of the assessment of successful (scientific) communication.

We have differentiated various levels of communication success measurement and have pointed out the methods available to the assessment of communication success. As we have seen, the more refined the intended level of communication success, the more difficult its attainment, and the more complex its evaluation. For that reason, success measurement in academia has tended to focus and rely on mere output metrics—a questionable foundation for the governance of universities. New media, though, increasingly allow for more diverse and sophisticated metrics of communication success—and scientific impact.

While digital media have already revolutionized scientific impact assessment by rendering large-scale citation counts and the calculation of metrics such as the h-index feasible, webometrics further contribute to success measurement in scientific communication. Analyzing page views, bookmarks, downloads and hyperlinks add richness to our understanding of audience reactions and judgments.

Social media will further complement these significant advances in success measurement by allowing access to the conversations and structures of various communities. In future, whenever a university pursues goals like gaining access to regional, staff or funding markets, establishing a certain position in an academic field, attracting the attention and interest of specific academics (or students, corporations, political supporters, etc.), connecting to important stakeholders, increasing its standing within the scientific community, promoting the recognition of its members or its research, social media analytics will be able to inform the ensuing evaluation of goal attainment.

By analyzing the structure of relevant communities and the position of the institution or its members within these communities, buy observing and analyzing ongoing conversations within relevant communities, universities will be able to learn more about the cognitive, affective and conative reactions of their target audiences than ever before. Of course, we are only beginning to understand the reliability, validity and limitations of metrics derived from social media analytics—and thereby the quality of their contribution to success measurement and university governance. This chapter should therefore be read as a call for further research into these matters. Yet, as the propositions set forth in this chapter indicate, there is ample reason to expect that, thanks to the affordances of social media, universities will shortly widen the scope of their impact assessment beyond impact factors and h-indices. As of today, there is ample room for the development of a much

(continued)

richer and more varied understanding of scientific success and impact on relevant stakeholders.

References

Adler J et al (2012) Warum wir aus dem Handelsblatt BWL-ranking ausgestiegen sind – ein offener Brief an das Handelsblatt. http://handelsblattranking.wordpress.com/2012/08/29/handelsblatt-ranking. Accessed 7 June 2014
Aguinis H, Suárez-González I, Lannelongue G, Joo H (2012) Scholarly impact revisited. Acad Manag Perspect 26(2):105–132
Bechar-Israeli H (1995) From <bonehead> to <clonehead>: nicknames, play and identity on internet relay chat. J Comput Mediat Commun 1(2)
Blumer H (1969) Symbolic interactionism – perspective and method. Prentice-Hall, Englewood Cliffs
Blumer H (1991) George Herbert Mead. In: Plummer K (ed) Foundations and history, vol 1, Symbolic interactionism. Edward Elgar, Aldershot, pp 143–176
Bollen J, Van De Sompel H, Hagberg A, Chute R (2009) A principal component analysis of 39 scientific impact measures. PLoS ONE 4(6):e6022
Bourdieu P (1990) Homo academicus. Stanford University Press, Palo Alto
Boyd D, Ellison NB (2007) Social network sites: definition, history, and scholarship. J Comput-Mediat Commun 13(1):210–230
Carroll AB, Buchholtz AK (2006) Business & society – ethics and stakeholder management, 6th edn. South-Western, Mason
Coase RH (1937) The nature of the firm. Economica 4(16):386–405
Coleman JS (1990) Foundations of social theory. Harvard University Press, Cambridge
Crane D (1972) Invisible colleges: diffusion of knowledge in scientific communities. University of Chicago Press, Chicago
Cyert RM, March JG (1963) A behavioral theory of the firm. Prentice Hall, Englewood Cliffs
Demsetz H (1988) The theory of the firm revisited. J Law Econ Org 4(1):141–161
Ellison N, Steinfield C, Lampe C (2007) The benefits of facebook “friends”: social capital and college students’ use of online social network sites. J Comput-Mediat Commun 12(4):1143–1168
Fishbein M, Ajzen I (1975) Belief, attitude, intention and behavior: an introduction to theory and research. Addison-Wesley, Reading
Fiske ST, Taylor SE (1991) Social cognition, 2nd edn. McGraw-Hill, New York
Freeman RE (1984) Strategic management – a stakeholder approach. Pitman, Boston
Freeman LC (2004) The development of social network analysis – a study in the sociology of science. Booksurge, North Charleston
Garfield E (2003) The meaning of the impact factor. Int J Clin Health Psychol 3(2):363–369
Giddens A (2004) The constitution of society. Polity, Cambridge
Grunig JE (1993) Image and substance: from symbolic to behavioral relationships. Public Relat Rev 19(2):121–139
Gruzd A, Goertzen M (2013) Wired academia: why social science scholars are using social media. In: Proceedings of the 46th Hawaii conference on system sciences, Grand Wailea, Maui, Hawaii, 7–10 Jan 2013, pp 3332–3341
Hirsch JE (2005) An index to quantify an individual’s scientific research output. Proc Natl Acad Sci USA 102(46):16569–16572
Hovland CI, Janis IL, Kelley HH (1953) Communication and persuasion – psychological studies of opinion change. Yale University Press, New Haven

Jeng W, He D, Jiang J (2014) User participation in an academic social networking site: a survey of open group users on Mendeley. J Assoc Inf Sci Technol
Kane GC, Alavi M, Labianca G, Borgatti SP (2014) What's different about social media networks? A framework and research agenda. MIS Q 38(1):274–304
Katz D, Kahn RL (1966) The social psychology of organizations. Wiley, New York
Katz E, Lazarsfeld PF (1955) Personal influence: the part played by people in the flow of mass communication. Free Press, New York
Kousha K, Thelwall M, Rezaie S (2010) Using the web for research evaluation: the integrated online impact indicator. J Inf 4(1):124–135
Lindenmann WK (2003) Public relations research for planning and evaluation. Institute for Public Relations, Gainesville
Mahrt M, Weller K, Peters I (2013) Twitter in scholarly communication. In: Weller K, Bruns A, Burgess J, Mahrt M, Puschmann C (eds) Twitter and society. Peter Lang, New York, pp 399–410
March JG, Simon HA (1959) Organizations. Wiley, New York
Mead GH (1965) Mind, self, and society: from the standpoint of a social. University of Chicago Press, Chicago
Nentwich M, König R (2012) Cyberscience 2.0 – research in the age of digital social networks. Campus Verlag, Frankfurt
Petty RE, Cacioppo JT (1986) The elaboration likelihood model of persuasion. Adv Exp Psychol 19:123–205
PLoS Editors (2006) The impact factor game. PLoS Med 3(6):e291
Priem J (2013) Scholarship: beyond the paper. Nature 495:437–440
Priem J, Hemminger BH (2010) Scientometrics 2.0: new metrics of scholarly impact on the social web. First Monday 15(7–5)
Procter R, Williams R, Stewart J, Poschen M, Snee H, Voss A (2010) Adoption and use of Web 2.0 in scholarly communications. Philos Trans A Math Phy Eng Sci 368(1926):4039–4056
Pscheida D, Albrecht S, Herbst S, Minet C, Köhler T (2014) Nutzung von social media und onlinebasierten Anwendungen in der Wissenschaft. http://www.qucosa.de/fileadmin/data/qucosa/documents/13296/Science20_Datenreport_2013_PDF_A.pdf. Accessed 19 Apr 2014
Puckett C, Hargittai E (2012) From dot-edu to dot-com: predictors of college students' job and career information seeking online. Sociol Focus 45:85–102
Rose AM (1962) A systematic summary of symbolic interaction theory. In: Rose AM (ed) Human behavior and social processes – an interactionist approach. Routledge, London, pp 3–19
Rosenberg MJ, Hovland CI (1960) Cognitive, affective, and behavioral components of attitude. In: Hovland CI, Rosenberg MJ (eds) Attitude organization and change – an analysis and consistency among attitude components. Yale University Press, New Haven, pp 1–14
Seglen PO (1997) Why the impact factor of journals should not be used for evaluating research. BMJ 314(7079):498–502
Thelwall M (2012) Journal impact evaluation: a webometric perspective. Scientometrics 92 (2):429–441
Thelwall M, Kousha K (2014) ResearchGate: disseminating, communicating and measuring scholarship. J Assoc Inf Sci Technol 65(4):721–731
Vaughan L, Hysen K (2002) Relationship between links to journal web sites and impact factors. Aslib Proc 54(6):356–361
Vaughan L, Shaw D (2003) Bibliographic and web citations: what is the difference? J Am Soc Inf Sci Technol 54(14):1313–1322
von Hayek FA (1945) The use of knowledge in society. Am Econ Rev 35(4):519–530
von Mises L (1996) Human action – a treatise on economics, 4th edn. Fox & Wilkes, San Francisco
Weick KE (1995) Sensemaking in organizations. Sage, Thousand Oaks
Zhao S, Grasmuck S, Martin J (2008) Identity construction on Facebook: digital empowerment in anchored relationships. Comput Hum Behav 24:1816–1836

Double-Blind Peer Review: How to Slaughter a Sacred Cow

Margit Osterloh and Alfred Kieser

Abstract The performance evaluation system in academia has been much criticized during the last years. But there are few suggestions how to improve it. In particular double blind pre-publication peer review has become a sacred cow that has not been touched. We analyze the flaws of the present system and discuss open post-publication peer review as a promising alternative.

In recent times quantitative performance evaluation in research has come under attack while qualitative pre-publication peer reviews upon which quantitative evaluation is built is still considered a sacred cow. Quantitative performance evaluation was introduced in universities and research institutions in the wake of New Public Management during the 1990s in order to make comparisons of scholarly performance more objective, to reduce cronyism, to activate competition and to make scholars accountable towards the taxpayers. But in the meantime there exists a considerable literature about the shortcomings and dysfunctional effects of quantitative performance evaluation, in particular of rankings (e.g., Adler and Harzing 2009; Alvesson and Sandberg 2013; Dunbar and Bresser 2014; Helbing and Balietti 2011; Kieser 2012; Lawrence 2003; Starbuck 2005, 2015; Walsh 2011). Even a popular medium like "The Economist" recently states that "Science goes wrong" (The Economist 2013a).

We discuss three main causes triggering complaints about the present system of performance evaluation: the quality of the double blind pre-publication peer review system upon which quantitative performance evaluation is piggy-packed, the validity of the performance indicators, and the negative unintended consequences of "governance by numbers". We then evaluate a new proposal for an open post-publication review system.

M. Osterloh (✉) • A. Kieser
Zeppelin University, Friedrichshafen, Germany
e-mail: margit.osterloh@business.uzh.ch

I.M. Welpe et al. (eds.), *Incentives and Performance*,
DOI 10.1007/978-3-319-09785-5_19

1 Questionable Quality of the Present Peer Review System

Peer reviews are the founding stones of academic research evaluation. In particular the double blind review system is treated as almost sacred (Judge et al. 2007, p. 503). However, the quality of this system has come under scrutiny (e.g., Abramo et al. 2009; Bedeian 2004; Frey 2003; Gillies 2005, 2008; Helbing and Balietti 2011; Kieser 2012; Starbuck 2005, 2006; Tsang and Frey 2007). These studies have disclosed severe problems (Campanario 1998a, b; Osterloh and Frey 2014).

Low interrater reliability: The correlation between the judgments of two peers is low. In 16 studies, Starbuck (2005, 2015) found a mean correlation between reviewers of only 0.18. In clinical neuroscience the correlation between reviewers "was little greater than would be expected by chance alone" (Rothwell and Martyn 2000, p. 1964). Luck of the referee drawn plays a big role (Bornmann and Daniel 2009).

Low prognostic quality: Assessment of promotion decisions in three prestigious management departments estimated an error rate of 60 % (Starbuck 2015). Reviewers' ratings of manuscript quality correlated only 0.24 with later citations (Gottfredson 1978). Many articles published in "top" journals are rarely cited, which means that the reviewers do not judge the future impact in a satisfactory way (Hudson 2013; Laband 2013). This is the reason why the rating of a journal should not be used as an indicator of quality of a single article published in the respective journal (Baum 2011; Kieser 2012).

Low consistency over time: Highly ranked journals rejected many papers that thereafter received distinguished prizes, among them the Nobel Prize (Campanario 1996; Gans and Shepherd 1994; Lawrence 2003).

Confirmation bias: Referees score papers according to whether the results conform or conflict with their preferences (Bedeian 2004; Campanario 1998b). Papers threatening the previous work of reviewers tend to be rejected (Lawrence 2003). As a consequence, unorthodox research has a small chance of being published.

Superficial and coercive reviews: Due to the anonymity of reviewers as well as the lack of incentives for investing in thorough reviews, reviewers often work superficially (Eisenhart 2002). In addition, their comments are not perceived as advice, but as coercion to get a paper published (Frey 2003). The double blind peer review system induces reviewers not to behave as peers but as rulers who never receive a critical feedback (Tsang and Frey 2007). As a consequence, a high percentage of authors revise their manuscript, even when they are convinced that the reviewers are wrong (Bedeian 2003).

Lack of transparency: Editors conceal important information, such as which school of thought the reviewer is related to.

Delays: Publication to journals often is delayed by months, if not years. Since scientific papers are the major media of scholarly communication, the progress of science is slowed down considerably.

The current system is very expensive since scientific publishers are predominantly for-profit organizations that charge high to outrageous prices for their services. It is doubtful whether the benefit of the services provided to the science system that is essentially financed by the tax payer justifies these costs.

Based on this evidence an overview article on peer reviews states bluntly "Journal peer review is thus an unreliable quality control mechanism" (Campanario 1998b, p. 299). Nevertheless the present system of anonymous peer reviews is seldom put into question (e.g., Eisenhart 2002). In many articles that deal with the problems of academic performance evaluation the problems of double blind peer reviews are not even mentioned (e.g., Alvesson and Sandberg 2013; Dunbar and Bresser 2014; Flickinger et al. 2014; Hudson 2013; Laband 2013; Walsh 2011). Double blind peer reviews remain a sacred cow.

Nevertheless we emphasize that peer judgements are indispensable for scientific progress, with the exception of double blind reviews. The fact that they do not accord often is not detrimental, but—in contrary—may indicate productive and solid research (Campanario 1998a). However, this is the case only if diversity of judgements and findings fuel an open scholarly discourse within the "republic of science" (Polanyi 1962). Blind reviews in a publish-or-perish career system oppress such discourses.

2 The Seductive Power of Numbers

Rankings today are accepted as valid performance indicators that shape decisions on hiring, tenure, income, and allocation of resources though they are piggybacked on peer judgements that are highly controversial and sometimes heavily flawed. Why could that happen? First, some authors believe that by aggregating independent judgements individual reviewers' biases can be mitigated: Aggregation allows for error compensation, enables a broader perspective (e.g., Abramo et al. 2009; Weingart 2005), and thus might represent the "collective wisdom" of the scientific community more correctly than a low number of peer reviews (Laband 2013).[1]

Second—and more important—numbers are seductive (Nkomo 2009). It is more convenient to read rankings than articles: "numbers travel better and faster than words" (Heintz 2010, p. 167). Numbers are "immutable mobiles" (Latour 1988). They can be moved and combined easily, and they can be aggregated to complex constructions. During this process numbers lose their context-embeddedness. At the same time the necessity to know the context to interpret numbers correctly is downplayed. A self-vindication of numbers sets in (Power 2004) that fabricates the world it pretends to represent. Moreover, numbers appear as "hard facts" that

[1] Whether peer reviews are really independent is questionable as long as they are part of research communities that share assumptions.

can be used in the form of output-control. In contrast to process or input control this kind of control can be applied by non-experts that do not know and are not involved in the context (Osterloh 2010; Ouchi 1979). This is the reason why journalists, politicians, and managers love numbers. Also researchers use such indicators when trying to inform themselves in areas they are not involved in. Sometimes they are even convinced that they should trust aggregated indicators more than their own judgement (Laband 2013).

It follows that these numbers when publicly communicated shape the cognition of the users. In this way, "governance by numbers" has changed the behaviour of scholars even when they criticize it or do not believe in its legitimacy (Flickinger et al. 2014; Sauder and Espeland 2009). They also contribute to identity constructions: "Who am I? I am a person who has published in this or that journal." (Alvesson and Sandberg 2013, p. 136) Such "journal fetishism" (Willmott 2011) leads scholars to care more about the publication outlet than the content of the research.

Such reactions are aggravated by a lock-in effect, i.e., a self-vindicaton of numbers at the institutional level. This effect is closely related to Goodhart's Law (1975) in monetary policy, and to the Lucas Critique (1976) in econometric modelling. Examples are editors who encourage authors to cite their respective journals in order to raise their impact rankings (Garfield 1997; Monastersky 2005; Smith 1997),[2] or universities like the King Saud University in Riyadh, that offer cash to highly cited researchers for adding the university's name to their research papers. In this way, the King Saud University has improved its position in the Webometric ranking from position 2,910 in 2006 to position 186 in 2011 (Bhattacharjee 2011).

Self-vindication of numbers happens within all kinds of "governance by numbers" be it in the area of research, politics or economics. It makes performance indicators become a dependent variable (March and Sutton 1997). But with research this self-vindication has more serious consequences. In this field there exists a fundamental uncertainty. Success in research only rarely corresponds with success in the commercial market (Bush 1945; Nelson 2004). In addition, research often produces serendipity effects; that is, it provides answers to unasked questions whose usefulness is revealed only after a long time (Stephan 1996). Therefore peer judgements are decisive to determine whether research is done in a solid way. As we have argued, the present double blind peer reviews do not fulfil this task satisfactorily. Whether by aggregation of peer judgements individual biases are mitigated is highly questionable. Among other things it is dependent on how the indicators are constructed and used.

[2] According to Wilhite and Fong (2012) journals published by commercial companies show greater use of coercive tactics. These authors also find that highly ranked journals are more likely to coerce, but the direction of causality is unclear.

3 The Validity of Scholarly Performance Indicators

The validity of scholarly performance indicators based on publications and citations is put into question for three reasons (Kieser 2012; Osterloh and Frey 2014).

First, there exist technical and methodological problems, e.g., when attributing publications and citations to authors and institutions. Small changes in measurement techniques and classifications can have considerable consequences for the position in rankings (Frey and Rost 2010; Ursprung and Zimmer 2006). There are also selection problems. Usually only journal articles are selected for consideration in citation-based metrics thus discrediting the contributions of books and articles in edited books. Other difficulties include the low representation of small research fields and non-English journals. Moreover, citations can have a supportive or negative meaning or merely reflect herding, and incorrect citations are endemic (Simkin and Roychowdhury 2005; Woelert 2013). The comparison of numbers of publications, citations and impact factors not only between disciplines but also between sub-disciplines does not make sense (Bornmann et al. 2008; Kieser 2012). Therefore, using citations or citation-based rankings as indicators for scholarly performance is highly problematic.

Second, using impact factor as a proxy for the quality of an article published in a journal is very common, but leads to large error probabilities (Adler et al. 2008). The "extreme variability in article citedness permits the vast majority of articles—and journals themselves—to free-ride on a small number of highly cited articles" (Baum 2011, p. 449). Many top quality articles are published in non-top journals, and many articles in top journals generate very few citations (Campbell 2008; Kriegeskorte 2012; Laband and Tollison 2003; Oswald 2007; Singh et al. 2007; Starbuck 2005). Nevertheless this indicator has become internationally accepted (e.g., Abramo et al. 2009; Archambault and Larivière 2009; Jarwal et al. 2009). In recent times more and more critical voices tell us that journal based metrics such as impact factors should not be used as a surrogate measure of the quality of an individual article or of the excellence of a researcher (e.g., Alberts 2013; Baum 2011; Kieser 2012). Nevertheless, this performance indicator is widely used and often decisive for scholarly careers.

Third, scholars react to indicators by changing their behaviour in the form of goal displacement and counterstrategies. Researchers learn and are advised how to maximize different metrics (e.g., Dumas et al. 2011; Lalo and Mosseri 2009). As a consequence, indicators not only lose their ability to discriminate between good and bad performance, but they distort good performance (Frost and Brockmann 2014; Meyer and Gupta 1994). Such effects in research have recently been widely discussed (e.g. Osterloh and Frey 2014). Examples are:

Replication studies are rarely done because such studies cannot be published in prestigious journals (The Economist 2013b).

Studies that do not support hypotheses are hard to publish though they contribute considerably to scientific progress.[3] As a consequence, HARKing—Hypothesizing After Results are Known—has become common (Bedeian et al. 2010; Starbuck 2015).

Intrinsic motivation and curiosity are crowded out by extrinsic rewards. The "taste for science" Merton (1973) is replaced by a "taste for publication" or a "taste for rankings". As a consequence, incremental "game-spotting research" flourishes that is conducted rigorously, but lacks intellectual boldness (Alvesson and Sandberg 2013).

Authors trying to please reviewers change their manuscripts to get them accepted in a way that contradicts their convictions. Frey (2003) calls this behaviour "publication as prostitution".

Cheating, e.g., by withholding data that contradict own previous research or by fabricating results is a mounting concern among scholars. Such "Win at all Costs" behaviour was revealed by a survey among US business schools. The authors compound such behavior "by the unprecedented importance attached to publishing in 'A-level' journals" (Bedeian et al. 2010, p. 720). Researchers are caught in a "Publication Impossibility Theorem System" or PITS (Frey 2009). They are expected to publish in A-journals, but for the overwhelming majority this is impossible. Very few slots in such journals heat up frustrating tournaments.

4 Is It Possible to Steer Away from Sacred Cows?

There exist some suggestions how to deal with the flaws of the present system of performance evaluation. First, the individual researcher is addressed. For example, Starbuck (2015) advises scholars perceiving unreliable research evaluations as liberating. They should "dare not to depend on editors". Other authors recommend training and workshops about correct scholarly behavior (Alvesson and Sandberg 2013), or deeper, reflexive conversations about what we do as scholars (Nkomo 2009). Second, some suggestions are made on the level of institutions, e.g., to create more and better indicators (e.g., Abramo et al. 2009; Adler and Harzing 2009; Starbuck 2009), to rely more on citations (e.g., Alvesson and Sandberg 2013), or to allow to submit a paper to several journals at the same time (Dunbar and Bresser 2014). But these suggestions do not deal with the problems of double blind peer reviews.

If one agrees that in research peer judgements as well as scholarly debate are the founding stones of scholarly performance evaluation, and at the same time the founding stone of scientific progress, than we need something different than the double blind peer review system. What is needed is a peer evaluation system that builds on reading and discussing papers instead of counting numbers (Lawrence 2008; San Francisco Declaration on Research Assessment (DORA) 2012). In

[3] According to Popper (2005) falsification of hypotheses is even the only avenue to scientific progress.

particular we need a system that first takes reviewer biases and conflicting views into account, acknowledging that diversity of scholarly views is necessary for scientific progress as long as an open discourse in the scientific community flourishes; second, gives advice without crowding out intrinsically motivated scholarly curiosity; third, is transparent, i.e., provides information to the public scholarly discourse which in the old system is only provided to the editor; fourth, is non-invasive, i.e., robust concerning reactivity; fifth, works without undue delays; sixth, is cost-efficient; seventh, provides advice on how to cope with the huge amount of scholarly publications without relying solely on numbers.

To meet these conditions, Osterloh and Frey (2014) suggest relying in the first place on input control, i.e., rigorous selection and socialization of scholars. Nevertheless selection of candidates, as well as selection of publications that are to be read, needs evaluations. We agree with Kriegeskorte (2012, p. 1):

> For better or worse, the most visible papers determine the direction of each field and guide funding and public policy decisions. Evaluation, therefore, is at the heart of the entire endeavor of science. As the number of scientific publications explodes, evaluation and selection will only gain importance. A grand challenge of our time, therefore, is to design the future system, by which we evaluate papers and decide which ones deserve broad attention.

5 Open Post-Publication Peer Review: The Future of Scientific Publishing?

As we have seen, the review process in its present form is invalid, intransparent, unreliable, inefficient, and hinders scholarly discourses that are at the heart of scientific progress. Reforms of the current system are urgently needed. Open-post-publication reviewing (OPR) as suggested Kriegeskorte (2012; see also Kriegeskorte and Deca 2012) is a promising suggestion that meets the conditions mentioned. It has the potential to revolutionize scientific publishing. Our following discussion of OPR closely follows Kriegeskorte's (2012) concept complemented by critical comments.

The openness of the new system: OPR means that any scientist can at any time publish a peer review on any published paper by submitting his or her comments to an online public repository. Reviews can take the form of review essays. According to Kriegeskorte they can also take the form of numerical quality ratings, which we see as problematic. They can be signed by authors or be submitted anonymously. The repository will link each paper to all its reviews, so that each paper is automatically accompanied with the evaluative information from different reviewers. Any scientist can freely submit a review on any paper and anyone can freely access any review. Papers already published in the current system can also be commented in post-publication reviews.

We find that the choice between signed and anonymous reviews is not unproblematic but, as Kriegeskorte holds, it is unavoidable:

> There is some evidence that the threat of revealing the reviewer's identity to the authors or of making a review public may just deter reviewers and do little to improve review quality. This highlights the need to give reviewers a choice of whether or not to sign. Moreover, defining reviews as open letters and mini-publications will create a different culture, in which scientists define themselves not only through their own work, but also through others' work they value. Signed evaluations have the advantage that they attach the reviewer's reputation to the judgment, thus alleviating abuse of reviewer power. Anonymous reviews have the advantage that they enable reviewers to criticize without fear of negative consequences. Both types are needed, and a scientist will make this choice on a case-by-case basis. The anonymous option will encourage communication of critical arguments. But to the extent that an argument is objective, sound, and original, a scientist will be tempted to sign in order to take credit for his or her contribution. In analyzing the review information to rank papers, signed reviews can be given greater weight if there is evidence that they are more reliable. (Kriegeskorte 2012, p. 7)

The first reviews are initiated by an editor chosen by the author: The first reviews of a paper can be of crucial importance. Kriegeskorte therefore suggests introducing an editor who selects the first reviewers. However, in contrast to the traditional system, the editor is chosen by the author and the first reviewers sign their reviews. Soon after publication, the author asks a senior scientist in his or her field to edit the paper. The senior scientist's role will be acknowledged in the paper. The editor has to find two to four reviewers who are willing to publicly review the paper. Of course, the reputation of the editor and of the first reviewers has an impact on the attention a paper will receive. The author may also inform other scientists of the publication and ask them to review it. Author- and editor-requested reviews will be marked as such.

Paper evaluation functions (PEFs): The system will allow any reader and any organization to define a PEF based on content and quality criteria and will automatically be informed about papers' fit to these criteria. The PEF could for example, filter out anonymous reviews or weight evidence for central claims over potential impact of the results (e.g., may even take impact factors into consideration, though we would not recommend to do so).

Individuals or organizations, e.g., associations for academic fields, may define PEFs according to their own priorities and publish the resulting evaluation through a web-portal. The free definability of PEFs will create a plurality of perspectives of the literature as long as no single PEF has gained dominance. The permanent evolution of multiple PEFs impedes "gaming the game" (Macdonald and Kam 2007) because PEFs can be adjusted in response to attempts to game the system. In general, manipulation exercises can be detected since, at least with signed reviews, as the community can identify who, together with whom, is taking a positive or negative stance for or against a paper.

> Web-portals can define PEFs for subcommunities—for scientists too busy to define their own. A web-portal can be established cheaply by individuals or groups whose members share a common set of criteria for paper prioritization. (Kriegeskorte 2012, p. 7)

Will it be possible to motivate reviewers to write reviews of high quality? Because public reviews are directed to the community, their impact to a large extent depends on the persuasiveness of their arguments. This sharply contrasts with the secret traditional peer review that often reflects political intentions of reviewers who see themselves as the author's competitors or as supporters of a representative of a theory which they find attractive. It can be assumed that the motivation to review a paper is greater in the new system than in the current one since the signed review is a publication with which reputation can be earned. Public reviews make reviewing a more meaningful and motivating activity. Scientists asked to write a pre-publication review for a traditional journal might also ask themselves whether they should publish their (originally secret) review once the respective paper has been published. Signed reviews will be citable publications. As the new system develops, one can imagine that reviews will be cited and quoted in articles and other reviews. One can imagine that exciting public discourses will develop between authors and reviewers and among reviewers.

Kriegeskorte (2012, p. 8) argues that control through the scientific community will improve the quality of evaluations:

> A core feature of this proposal is a clear division of powers between the OE[4] system, which accumulates reviews and ratings and links them to the papers they refer to, and the PEFs, which combine the evaluative evidence so as to prioritize the literature from particular perspectives. This division of powers requires that the evidence accumulated by the OE system is publicly available, so that independent groups and individuals can analyze it and provide PEFs. This division of powers ensures transparency and enables unrelated groups and individuals to freely contribute to the evaluative evidence and to its combination for prioritizing papers. For example, if a group of scientists started doing mutual favors by positively evaluating each other's papers, an independent group could build a PEF that uses only signed evaluations and downweights evaluations from individuals within cliques of positive mutual evaluation. Conversely, when a web-portal claims to combine the evaluative evidence by a given PEF to compute its paper ranking, anyone can re-implement that algorithm, run it on the public evaluative evidence, and check the ranking for correctness. This fosters a culture in which we keep each other honest, and in which public interest and self-interest are aligned.

A minimalist quantitative evaluation mechanism for rapid ratings: Kriegeskorte (2012, p. 8) recommends the provision of a minimalist open evaluation formula that reminds one of minimal "like" and "dislike" buttons but goes beyond them. He admits that such buttons are not suitable for evaluating scientific papers. Therefore, assuming that reviewers are familiar with impact factors, he suggests using a one-dimensional rating scale for the rapid review called "desired impact in impact factor units" that ranges from 0 to 12. We do not agree with this idea. This suggestion leads to the seductive power of numbers mentioned, and neglects that discourses, not counting, are at the heart of scientific progress. It would be misleading to re-introduce impact factors, which Kriegeskorte (2012, p. 3) himself characterizes as "impoverished evaluative signals" that "are kept secret and contribute to the reception of a paper only after being reduced to a categorical quality

[4] OE = Open Evaluation.

stamp". This would be true not only for a journal label, but also for a "desired impact" label.[5]

Revisions: The author can submit a revision of his or her paper. A revision is a separately published document linked to the previous version of the paper and accompanied by a "justification statement" that addresses the changes (typically in response to reviewers' comments). The justification statement is needed for the revision to replace the previous version. The authors of signed reviews are automatically informed about revisions and can revise their reviews in the light of the revision. Revisions may be limited to a certain number per year.

In the long run, rankings may become superfluous: The review phase is open ended and so important papers will accumulate more reviews over time. They will trigger more intensive discourses between reviewers. They will also more often be identified as important in evaluative web-portals.

A new culture of scientific publishing: Kriegeskorte (2012, p. 12) optimistically assumes that an OPR could create a new culture of scientific publishing overcoming the drawbacks of the current system:

> Open evaluation goes hand in hand with a new culture of science. This culture will be more open, transparent, and community controlled than the current one. We will define ourselves as scientists not only by our primary research papers, but also by our signed reviews, and by the prior work we value through our public signed ratings. The current clear distinction between the two senses of 'review' (as an evaluation of a particular paper and as a summary and reflection upon a set of prior papers) will blur. Reviews will be the meta-publications that evaluate and integrate the literature and enable us as a community to form coherent views and overviews of exploding and increasingly specialized literatures.

After some years a PEF could collect contributions that have elicited the most inspiring debates. The original authors are given the opportunity to present their work to a broader audience as "state of the art". In contrast to the present system, this would be the outcome of a transparent and collaborative process.

6 Evaluation of the Open Post-Publication Evaluation System

A number of questions arise concerning the new system. Kriegeskorte (2012) raises the following questions:

Won't the literature be swamped with low-quality papers that are never evaluated?
This could happen. The new system will not prevent us from reading papers that have not been evaluated yet (Kriegeskorte 2012, p. 12/13). However, with the

[5] The "desired impact factor" Kriegeskorte mentions is different from a journal impact factor. Therefore, the problem of free-riding on a small number of highly cited articles in high-impact-journals is avoided with Kriegeskorte's approach. However, when certain PEFs will have gained prominence, counting exercises are likely to set in.

present peer review system the attention of readers is often enough misguided by criteria hard to trace. With evaluated papers the new system helps to select what to read on the basis of open and transparent criteria. Moreover, scholars are motivated to seek informal feedback before publishing their paper, because a published paper cannot be erased from the record (Kriegeskorte 2012, p. 15).

What will happen to papers that will end up in "the twilight zone of unevaluated papers"? Kriegeskorte (2012, p. 13) answers that "as readers, we do not mind this, because twilight papers will not come to our attention unless we explicitly search for them". On the other hand, scholars have to learn to connect better with peers by informal feedbacks to find an initial audience and a peer-to-peer editor. As a disadvantage, there is no doubt that this gives room for some scholarly "marketing" and even cronyism.

How can reviews and reviewers be evaluated? Kriegeskorte (2012, p. 14/15) makes two suggestions. In the first one he recommends using quantitative criteria, e.g., the weighting of reviews according to the reputation of the reviewers (measured, e.g., by the H-index) or by estimating the predictive power of their past reviews by relying on meta-reviews of their past reviews. We are skeptical concerning this suggestion since it activates the "seductive power of numbers" as well as reactive measures. It does not encourage scholarly debates and might strengthen mainstream research. The second suggestion is a qualitative one: A reviewer who signs a review or rating links his or her reputation to a paper. "Performance could be judged as high if the reviewer's judgment stands the test of time, and especially high if this evaluation was made early and/or diverged from existing evaluations when it was entered."(p. 15). We appreciate highly the idea to make reviewing in this way a competitive, public activity that at the same time promotes the scholarly discourse.

In sum, post-publication review has advantages and disadvantages. The most important disadvantages are, first, that it gives only few guidelines of how to select newly published papers for reading. However, as mentioned, the old system often misguides this selection by an intransparent review process and one-dimensional rankings. Second, the necessity to choose an editor who selects the first reviewers may trigger undue advertisement efforts and even may promote cronyism. However, in the new system cronyism is much harder to hide since judgments have to be substantiated by arguments. Third, we do not support Kriegeskorte's suggestion to use quick evaluation formulas. What is counted counts. It leads to empoverished evaluative signals and seduces peers not to use their own brain. Nevertheless also in the new system evaluations are in danger of the "seductive power of numbers" as soon as some communities or persons become dominant and their judgments are weighted.

The most important advantages of the open post-publication peer review system is that it strengthens open scholarly discourses about different views that are at the heart of scientific progress. It does so first, by enhancing the quality of peer reviews, making them part of the scholarly reputation system and by impeding intransparent power of reviewers. Second, it avoids the lock-in effect of misleading performance indicators and encourages to care for content instead of publication outlets. Third, it is more resistant to "gaming the game" efforts because it is based on comprehensible and transparent arguments instead of numbers.

References

Abramo GD, Angelo CA, Caprasecca A (2009) Allocative efficiency in public research funding: can bibliometrics help? Res Policy 38:206–215

Adler NJ, Harzing A-W (2009) When knowledge wins: transcending the sense and nonsense of academic rankings. Acad Manag Learn Edu 8:72–95

Adler R, Ewing J, Taylor P (2008) Citation statistics. 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). Stat Sci 24:1–14

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

Alvesson M, Sandberg J (2013) Has management studies lost its way? Ideas for more imaginative and innovative research. J Manag Stud 50(1):128–152

Archambault É, Larivière V (2009) History of the journal impact factor: contingencies and consequences. Scientometrics 79(3):639–653

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 18:449–466

Bedeian AG (2003) The manuscript review process: the proper roles of authors, referees, and editors. J Manag Inq 12:331–338

Bedeian AG (2004) Peer review and the social construction of knowledge in the management discipline. Acad Manag Learn Educ 3:198–216

Bedeian AG, Taylor SG, Miller AN (2010) Management science on the credibility bubble: cardinal sins and various misdemeanors. Acad Manag Learn Educ 9(4):715–725

Bhattacharjee Y (2011) Saudi universities offer cash in exchange for academic prestige. Science 334(6061):1344–1345

Bornmann L, Daniel HD (2009) The luck of the referee draw: the effect of exchanging reviews. Learned Publish 22(2):117–125

Bornmann L, Mutz R, Neuhaus C, Daniel D (2008) Citation counts for research evaluation: standards of good practice for analyzing bibliometric data and presenting and interpreting results. Ethics Sci Environ Politics 8:93–102

Bush V (1945) Science: the endless frontier: report to the president by Vannevar Bush, director of the office of scientific research and development. United States Government Printing Office, Washington, DC

Campanario JM (1996) Using citation classics to study the incidence of serendipity in scientific discovery. Scientometrics 37:3–24

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

Campanario JM (1998b) Peer review for journals as it stands today, part 2. Sci Commun 19 (4):277–306

Campbell P (2008) Escape from the impact factor. Ethics Sci Environ Politics 8:5–7

Dumas M, Garcia L, Kisselite K, Kungas P, Trepowski CP (2011) Homophily-weighted citation measures. Interdisciplines. http://www.interdisciplines.org/paper.php?paperID=111. Accessed 24 Mar 2014

Dunbar RLM, Bresser RK (2014) Knowledge generation and governance in management research. J Bus Econ 84:129–144

Eisenhart M (2002) The paradox of peer review: admitting too much of allowing too little. Res Sci Educ 32:241–255

Flickinger M, Tuschke A, Grueber-Muecke T, Fiedler M (2014) In search of rigor, relevance, and legitimacy: what drives the impact of publications? J Bus Econ 84:99–128
Frey BS (2003) Publishing as prostitution? – choosing between one's own ideas and academic success. Public Choice 116:205–223
Frey BS (2009) Economists in the PITS. Int Rev Econ 56(4):335–346
Frey BS, Rost K (2010) Do rankings reflect research quality? J Appl Econ 13(1):1–38
Frost J, Brockmann J (2014) When quality is equated with quantitative productivity – scholars caught in a performance paradox. Zeitschrift für Erziehungswissenschaft (in press)
Gans JS, Shepherd GB (1994) How are the mighty fallen: rejected classic articles by leading economists. J Econ Perspect 8:165–179
Garfield E (1997) Editors are justified in asking authors to cite equivalent references from same journal. Br Med J 314:1765
Gillies D (2005) Hempelian and Kuhnian approaches in the philosophy of medicine: the Semmelweis case. Stud Hist Phil Biol Biomed Sci 36:159–181
Gillies D (2008) How should research be organised? College Publication King's College, London
Goodhart C (1975) Monetary relationships: a new form of Threadneedle street. Papers in Monetary Economics 1, Reserve Bank of Australia
Gottfredson SD (1978) Evaluating psychological research reports: dimensions, reliability, and correlates of quality judgments. Am Psychol 33(10):920–934
Heintz B (2010) Numerische Differenz. Überlegungen zu einer Soziologie des (quantiativen) Vergleichs. Z Soziol 39(3):162–181
Helbing D, Balietti S (2011) How to create an innovation accelerator. http://arxiv.org/abs/1011.3794v3. Accessed 24 Mar 2014
Hudson J (2013) Ranking journals. Econ J 123:F202–F222
Jarwal SD, Brion AM, King ML (2009) Measuring research quality using the journal impact factor, citations and 'Ranked Journals': blunt instruments or inspired metrics? J High Educ Pol Manag 31(4):289–300
Judge TA, Cable DM, Colbert AE, Rynes SL (2007) What causes a management article to be cited – article, author, or journal? Acad Manag J 50(3):489–508
Kieser A (2012) JOURQUAL – der Gebrauch, nicht der Missbrauch, ist das Problem. Oder: Warum Wirtschaftsinformatik die beste deutschsprachige betriebswirtschaftliche Zeitschrift ist. Die Betriebswirtschaft 72:93–110
Kriegeskorte N (2012) Open evaluation: a vision for entirely transparent post-publication peer review and rating for science. Front Comput Neurosci 6:1–18
Kriegeskorte N, Deca D (eds) (2012) Beyond open access: visions for open evaluation of scientific papers by post-publication peer review. Special topic image. Frontiers, Lausanne
Laband DN (2013) On the use and abuse of economics journal rankings. Econ J 123:F223–F254
Laband DN, Tollison RD (2003) Dry holes in economic research. Kyklos 56:161–174
Lalo F, Mosseri R (2009) Bibliometric evaluation of individual researchers: not even right... not even wrong! Europhys News 40(5):26–29
Latour B (1988) Drawing things together. In: Lynch M, Woolgar S (eds) Representation in scientific practice. MIT Press, Cambridge, pp 19–68
Lawrence PA (2003) The politics of publication – authors, reviewers, and editors must act to protect the quality of research. Nature 422:259–261
Lawrence PA (2008) Lost in publication: how measurement harms science. Ethics Sci Environ Politics 8(1):9–11
Lucas RE (1976) Econometric policy evaluation: a critique. In: Brunner K, Meltzer AH (eds) Carnegie-Rochester conference series on public policy. The Phillips curve and labor markets. North Holland, New York, pp 19–46
Macdonald S, Kam J (2007) Ring a ring o' roses: quality journals and gamesmanship in management studies. J Manag Stud 44:640–655
March JG, Sutton RI (1997) Organizational performance as a dependent variable. Organ Sci 8 (6):698–706

Merton RK (1973) The sociology of science: theoretical and empirical investigation. University of Chicago Press, Chicago
Meyer MW, Gupta V (1994) The performance paradox. Res Organ Behav 16:309–369
Monastersky R (2005) The number that's devouring science. Chron High Educ 52(8):A12
Nelson R (2004) The market economy, and the scientific commons. Res Policy 33:455–471
Nkomo SM (2009) The seductive power of academic journal rankings: challenges of searching for the otherwise. Acad Manag Learn Educ 8:106–112
Osterloh M (2010) Governance by numbers. Does it really work in research? Analyse Kritik 32 (2):267–283
Osterloh M, Frey BS (2014) Ranking games. Eval Rev (in press)
Oswald AJ (2007) An examination of the reliability of prestigious scholarly journals: evidence and implications for decision-makers. Economica 74:21–31
Ouchi WG (1979) A conceptual framework for the design of organizational control mechanisms. Manag Sci 25:833–848
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. University of Chicago Press, Chicago, pp 49–72. Re-reprinted in Mirowski P, Sent EM (eds) (2002) Science bought and sold. Essays in the economics of science. The University of Chicago Press, Chicago, pp 465–485
Popper K (2005) The logic of scientific discovery. Routledge/Taylor & Francis e-Library, London
Power M (2004) Counting, control and calculation: reflections on measuring and management. Hum Relat 57:765–783
Rothwell PM, Martyn CN (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
San Francisco Declaration on Research Assessment (DORA) December 16, 2012. http://am.ascb.org/dora/files/SFDeclarationFINAL.pdf. Accessed 24 Mar 2014
Sauder M, Espeland WN (2009) The discipline of rankings: tight coupling and organizational change. Am Sociol Rev 74:63–82
Simkin MV, Roychowdhury VP (2005) Do copied citations create renowned papers? Ann Improb Res 11(1):24–27
Singh G, Haddad KM, Chow S (2007) Are articles in "top" management journals necessarily of higher quality? J Manag Inq 16:319–331
Smith R (1997) Journal accused of manipulating impact factor. Br Med J 314:463
Starbuck WH (2005) How much better are the most prestigious journals? The statistics of academic publication. Organ Sci 16:180–200
Starbuck WH (2006) The production of knowledge. The challenge of social science research. Oxford University Press, Oxford
Starbuck WH (2009) The constant causes of never-ending faddishness in the behavioural and social sciences. Scand J Manag 25:225–227
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
Stephan PE (1996) The economics of science. J Econ Lit 34:1199–1235
The Economist (2013a) How science goes wrong. October 19th–25th 2013, p 11
The Economist (2013b) Trouble at the lab. Scientists like to think of science as self-correcting. To an alarming degree, it is not. October 19th–25th 2013, pp 21–24
Tsang EWK, Frey BS (2007) The as-is journal review process: let authors own their ideas. Acad Manag Learn Educ 6:128–136
Ursprung HW, Zimmer M (2006) Who is the "Platz–Hirsch" of the German economics profession? A citation analysis. Jahrbücher für Nationalökonomie und Statistik 227:187–202

Walsh J (2011) Embracing the sacred in our secular scholarly world. 2010 presidential address. Acad Manag Rev 36(2):215–234

Weingart P (2005) Impact of bibliometrics upon the science system: inadvertent consequences? Scientometrics 62:117–131

Wilhite AW, Fong EA (2012) Coercive citing in academic publishing. Science 335:542–543

Willmott H (2011) Journal list fetishism and the perversion of scholarship: reactivity and the ABS list. Organization 18:429–442

Woelert P (2013) The "Economy of Memory": publications, citations, and the paradox of effective research governance. Minerva 51:341–362

Part VI
What Research Organizations Can Learn from Knowledge-Intensive Business

The Professional Partnership: An Alternative Model for Public University Governance Reform

Nancy R. Edwards and Berthold U. Wigger

Abstract We propose an alternative model for university governance reform; namely, the professional partnership. To this end, we provide a theoretical foundation to support the propositions that the profession, as institution, and the partnership, as organizational form, have unique governance characteristics that are more effective for knowledge-intensive organizations, such as universities, than the corporate model of governance, on which the New-Public-Management-inspired governance model for German public universities, the so-called Neues Steuerungsmodell, is based.

Much of the past and current debate on university governance reform centers on the adoption of the corporate model of governance. Like the traditional model of public administration, the corporate model implies the management and governance of a large, hierarchical organization in which readily measurable outcomes result from work that can be—and often is—standardized and easily defined in corporate manuals. The theoretical foundation of corporate governance is agency theory, which is problematic for the German public university context, to the extent that no clear and uncontested agency relationships can be defined, with the result that artificial agency relationships must be constructed in order to operationalize the corporate model and its governance structures and instruments. Given the unsatisfactory theoretical underpinnings of this approach and the questionable results rendered by the corporate model, we take a different approach; namely we extend Williamson's transaction cost theory in offering a more theoretically-robust approach to the governance of knowledge-intensive services. Specifically, we extend transaction cost theory in several important ways that recognize the unique attributes of knowledge as a commodity, explicitly account for the knowledge dimension of human assets, account for the unique characteristics of knowledge workers, and, taking all of this together, extends the scope of viable governance

N.R. Edwards (✉) • B.U. Wigger (✉)
Karlsruhe Institute of Technology, Karlsruhe, Germany
e-mail: nancy.edwards@ucf.uni-freiburg.de; berthold.wigger@kit.edu

I.M. Welpe et al. (eds.), *Incentives and Performance*,
DOI 10.1007/978-3-319-09785-5_20

structures to include the broader professionalization of knowledge workers and the partnership as an organizational form and governance structure.

1 Characteristics of Knowledge Work and Knowledge Workers

A key characteristic of knowledge work is that it chiefly involves non-routine, non-standardized tasks that require domain-specific knowledge. To this end, knowledge workers "rely on a sophisticated combination of theoretical knowledge, analytical tools and tacit or judgmental skills that are very difficult, but not impossible, to standardize, replicate and incorporate within formalized organizational routines" (Reed 1996, p. 585). They are typically engaged in work that involves solving complex and unique problems—often for individual clients—by offering creative and innovative solutions, which make bureaucratic approaches to monitoring and control difficult (Alvesson 1993; Greenwood et al. 1990).

The other key characteristic of knowledge work is that it is often difficult for non-experts to evaluate. Credence goods represent such a special case. Problems concerning credence goods arise when an expert provider of services, who is more knowledgeable about the needs of the consumer than the consumer himself is, determines the quantity and quality of service that is provided (Emons 1997; Dulleck and Kerschbamer 2006). As noted by Emons (1997, p. 107; emphasis in the original), "aggravating this special feature is the fact that even *ex post,* consumers can hardly determine the extent of the service that was required ex ante. It is often difficult, if not impossible, to find out whether repairs were really needed or whether necessary treatments were not performed." Many knowledge-intensive services, including higher education and research, may be characterized as credence goods and therefore pose a unique governance problem.

We assume a definition of knowledge workers as experts in a domain who possess a high degree of domain-specific knowledge. This definition consequently includes the traditional professions (i.e., law, accountancy, medicine, etc.) as well as non-professional occupations that require expert-level knowledge (e.g., management consultancy, academia, etc.), reflecting the convergence in characteristics ascribed to both professional and non-professional knowledge workers (Alvesson 1993).

The key questions to be answered now are what differentiates knowledge workers from other categories of workers and what are the implications for the governance of organizations that employ mostly knowledge workers? As Foss (2007) argues, assumptions about the knowledge and motivation of economic actors are crucial for any informed discussion of governance and organization, yet have not been addressed in a systematic way in the case of knowledge workers. In reviewing the literature on knowledge workers, three key characteristics have been identified: a preference for collegiality in decision making; the performance of work that is non-routine and non-standardized; and, a preference for autonomy and freedom from external constraint in the work environment (Greenwood and Empson 2003; Moe 1995; von Nordenflycht 2010).

Collegiality in decision-making reflects the extent to which decisions in an organization are based on a flat hierarchy in which workers are active participants in key managerial decisions. Several authors confirm that the distribution of authority in knowledge-intensive firms reflects an emphasis on collegiality, peer evaluation, autonomy, informality and flexibility (Greenwood et al. 1990; Starbuck 1992). Lending further support, empirical research conducted by Sveiby and Simons (2002) found that a collaborative atmosphere even enhances the effectiveness of knowledge work. In the context of professional service firms, collegiality is reflected in members' "broad participation in strategic decisions, rotating executive positions and high individual autonomy in the production process" (von Nordenflycht 2007, p. 431).

According to Reed (1996, p. 585), knowledge workers, "specialize in complex task domains which are inherently resistant to incursions by the carriers of bureaucratic rationalization and control". This relates not only to the belief in self-regulation (i.e., that only other qualified experts are able to evaluate their work and, consequently, a preference for collegiality in decision making and control), but also to the desire for autonomy in organizing and conducting their work. As explained by Hall (1968, p. 93), autonomy "involves the feeling that the practitioner ought to be able to make his own decisions without external pressures from clients, those who are not members of his profession, or from his employing organization".

In summary, salient distinguishing characteristics have been identified that differentiate knowledge work and knowledge workers from other categories of work and workers. In what follows, we formulate three propositions, from which we derive our chief argument; namely, in favor of the professional partnership as an alternative approach to the governance of knowledge work and knowledge workers, in general, and for public university governance reform in particular.

2 The Profession as Institution and Mode of Governance

Proposition 1 Professionalization offers more effective governance of knowledge work and knowledge workers than the corporate model.

In order to discuss professionals in general or, in the university context, academics in particular as being a distinctive group apart from other (knowledge) workers, a definition of the professions must be established on which further distinguishing characteristics can be elaborated. Wilensky (1964, p. 138; emphasis in the original) defines professions according to the following two criteria. Firstly, "the job of the professional is *technical*—based on systematic knowledge or doctrine acquired only through long prescribed training. Secondly, professionals adhere to a set of *professional norms*; i.e., standards of conduct that inform their behavior vis-à-vis their clients and their colleagues." While the first characteristic overlaps with our broader definition of knowledge workers, the second characteristic can be viewed as distinguishing professionals from other knowledge workers. Thus, in addition to the characteristics common to knowledge workers discussed above (i.e., collegiality in decision-making, the nature of knowledge work, and preference for autonomy) members of recognized professions exhibit additional, unique characteristics that have important governance implications; namely, self-regulation and social control. These are manifested in the establishment and enforcement of strict requirements for entry into the profession (i.e., educational standards), the selection, training, promotion and socialization of aspiring entrants, resulting in a strong allegiance to the profession, the strict adherence to a code of ethics, and the trusteeship norm (Goode 1957; Greenwood et al. 1990; Hall 1968; von Nordenflycht 2007, 2010; Parsons 1939; Scott 1965; Starbuck 1992; Wilensky 1964).

Underpinning the self-regulation and social control among members of a profession is the adherence to strict codes of professional ethics. According to Starbuck (1992, p. 717) "professionals' ethical codes require them to serve clients unemotionally and impersonally, without self-interest." This echoes Wilensky's (1964) discussion of professional norms, which dictate that professionals have an obligation to perform technically competent, high-quality work while adhering to a service ideal, stipulate that, when the two are at odds, "devotion to the client's interests more than personal or commercial profit should guide decisions" (Wilensky 1964, p. 140). In addition, Starbuck (1992, p. 717) notes that "professionals identify strongly with their professions, more strongly than with their clients or their employers. They not only observe professional standards, they believe that only members of their professions have the competence and ethics to enforce these standards".

Professional codes of ethics are generally considered to be more severe than the laws and regulations with which professionals must comply, thus bestowing upon them a high status within the larger society. However, maintenance of this social standing requires credible enforcement of codes of ethics in order to sustain the legitimacy of self-regulation and shield professionals from the scrutiny of laymen[1] (Goode 1957; Hall 1968). Given the preference for autonomy and the belief that only other expert-professionals are competent to evaluate their work, it is in the interest of professionals to comply with and mutually enforce their codes of ethics.

Related to their adherence to a strong code of ethics and professional norms, professionals possess an attitudinal trait that further distinguishes professionals from other knowledge workers is their public service ethos (Hall 1968), also referred to by von Nordenflycht (2010) as the trusteeship norm. The trusteeship norm encompasses the notion that "professionals have a responsibility to protect the interests of clients and/or society in general," which lies "at the core of professional

[1] As explained by Goode (1957, p. 198), "in exchange for protection against the larger lay society, the professional accepts the social control of the professional community."

codes of ethics and is often contrasted against a “commercial” or “economic” ethos that allows unfettered pursuit of self-interest” (von Nordenflycht 2010, p. 163).

Codes of ethics and professional norms explicitly recognize and control for the information asymmetries inherent in most knowledge-intensive service transactions. Alluding to the specific credence goods nature of professional services, Goode (1957, p. 196) explains that, “socialization and social control in the professions are made important by the peculiarly exploitative opportunities the professions enjoy. The problems brought to the professional are usually those the client cannot solve, and only the professional can solve. The client does not usually choose his professional by a measurable criterion of competence, and after the work is done, the client is not usually competent to judge if it was properly done”. In a similar vein, Wilensky (1964, p. 140) explains that, “the client is peculiarly vulnerable; he is both in trouble and ignorant of how to help himself out of it. If he did not believe that the service ideal were operative, if he thought that the income of the professional were a commanding motive, he would be forced to approach the professional as he does a car dealer—demanding a specific result in a specific time and a guaranty of restitution should mistakes be made. He would also refuse to give confidences or reveal potentially embarrassing facts. The service ideal is the pivot around which the moral claim to professional status revolves”.

Finally, self-regulation and social control are anchored in the socialization process that aspiring entrants to a profession must undergo. Those aiming to enter a profession must meet the high educational requirements, as well as undergo extensive training during which intensive socialization occurs (Goode 1957). All of this culminates in a strong sense of allegiance to the profession. Members of professions have been described as exhibiting an esprit de corps and a sense of “being in the same boat”, which are said to be “fostered by such things as control of entry to the occupation, development of a unique mission, shared attitudes toward clients and society, and the formation of informal and formal associations” (Bucher and Strauss 1961, p. 330). According to Starbuck (1992), professionals identify more strongly with their profession than with either their clients or their employers, which Hall (1968, p. 93) refers to as “a sense of calling to the field.” As described by Goode (1957, p. 195; emphasis added), “typically a profession, through its association and its members, controls admission to training and requires far more education from its trainees than the containing community demands. Although the occupational behavior of members is regulated by law, *the professional community exacts a higher standard of behavior than does the law*. Both of the foregoing characteristics allow the professions to enjoy more prestige from the containing community than can other occupations. Thus, professionals stand at the apex of prestige in the occupational system”.

Thus far we have identified and discussed the unique attributes of professionals—both attitudinal and behavioral—that differentiate them from other labor market participants. Based on the foregoing discussion, we conclude that these attributes are clearly inconsistent with the assumptions about human behavior that underlie transaction cost theory—specifically, self-interested opportunism. An additional property that differentiates professionals from the archetypal homo economicus is that professionals have become institutionalized.

As with many social science theories and constructs, there lacks a clear and uncontested definition of what constitutes an *institution*. A key figure in the field of sociology and institutional theory, Scott (1995, p. 33) defines institutions as being composed of "cultural-cognitive, normative, and regulative elements that provide stability and meaning to social behavior" and have attained a high degree of resilience and legitimacy. Important work in sociology (see for example, Alvesson 1993, 2000; DiMaggio and Powell 1983; Meyer and Rowan 1977) has discussed professions as institutions. This interest in professions as institutions has stemmed in large part from the fact that "many of the traits that make the professions sociologically interesting grow from the dimension of *community*" (Goode 1957, p. 195; emphasis added).

Characteristic of each of the established professions, and a goal of each aspiring occupation, is the 'community of profession.' Each profession is a community without physical locus and, like other communities with heavy in-migration, one whose founding fathers are linked only rarely by blood with the present generation. It may nevertheless be called a community by virtue of these characteristics: (1) Its members are bound by a sense of identity. (2) Once in it, few leave, so that it is a terminal or continuing status for the most part. (3) Its members share values in common. (4) Its role definitions vis-à-vis both members and non-members are agreed upon and are the same for all members. (5) Within the areas of communal action there is a common language, which is understood only partially by outsiders. (6) The Community has power over its members. (7) Its limits are reasonably clear, though they are not physical and geographical, but social. (8) Though it does not produce the next generation biologically, it does so socially through its control over the selection of professional trainees, and through its training processes it sends these recruits through an adult socialization process (Goode 1957, p. 194).[2]

From this point of view, a professional community has its own symbolic systems, relational systems, routines, and artifacts, which support the three pillars—regulative, normative, and cultural/cognitive—of the 'profession as institution' (Scott 1995). Powell explains the three pillars as follows: "regulative elements emphasize rule setting and sanctioning, normative elements contain an evaluative and obligatory dimension, while cultural/cognitive factors involve shared conceptions and frames through which meaning is understood" (2007, p. 2). According to Powell, "each of Scott's pillars offered a different rationale for legitimacy, either by virtue of being legally sanctioned, morally authorized, or culturally supported. These two key treatments of institutional mechanisms underscore that it is critical to distinguish whether an organization complies out of expedience, from a moral obligation, or because its members cannot conceive of alternative ways of acting" (2007, p. 2).

[2] Ouchi (1979) distinguishes between three different governance models: markets, bureaucracies and "clans", where clans "rely upon a relatively complete socialization process which effectively eliminates goal incongruence between individuals" (1979, p. 833). In this respect, professionalization and the resulting professional communities as discussed here resemble clans.

The institutional effects of professionalization on professionals differentiate them in important ways from non-professional knowledge workers. As explained by Wilensky (1964, p. 141), "the degree of professionalization is measured not just by the degree of success in the claim to exclusive technical competence, but also by the degree of adherence to the service ideal and its supporting norms of professional conduct". Earlier work by Parsons (1939, p. 467) provides a more detailed discussion: "The professional type is the institutional framework in which many of our most important social functions are carried on, notably the pursuit of science and liberal learning and its practical application in medicine, technology, law and teaching". He elaborates further on this point as follows. "The institutional pattern governing professional activity does not, in the same sense, sanction the pursuit of self-interest as the corresponding one does in the case of business [...]. Business men are, for instance, expected to push their financial interests by such aggressive measures as advertising. They are not expected to sell to customers regardless of the probability of their being paid, as doctors are expected to treat patients" (Parsons 1939, p. 463). In Parsons' view, "success" and "achievement" are institutionally defined and differ fundamentally between business and the professions. These differences also form the institutional constraints on socially accepted and expected behavior for the members of these two distinct groups (i.e., businessmen and professionals). The fundamental conclusion that can be drawn is the following: transaction cost theory's self-interested and opportunistic homo economicus appears to be subject only to *external* regulative constraints on his behavior, whereas professionals are subject to both *external* and *internal* regulative, normative and cognitive constraints.

3 The Partnership as Organizational Form and Mode of Governance

Proposition 2 The partnership, as an organizational form, offers more effective governance of knowledge work and knowledge workers than the corporate model.

As an organizational form, when viewed from a legal and financial perspective, a partnership is created by a contractual agreement between two or more individuals who agree to share the risks and the profits resulting from the operation of a business. It has been observed that, "partnerships tend to occur among individuals of similar type and quality" (Kandel and Lazear 1992, p. 813). For the governance of knowledge-based organizations and of knowledge-intensive service transactions that are rendered to non-expert clients, the partnership, as an organizational form, possesses a number of governance features that distinguish it favorably from both market and hierarchy.[3]

[3] The partnership model has also been discussed under different labels in the management literature. Burns and Stalker (1961) distinguish between mechanistic and organic organizations. Bureaucracies are inherently mechanistic as they rely upon ramified hierarchies, explicit formal rules and large worker heterogeneity. Partnerships, in contrast, are organic by nature as they exhibit flat hierarchies, few formal rules and little heterogeneity among its members.

Firstly, given that an individual partner is, on the one hand, a mutual beneficiary of her partners' actions but is also, on the other hand, mutually liable, she has very strong incentives to monitor and control her partners' behavior.[4] This results in an internal agency problem that is resolved with a form of governance referred to by Fama and Jensen (1983) as a strong "mutual monitoring system". The unlimited personal liability to which partners are exposed thereby reinforces peer mutual monitoring and militates against self-interested opportunistic behavior (Kandel and Lazear 1992; Ribstein 2009). This is consistent with the preference for collegiality and the belief that only other experts or professionals are able to accurately and critically evaluate the non-routine, non-standardized work in which knowledge workers are engaged. It is also consistent with the self-regulation of the professions, which depend on strict compliance with own codes of ethics and professional norms for sustaining their elevated social status and maintaining their reputational capital. This combination of both the ability and the incentive to engage in peer mutual monitoring results in significantly more effective governance than, for example, profit participation in the corporate model (Kandel and Lazear 1992).

From the perspective of non-expert third parties—most notably clients—this goes a long way toward reducing the risks associated with the information asymmetries that complicate market contracting and lead to high transaction costs. In this way, the partnership optimizes transaction governance while significantly reducing transaction costs. Where credence goods transactions are concerned, we expect the differential to be even greater.

Secondly, the unification of ownership, management and control, combined with the tournament system of promotion to partner create incentives that promote knowledge creation and knowledge sharing by all members of a partnership—both among partners and non-partner professional staff. Both Foss (2007) and Hackett (2000) discuss the problems encountered by many knowledge-based organizations and identify a tendency among knowledge workers in traditional private and public bureaucratic organizations to hoard knowledge, which typically results from corporate-style performance-based incentives, as one of the most serious obstacles that such organizations must overcome. Sharing of knowledge and other resources (especially those that contribute to fixed costs), can allow knowledge workers who work together in a partnership context to realize economies of scale and scope in the services they offer, which leads to both lower production and transaction costs than would be realizable via either hierarchy (which discourages knowledge sharing) or market-based transactions between individual self-employed experts working by themselves (where knowledge-sharing is precluded due to competition).

Finally, the partnership is highly consistent with the professional allegiance to a profession. It is also more consistent with related professional norms—such as the

[4] In contrast to the limited liability enjoyed by shareholders, owing to the fact that the corporation is a legally separate entity, the partnership is not a legal entity separate from its partner-owners. Consequently, each partner bears unlimited personal liability for the debts of the partnership, even if they were incurred by another partner (on behalf of the partnership).

trusteeship norm and the service ideal—than either market contracting or hierarchical organization. As noted by Goode (1957, p. 197), "the professional who is also a bureaucrat becomes less directly dependent on the professional community for his career advancement, so that the ordinary sanctions of that community may have less impact." Subsequent empirical research by Hall (1968) lends further support to the claim that a bureaucratic organizational structure (i.e., akin to Williamson's "hierarchy") is incompatible with a professionalized workforce. The formal processes and hierarchical control structures associated with corporate governance and traditional public administration are therefore likely to conflict with informal processes and "collegial controls" that underlie the governance of professional partnerships (Moe 1995). Taken together with the unique attributes of knowledge workers and professionals—i.e., a preference for self-regulation, autonomy in conducting work, the non-routine, non-standardized nature of work performed (and the difficulty faced by non-experts in monitoring and controlling work), and the preference for collegiality in decision-making, a hierarchically structured organization, like a traditional bureaucracy, may well have the effect of undermining the desirable governance features associated with the institutional aspects of professionalization.

4 The Case for the Professional Partnership

Proposition 3 As embodied in the professional partnership model, when combined, these two complementary approaches optimize the governance of knowledge-based organizations, while minimizing governance costs.

On the basis of the foregoing discussion, we make a case for the partnership as the ideal organizational structure, arguing that it maximizes the governance effects related to the professions as well as those associated with the organizational structure as such.

The partnership as an organizational form is strongly linked to the professions, which may account for its historical prevalence among professional services, such as law and accountancy firms (e.g., Kandel and Lazear 1992; Ribstein 2009). In many respects, it reflects the defining institutional features of the professions. Firstly, it reflects the belief held by most professionals that non-professionals are not fit to monitor and control the work of professionals, therefore resulting in a preference for mutual peer monitoring of professionals by professionals.[5] Secondly, it reflects the preference for autonomy in conducting work and the preference for collegiality in decision-making at the managerial-level, given the relative equality and status of the partners. Thirdly, it is consistent with the trusteeship norm. That is, the view held by professionals that commercial interests—both own and those of

[5] For a detailed discussion of peer review in academia see Osterloh and Kieser (2015).

clients—should be secondary to compliance with professional standards and codes of ethics.[6] This is an important contrast to the corporate form, in which a separation of ownership control has led to the goal of shareholder value maximization and performance-based financial incentives for employees to narrowly pursue that goal. Lastly, and relatedly, the tournament system of promotion to partner that is frequently utilized by professional partnerships provides incentives for professional staff who aspire to become a partner to work towards the best interest of the partnership, rather than pursuing their own individual financial interests. Thus, the governance features of the partnership as an organizational form complements and reinforces many of the key governance features of the profession as an institution which, we believe, should result in optimal governance at minimum cost.

Professional service firms (PSFs) are businesses characterized as employing highly skilled individuals who provide knowledge- and human-capital intensive services. While these individuals typically belong to a profession, that need not be the case. Examples of PSFs frequently cited in the literature include law firms, accounting firms, management consultancies, medical practices and advertising agencies. Taking a resource-based view of the firm, human capital is the key source of competitive advantage in a PSF (Empson 2001).

While sometimes organized as corporations with outside shareholders, the archetypal PSF is organized as a professional partnership.[7] This choice of organizational form is significant, in that a partnership has unique governance features that distinguish it from the corporate model of governance. Namely, in contrast to the corporation, with its clear separation of ownership and control, a partnership combines ownership, management and operations (Greenwood et al. 1990; von Nordenflycht 2007; Ribstein 2009). As a consequence, authority is shared by all owner-managers. Employee-owners are actively involved in formulating and executing strategy, while at the same time representing the organization's key strategic assets.

Professions, as occupational groups, have been a focus of sociological research, which has identified key traits that distinguish them from other occupational groups (von Nordenflycht 2010; Scott 1965). In general, professionals are characterized as having a high degree of bargaining power vis-à-vis employers, given both the scarcity and the transferability of their knowledge, skills, and abilities, combined with a strong preference for "autonomy and freedom from external constraint" in their work environment (Greenwood and Empson 2003, p. 916). They are typically engaged in work that involves solving complex and unique problems for individual clients, which makes bureaucratic approaches to monitoring and control—a key characteristic of large corporations and public organizations—infeasible.

[6] In some cases, such as hospitals, organizing as a nonprofit is a way to minimize commercially oriented governance (Hansmann 1996).

[7] Some self-regulating professions, such as law and accountancy, do not allow outside ownership; only licensed professionals may participate in the ownership of the firm.

Greenwood and Empson (2003) provide a number of arguments to support their assertion that the internal agency costs in a professional partnership or a PSF are, in general, much lower than the external agency costs associated with the separation of ownership and control in the modern firm: First, partners are more knowledgeable about the business of the firm than are investors in public corporations, enabling them to monitor more efficiently the behavior of their agents. Second, the proximity of partners to managers provides opportunities to exercise influence in a way not available to more dispersed shareholders. Third, managers are likely aware of the scrutiny of their colleagues.

Thus, the professional partnership avoids the agency problems arising from information asymmetries that plague the relationship between managers and shareholders of the firm. An additional benefit of the partnership identified by Ribstein (2009) is its flow-through entity status. Namely, earnings must legally be distributed to partners, thereby preempting the agency problem associated with control over free cash flows that increases agency costs in firms (Easterbrook 1984; Jensen 1986; Kochhar 1996; Lehn and Poulsen 1989).

Also unique to the professional partnership is the influence of personnel policy on organizational finance. Becoming a partner in most professional service firms is highly sought after by junior staff, given the financial rewards, the social prestige, and the ability to actively participate in decision-making that it entails (Greenwood et al. 2007; Greenwood and Empson 2003; von Nordenflycht 2007; Ribstein 2009). The 'up-or-out' system of career advancement, which is often utilized in professional service firms, exploits the 'tournament system' of motivation (Becker and Huselid 1992). This approach to career advancement supports the socialization that underpins many of the important norms that distinguish the profession as an institution and mode of governance.

In a discussion of incentives in the public university context, research on the interaction between motive dispositions and forms of incentives must be mentioned. Employment in the public sector has traditionally been associated with job security, career advancement via a merit system and salary protection based on a system of collective labor bargaining (Chen 2012). While an elaborate system of rules and formal procedures is seen as decreasing managerial flexibility and autonomy in the public sector (Chen 2012), which may be viewed as a drawback, at the same time public sector managers have been shielded from the market pressures and individual accountability for organizational performance that is faced by managers in the private sector. Public universities and, specifically, professors have been no exception.

The recent introduction of performance-based management techniques adopted from the corporate model of governance was intended to change this bureaucratic culture by stimulating professors to think and act more like private sector managers. A recent meta-analysis by Weibel et al. (2010), who researched relationship between individual performance and performance-related pay in the public sector, found that intrinsic motivation was greatly reduced by the use of performance-based pay. These authors analyzed the results of 46 experimental studies and concluded that intrinsic motivation accounts for much greater performance by public sector managers than does extrinsic motivation (specifically, the promise of a financial reward). Relatedly, a 1982 survey of US federal government and private sector 'middle managers'[8] revealed clear differences in reward preferences between the two groups. Specifically, the government managers rated the importance of financial reward as a career goal lower than did their private sector counterparts and rated the importance of "helping other people" and doing work that is "worthwhile to society" higher than private sector managers (Rainey 1982, p. 290). The work of Weibel, Rost and Osterloh tends to corroborate Rainey's findings.

This leads to a further criticism of the use of corporate governance tools like performance-based pay by those who emphasize the social role played by public sector organizations. These critics argue that financial goals are frequently at odds with social values, such as equity and access, which public service organizations should prioritize. This argument is particularly salient for such merit goods as education and health care. The normative disagreement about the role played by public sector organizations in society has galvanized political and philosophical opposition to the adoption of the corporate model, but with a notable lack of viable alternatives on offer. Given that professors can be described as knowledge workers, with all of the associated characteristics, and as intrinsically motivated public sector workers, professionalization is an incentive-compatible mode of governance that should promote both the enhanced performance of professors and the broader social interests served by education and basic research. Indeed, both theory and evidence suggest that current experiments with the corporate model of performance-based compensation will not reap the expected benefits and may even reduce motivation and performance.

Finally, research suggests that the unique governance features associated with the professional partnership and the PSF are also associated with organizational performance that is superior to the performance of the corporation. An empirical study of management consultancies by Greenwood et al. (2007) found that partnerships and privately held firms outperformed publicly-traded firms in the consulting industry, regardless of the level of organizational complexity or geographical scope of operations. The authors conclude that organizations that are managed and controlled by owners perform on average better than those organizations that are characterized by a separation of ownership and management. Indeed, the fact that owners place their own capital at risk is considered to be a 'signal' of quality and integrity to clients and other stakeholders (Van Lent 1999, p. 240).

In summary, the professional partnership combines the unique and complementary governance features of the profession as institution and the partnership as organizational form, offering the following benefits. Firstly, they are more consistent and compatible with the characteristics of knowledge work and knowledge

[8] "A 'middle manager' was defined as a person below the level of vice-president or assistant agency director, with at least one level of supervision below him or her" (Rainey 1982, p. 292).

workers than are either the corporate model or the public bureaucracy model of governance. Secondly, it reflects the lack of clear agency relationships and treats knowledge workers as experts who collegially monitor each other's work via a system of peer mutual monitoring. This is consistent with the credence goods nature of work conducted particularly in the university setting, which does not lend itself to monitoring and evaluation by non-expert managers. Thirdly, the collegial work environment, autonomy in working and participation in decision-making are more consistent with the intrinsic motivation that characterizes knowledge workers and sheds a dubious light on the use of corporate-style performance-based rewards, which research has shown may actually decrease work performance and motivation. Lastly, and relatedly, it reflects research findings that indicate higher levels of work motivation in knowledge workers working in a knowledge-based organization as opposed to a mixed or non-professional, hierarchical organizational setting. This may also explain, at least in part, the superior performance of PSFs.

In applying the PSF governance model to universities there is, however, one obstacle that must be taken into account. While private PSFs operate in markets in which market prices (ideally) provide signals of both opportunity cost in producing services and the willingness to pay of costumers or clients, comparable prices do not exist in the university context. The underlying reason is that universities, in engaging in science, contribute to a public good that will not be provided in private markets due to market failure.[9] Private PSFs, in contrast, provide knowledge-intensive services that are marketable, as these services are essentially private in nature (e.g., medical treatment or client-specific legal advice). This difference, however, does not preclude the PSF as a role model for university *governance*. Rather, it implies that universities need a different source of *funding*. While PSFs can fund themselves by marketing their services, universities rely on public funds. A further reason why the PSF model cannot be transferred one-to-one to universities concerns ownership. To the extent that universities are public, separation of ownership and control is a virtually unavoidable prerequisite of university governance. Nevertheless, this does not preclude the PSF model as an alternative governance approach. What matters for PSF incentives is not ownership in a literal sense; rather, it is the distribution of residual rights of control typically associated with private ownership (Hart et al. 1997; Wigger and von Weizsäcker 2000; Wigger 2004). In applying the partnership model to public universities the question is thus, to what extent can residual rights of control be attributed to professors as partners?

[9] See, e.g., Nelson (1959, 2004), Arrow (1962) and Dasgupta and David (1994) on market failure in science. The lack of market prices may also explain the widespread use of research rankings in academia. Like market prices they are easy to gather and one-dimensional in nature. In contrast to market prices, they do not reflect costs of research and it is a matter of debate to what extent they reflect benefits for society. In addition, research rankings may invoke detrimental incentives, as Osterloh and Frey (2014) point out.

Conclusion
A chief aim of our research is to stimulate reformers to think beyond the limited scope of the corporate model of governance in the discourse on public university reform. In this vein, questions were posed, such as "Does it make sense to run a university like a firm?" and "Are there alternative private-sector models of management and governance more appropriate and effective for the university context than the corporate model?", questions to which we sought answers. To this end, the chief contribution of our work can be summarized as follows. Starting at the broader, abstract level of knowledge work and knowledge workers, arguments drawn from transaction cost theory were made that lay the foundation for two alternative, yet complementary, modes of governance borrowed from the private sector: the broader professionalization of knowledge workers and the partnership as an organizational form. Namely, the argument was made that partnerships contribute to lower transaction costs than can be achieved by either markets or hierarchies in the governance of knowledge-intensive service transactions for a number of reasons. Firstly, lower information asymmetries between expert-partners reduce the probability that opportunistic behavior can be successfully pursued. Secondly, the unification of ownership, management and operations in the persons of the partners contributes to a highly effective system of internal mutual peer monitoring.

Acknowledgements We are grateful for financial support provided by the German Federal Ministry of Education and Research (BMBF grant no. 01PW11015). All opinions expressed herein are those of the authors.

References

Alvesson M (1993) Organizations as rhetoric: knowledge-intensive firms and the struggle with ambiguity. J Manag Stud 30(6):997–1015

Alvesson M (2000) Social identity and the problem of loyalty in knowledge-intensive companies. J Manag Stud 37(8):1101–1123

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

Becker BE, Huselid MA (1992) The incentive effects of tournament compensation systems. Adm Sci Q 37(2):336–350

Bucher R, Strauss A (1961) Professions in process. Am J Sociol 66(4):325–334

Burns T, Stalker GM (1961) The management of innovation. Tavistock, London

Chen C (2012) Explaining the difference of work attitudes between public and nonprofit managers: the views of rule constraints and motivation styles. Am Rev Public Admin 42(4):437–460

Dasgupta P, David PA (1994) Toward a new economics of science. Res Policy 23:487–521

DiMaggio PJ, Powell WW (1983) The iron cage revisited: institutional isomorphism and collective rationality in organizational fields. Am Sociol Rev 48(2):147–160

Dulleck U, Kerschbamer R (2006) On doctors, mechanics, and computer specialists: the economics of credence goods. J Econ Lit 44(1):5–42
Easterbrook FH (1984) Two agency-cost explanations of dividends. Am Econ Rev 74(4):650–759
Emons W (1997) Credence goods and fraudulent experts. Rand J Econ 28(1):107–119
Empson L (2001) Fear of exploitation and fear of contamination: impediments to knowledge transfer in mergers between professional service firms. Hum Relat 54(7):839–862
Fama E, Jensen M (1983) Separation of ownership and control. J Law Econ 26(2):301–325
Foss NJ (2007) The emerging knowledge governance approach: challenges and characteristics. Organization 14(1):29–52
Goode WJ (1957) Community within community: the professions. Am Sociol Rev 22(2):194–200
Greenwood R, Empson L (2003) The professional partnership: relic or exemplary form of governance? Organ Stud 24(6):909–933
Greenwood R, Hinings CR, Brown J (1990) "P2-Form" strategic management: corporate practices in professional partnerships. Acad Manag J 33(4):725–755
Greenwood R, Deephouse DL, Li SX (2007) Ownership and performance of professional service firms. Organ Stud 28(02):219–238
Hackett B (2000) Beyond knowledge management: new ways to work and learn. Research Report 1262-00-RR. The Conference Board, New York
Hall RH (1968) Professionalization and bureaucratization. Am Sociol Rev 33(1):92–104
Hansmann H (1996) The ownership of enterprise. Belknap Press of Harvard University Press, Cambridge
Hart O, Shleifer A, Vishny RW (1997) The proper scope of government: theory and an application to prisons. Q J Econ 112:1127–1161
Jensen MC (1986) Agency costs of free cash flow, corporate finance, and takeovers. Am Econ Rev 76(2):323–329
Kandel E, Lazear EP (1992) Peer pressure and partnerships. J Polit Econ 100(4):801–817
Kochhar R (1996) Explaining firm capital structure: the role of agency theory vs. transaction cost economics. Strateg Manag J 17:713–728
Lehn K, Poulsen A (1989) Free cash flow and stockholder gains in going private transactions. J Financ 44(3):771–787
Meyer JW, Rowan B (1977) Institutionalized organizations: formal structure as myth and ceremony. Am J Sociol 83(2):340–363
Moe TM (1995) The politics of structural choice: toward a theory of public bureaucracy. In: Williamson OE (ed) Organization theory: from Chester Barnard to the present and beyond. Oxford University Press, Oxford, pp 116–153
Nelson RR (1959) The simple economics of basic scientific research. J Polit Econ 67(3):297–306
Nelson RR (2004) The market economy, and the scientific commons. Res Policy 33:455–471
Osterloh M, Frey BS (2014) Tanking games. Eval Rev
Osterloh M, Kieser A (2015) Double-blind peer review; how to slaughter a sacred cow. In: Welpe IM, Wollersheim J, Ringelhan S, Osterloh M (eds) Incentives and performance – governance of research organizations. Springer, Cham
Ouchi WG (1979) A conceptual framework for the design of organizational control mechanisms. Manag Sci 25(9):833–848
Parsons T (1939) The professions and social structure. Soc Force 17(4):457–467
Powell WW (2007) The new institutionalism, Stanford. http://www.stanford.edu/group/song/papers/NewInstitutionalism.pdf. Accessed 20 Mar 2014
Rainey HG (1982) Reward preferences among public and private managers: in search of the service ethic. Am Rev Pub Admin 16(4):288–302
Reed MI (1996) Expert power and control in late modernity: an empirical review and theoretical synthesis. Organ Stud 17(4):573–597
Ribstein LE (2009) Partnership governance of large firms. Univ Chicago Law Rev 76:289–309
Scott WR (1965) Reactions to supervision in a heteronomous professionalorganization. Adm Sci Q 10(1):65–81

Scott WR (1995) Institutions and organizations. Sage, Thousand Oaks

Starbuck WH (1992) Learning by knowledge intensive firms. J Manag Stud 29(6):713–740

Sveiby K-E, Simons R (2002) Collaborative climate and effectiveness of knowledge work – an empirical study. J Knowl Manag 6(5):420–433

Van Lent L (1999) The economics of an audit firm: the benefits of partnership governance. Br Account Rev 31(2):225–254

von Nordenflycht A (2007) Is public ownership bad for professional Service firms? Ad agency ownership, performance, and creativity. Acad Manag J 50(2):429–445

von Nordenflycht A (2010) What is a professional service firm? Toward a theory and taxonomy of knowledge-intensive firms. Acad Manag Rev 35(1):155–174

Weibel A, Rost K, Osterloh M (2010) Pay for performance in the public sector – benefits and (hidden) costs. J Public Adm Res Theory 20(2):387–412

Wigger BU (2004) Public versus private production of government services. Zeitschrift für Betriebswirtschaftslehre 74(1):53–67

Wigger BU, von Weizsäcker RK (2000) Öffentliche oder private Umsetzung des staatlichen Bildungsauftrags? – Eine Analyse auf der Grundlage der Theorie unvollständiger Verträge. In: von Weizsäcker RK (ed) Schul- und Hochschulorganisation. Duncker & Humblot, Berlin

Wilensky HL (1964) The professionalization of everyone? Am J Sociol 70(2):137–158

Incentivizing Innovation in Knowledge-Intensive Companies: Conceptual Analysis of the Fit Between Reward Programs and Organizational Contexts

Sven Grzebeta

Abstract In this conceptual study, I analyze the fit between corporate programs incentivizing innovation and knowledge-intensive organizational contexts. Based on the personal characteristics and contextual factors determining creative work behaviour the levers for rewarding innovation in knowledge-intensive business contexts are discussed. Drawing on fit theory, I corroborate the hypothesis that a "one size fits all" approach for programs incentivizing innovation does not fully leverage the creative potential of organizations. I analyze the dimensions and implications of the fit between incentive programs for innovation and knowledge-intensive organizational contexts and conclude with selected practical recommendations to improve this fit.

1 Introduction

For the past 30 years, innovation has been a buzz word both in the theory and practice of business management. Whenever firms find themselves under increased pressure—be it from competition, regulation or social change—it is likely that at some stage there will be a call for more innovation. For knowledge-intensive companies, innovation is at the same time natural and of critical importance because enhancing their knowledge stock and finding new applications for it is part of their organizational DNA (Muller and Doloreux 2009; Starbuck 1992; Swart and Kinnie 2003). While crucial competitive advantages are to be gained from innovation (Kanter 1983; Sheehan and Stabell 2007), finding and implementing successful business innovations is a complex and elusive task (Ahmed and Shepherd 2010; Burgelman and Sayles 1986). Typically, the management board of a company will decide to implement an innovation program of some sort to increase the company's innovativeness. In order to ensure that management and staff contribute

S. Grzebeta (✉)
Offenbach, Germany
e-mail: sg@grzebeta.de

I.M. Welpe et al. (eds.), *Incentives and Performance*,
DOI 10.1007/978-3-319-09785-5_21

to such a program's success (or, at least, do not act against it) *incentives* are introduced. Such incentives may have a positive effect on people's behavior within the organization, and, if used strategically, they can in fact be a powerful governance tool (Aguinis 2013; Lerner and Wulf 2007). For incentives to be effective it is important, both for the researcher and the practitioner, to understand the interactions of incentives, creativity, co-operation and business innovation in knowledge-intensive companies thoroughly.

Starbuck (1992, p. 715) defines a knowledge-intensive firm as one in which "knowledge has more importance than other inputs". To make a knowledge-intensive firm special it must, to a significant extent, depend on "esoteric" and "exceptional expertise" (Starbuck 1992, p. 716). Most of the recent literature accepts West and Farr's (1990, p. 9) definition of the term "innovation" as "the intentional introduction and application within a role, group or organization of ideas, processes, products or procedures, new to the relevant unit of adoption, designed to significantly benefit the individual, the group, organization or wider society". For the purpose of this article, I follow these definitions. The ability to generate new ideas is usually identified with the personal characteristic of *creativity* (Ekvall 1971; Mumford and Gustafson 1988; Scott and Bruce 1994). An impressive body of research on applied creativity has been produced over the past three decades (Amabile 1983; Egan 2005; Mumford and Gustafson 1988; Zhou and Shalley 2003), accompanied by studies on how it is impacted by incentives (Aguinis et al. 2013; Becker 1987; Toubia 2006), on the relationship between productive and creative goals (Miron et al. 2004), and on the impact of teamwork and knowledge sharing on creativity (Sung and Choi 2012). There is general agreement in the literature that both individual and organizational factors influence the degree of creativity that an individual shows in a given situation (Amabile 1996; Oldham and Cummings 1996; Scott and Bruce 1994; Woodman et al. 1993). It is also generally acknowledged that goals and incentives are important antecedents to individual creativity (Farr and Ford 1990; Shalley 1995). Azoulay et al. (2011) found that in an academic research setting, rewarding longer-term rather than short-term success and giving researchers freedom to experiment increases both the frequency and novelty of published research findings.

However, when investigated empirically, existing models only explain up to 34 % (Oldham and Cummings 1996) and 37 % (Scott and Bruce 1994) of variance in innovative behavior. Thus, a large part of the observed phenomenon remains unaccounted for. Moreover, the interaction effects of personal characteristics, job-related tasks, incentives and knowledge-intensive organizational context have not yet been explored systematically. Although there is evidence that performance management and incentive schemes can have a positive effect on innovation in knowledge-intensive organizations (Azoulay et al. 2011; Lerner and Wulf 2007), the implications for the various organizational settings in the real world have not yet been fully explicated. This is particularly painful for the practitioner who is looking for guidance when designing a concrete incentive scheme for innovation in a given knowledge-intensive business context. From an academic perspective, it is desirable to develop models which explain an even larger part of the variance in observed innovative behavior.

In the worst case, the incentive scheme for new ideas is counterproductive as employees shift their focus from the timely delivery of projects to the creation of new ideas or as they get badly frustrated by seeing that their good ideas do not get support for implementation from the organization. Thus, the innovation program must be adequate to the organizational context in which it is placed (Becker 1987). However, the implications of the *fit* of an incentive scheme for innovation in knowledge-intensive companies *with its organizational context* have not been analyzed to date. Therefore, I undertake a conceptual analysis of the fit between corporate programs incentivizing creativity and knowledge-intensive organizational contexts. The key question is: What is the best approach to designing an effective incentive program for innovation in a knowledge-intensive business context? My hypothesis is that a "one size fits all" approach for programs incentivizing innovation is not adequate to systematically leverage the creative potential of knowledge-intensive companies. My approach draws on existing research and anecdotal evidence.

I start by exploring the relation of knowledge, ideas and innovation in knowledge-intensive companies. Next, I undertake an exposition of the levers for rewarding creativity and innovation in a knowledge-intensive corporate context. I then propose to analyze the problem using a systems approach to fit theory. Based on the research findings from the literature concerning incentives, creativity and innovation I analyze the dimensions and implications of the fit between an incentive program for innovation and its organizational context in knowledge-intensive companies. I conclude with selected practical recommendations to improve this fit in practice.

2 Generating Knowledge Through Ideas: Innovation in Knowledge-Intensive Companies

Although innovation can be critical for the survival of companies of all types and in all sectors (Christensen 1997) it is of particular importance for knowledge-intensive companies (Lei et al. 1999; Swart and Kinnie 2003). Knowledge-intensive firms must engage in a continuous process of learning (and unlearning) if they want to remain competitive since knowledge ages and may get outdated (Starbuck 1992; Lei et al. 1999). Up-to-date knowledge is the lifeblood of knowledge-intensive companies, and the application of this knowledge in ways which create value for their customers is a critical competency (Muller and Doloreux 2009; Sheehan and Stabell 2007). Such relevant knowledge is difficult to generate or to acquire and its management requires special skills and adequate organizational environments (Earl 1994; Lei et al. 1999; Swart and Kinnie 2003).

With respect to their competitively critical knowledge, firms may pursue an explorative or exploitative approach (Edvardsson 2008; Hansen et al. 1999). Knowledge-intensive companies may innovate in two ways specific to them: First, they may produce new insights by research and intelligence or even discover new domains of knowledge (Starbuck 1992). Secondly they may find new ways of applying their knowledge in commercial contexts (Sheehan and Stabell 2007). Although only the first approach is based on the explorative generation of new knowledge in the strict sense, identifying new opportunities to exploit existing knowledge also requires, somewhat paradoxically, exploration. Innovation and knowledge have a relationship of mutual reinforcement because fruitful research and the generation of new useful ideas require domain expertise (Amabile 1983), and in turn, new ideas generate new knowledge. However, it should be noted that the cumulative extension of knowledge, e.g., a lawyer reading and understanding new pieces of legislation, is not necessarily innovation. Rather, adding and refreshing knowledge in a linear way is routine business for knowledge-intensive companies.

As innovation requires new ways of thinking, the phenomenon of creativity has received great attention in organizational psychology (Amabile 1983, 1996; Mumford and Gustafson 1988). But creative ideas must be translated into tangible improvements in order to be effective (Miron et al. 2004; West 2002). For example, if a programmer in an IT company has an idea for a new software product, the realization of this idea is likely to require organizational support in terms of management decisions, resources and co-operation with colleagues. West (2002) even goes so far as to claim that the implementation of ideas, regarded as a team exercise, is much more critical for successful innovation than the generation of ideas. Therefore, idea generation and implementation of the solution constitute two halves which together make up innovation.[1]

It is important to differentiate between acquired knowledge, explicit ideas as to how to apply this knowledge and its successful application. Moreover, some innovations may be implemented by the idea originators themselves and others require the support of other employees or management, or the provision of resources by others.[2] Furthermore, ideas can be applicable to one's own job or work group (Axtell et al. 2000) or to improvements anywhere in the company (Van Djik and Van den Ende 2002). Finally, innovation may take place implicitly in the routine of one's daily work (Janssen 2000), or it may be formulated in an explicit idea (Van Djik and Van den Ende 2002). The latter approach is the basic principle underlying all types of employee suggestion schemes (Ekvall 1971) and a good practical way to capture creativity.

---

[1] Some authors have proposed more sophisticated phase models: idea extraction, idea landing and idea follow-up (Van Djik and Van den Ende 2002); idea generation, idea acceptance by coalition building, idea implementation, and transfer and diffusion (Kanter 1988).

[2] For a systematic typology of creative contributions cf. Unsworth (2001).

3 Fostering Innovation Through Incentives in Knowledge-Intensive Companies

If a knowledge-intensive company wants to become more innovative, it needs to ensure that both the generation of valuable ideas and their successful application are supported as much as possible (Kanter 1988). An organization may achieve this by adapting its culture accordingly (Ahmed 1998) and by promoting innovation explicitly with corporate programs (Ahmed and Shepherd 2010; Burgelman and Sayles 1986). Such programs work on the level of the organization (e.g., by defining strategy and setting the corporation's agenda) and on the level of the individual member (e.g., by goal setting or skill development). Three types of recommendations can be found in the literature: (1) Changes to the organizational structure to remove administrative barriers, increase departmental autonomy and accelerate decision-making processes (Lei et al. 1999); (2) Changes in management behavior (e.g., by goal setting, reward schemes, or alternative decision-making procedures) to promote exploration and rational risk-taking (Ahmed 1998; Azoulay et al. 2011; Gumusluoglu and Ilsev 2009); and (3) Fostering knowledge-sharing and open communication to support the diffusion of knowledge and the generation of breakthrough discoveries or ideas through cross-functional collaboration (Edvardsson 2008; Scarbrough 2003; Swart and Kinnie 2003).

Organizational structure, culture and leadership have implications far beyond a company's capability to innovate. Although an organizational context supportive of innovation may be the most effective long-term approach, changing an organization's cultural characteristics is a difficult and, in practice, slow process (Farr and Ford 1990; Lei et al. 1999). Therefore, it appears to be relevant to investigate alternative approaches to foster innovation. One alternative approach is represented by programs that aim at increasing a company's innovativeness by offering rewards[3] to the employees. Such programs can be implemented in a short period of time and can be expected to have a positive effect on the staff's behavior without much delay. However, one should consider that creativity can only partly be enhanced by means of incentives: specifically, whereas task motivation might be influenced by incentives in the short- to medium-term (Aguinis et al. 2013), the other personal characteristics which are required to produce creative behavior, namely domain expertise and creativity skills (Amabile 1983) cannot be influenced by incentives within a short timeframe (Amabile 1996). In particular with a view to

[3] Davila et al. (2006) differentiate between *incentives* and *recognition rewards*. They define incentives as contractually arranged links between performance and subsequent rewards, and recognition as voluntary and subjectively determined rewards which are not announced in advance. However, if recognition rewards are to have any motivational effect their existence must either be known or at least be reasonably expectable by the affected persons. Therefore, I use the terms "incentive" and "reward" more or less interchangeably in this article. "Reward" has a stronger focus on the actual presentation of the benefit, whereas "incentive" highlights the motivating effect of a contingent reward. By "recognition" I refer to the acknowledgement (verbal or other) by the organization of an achievement or a contribution.

the key role of highly specialized domain expertise in knowledge-intensive companies (Starbuck 1992) this is an important limitation. Another challenge is that highly-specialized experts may even show resistance to new ideas if these reduce the value of the specialists' current expertise or put their professional privileges at risk (Earl 1994; Starbuck 1992). A successful incentive program for innovation in a knowledge-intensive organization needs to reflect these constraints.

Becker (1987) provides a framework for the design of incentive systems for innovation. The strength of Becker's approach is that it takes a holistic view of the incentives working towards or against innovation in an organization. However, his conceptual framework is geared to traditional hierarchical models of the firm and he deals with incentives for innovative behavior only on the basis of conventional approaches to pay (base and general performance-related components). Hence he suggests different types of incentive schemes for managers and staff, in which only the former are granted performance-related pay, and he measures contributions to innovation on hierarchical entities such as work groups and business units, rather than on the level of informal teams or simply those individuals who made the innovation happen, regardless of rank or organizational position. Therefore, although his framework is still valid as an analytical tool, it needs to be updated to reflect recent approaches to innovation management (Ahmed and Shepherd 2010) and the current working environments in knowledge-intensive organizations (Edvardsson 2008).

Based on principal-agent theory and expected probability considerations, Toubia (2006) and Manso (2011) develop output-maximizing incentive models for explorative behavior which take the strategic considerations of actors into account. Although Toubia's insights may help to design adequate incentives in group creativity processes, and guidelines for rewarding executives may be formulated based on Manso's conclusions, one cannot derive parameters for a concrete incentive program directly from their theoretical models. Moreover, both approaches implicitly assume that managers and employees are rational economic agents who are mainly motivated extrinsically.

Employee expectations in knowledge-intensive organizations tend to differ from those prevalent in traditional production or service environments (Edvardsson 2008; Starbuck 1992). In particular, many knowledge workers value self-efficacy and opportunities for professional and personal growth at least as much as monetary rewards (Edvardsson 2008). Ekvall (1971) notes that motives such as intellectual stimulation and making a contribution to the company's success are, by and large, equally important to idea originators as monetary incentives. Given such motives in employees, developing and providing creative ideas is attractive by and in itself, even without extrinsic motivational factors. In a similar vein, Becker (1987) regards the enhancement of one's skills and the appreciation by competent colleagues as motivators which are effective without any incentives set by management. He differentiates between incentives for innovation "in a narrow sense", by which he refers to explicit rewards, and incentives "in a broader sense", which result from strategy definition and implementation, the organizational setup and personnel development programs. Non-monetary rewards include recognition of the employee's contribution and visibility of the employee (within the company or even to the general public), and awards such as certificates, trophies or privileges, and promotions (Van Djik and Van den Ende 2002). But as innovation in corporate contexts has a clear commercial goal, employees may show even higher motivation when offered financial rewards, which, if applied in the right way, may improve employees' performance (Aguinis 2013). Moreover, experiments and case studies show that creative output can be increased with adequate performance-based rewards (Davila et al. 2006; Toubia 2006). A major advantage of financial rewards is that they may scale with the *value* of the contribution to the organization.

However, existing research also highlights the risks and downsides of monetary rewards. Generally, extrinsic motivators may undermine intrinsic motivation (Amabile 2000). Amabile et al. (1986) show that *contracted-for* rewards can reduce the creativity that people show when executing a task. However, according to the same study, this is much less the case for *performance-contingent* rewards. Hence, incentive schemes should offer rewards according to the value added by the contribution, not just for executing a certain task such as submitting an idea. It should be noted that any program will in practice reward a combination of creative and productive performance because putting forward, promoting and implementing ideas are not only creative but also productive activities.

Based on the insights by Axtell et al. (2000), setting the parameters of an incentive program should be relatively straightforward. According to their analysis, individual level characteristics are the main determinants of idea generation, whereas factors on the organizational and team level can best explain variations in the implementation likelihood. One would simply need to understand the main and interaction effects of rewards on idea generation and implementation and optimize the input variables with a view to maximizing the expected value of implemented ideas. However, this approach has two complications. First, whereas it is quite straightforward to monitor and assess the production of creative ideas, it appears much more difficult to evaluate someone's contribution to their implementation. Often the desired behavior is not even pro-active support of an idea, but the mere absence of defensive or obstructive behavior of managers or employees who resist the change implied by the idea (Becker 1987; Miron et al. 2004). Therefore, reward programs should also aim at encouraging the support of ideas, no matter where they come from.

Second, Axtell et al.'s conclusion that management support is not a critical contributor to idea generation is counter-intuitive. Employees in knowledge-intensive organizations tend to appreciate self-direction (Edvardsson 2008), and anecdotal evidence points to the fact that they care a lot about the support of their ideas by the organization. Given a person's characteristics and preferences, the motivation to forward ideas depends on the expected payoff that the idea may bring to the originator in terms of self-efficacy, benefits from the proposed change, recognition, or reward (Amabile 2000; Farr and Ford 1990). If an employee has reason to believe that an idea will likely be rejected or not supported by the organization, his or her expected payoff from this potential innovation will be very low and the employee is unlikely to put much effort into developing and promoting it. Axtell et al. look at the impact of personal characteristics and organizational context factors on the implementation of *one's own* ideas. They find that in this case, contextual factors do not have a great effect on the likelihood of implementation, which is an important contribution to the academic understanding of innovative work behavior. If, however, in a broader innovation approach in knowledge-intensive firms the originator and the people promoting and implementing the idea are different persons, environmental factors may be determinants of creative work behavior. Vice versa, and intuitively, individual and job-level characteristics such as motivation to explore, breadth and depth of domain-relevant skill and the ability and willingness to help others with the implementation of their ideas should have a significant impact on the implementation likelihood of ideas.

4 Fit Between Incentive Programs to Foster Innovation and Organizational Context

Because each organization has its specific characteristics with regard to encouraging or inhibiting innovative work behavior (Scott and Bruce 1994; Oldham and Cummings 1996), incentive schemes for innovation may have a better or worse fit to their organizational context. This fit is important for three reasons: First, the organizational contexts of knowledge-intensive companies tend to be of a peculiar kind (Earl 1994; Starbuck 1992) and an effective incentive scheme for innovation needs to reflect these peculiarities. Second, a program incentivizing creativity must be designed in a way that reflects the organization's strategy, in particular with regard to exploration and exploitation of knowledge, and its desire and capabilities to implement ideas. Ideas which do not even have a chance of being implemented are not only without value, but they will also lead to serious frustration of the employees who put them forward. On the other hand, if a program leads to the implementation of the *wrong* ideas, it may even harm the organization economically. A third reason is that the program should be tuned so as to increase the likelihood of a good idea or insight to be applied within the organization.

Typically, the number and quality of implemented innovations are regarded as the main dependent variable in models of individual creativity in work contexts (Janssen 2000; Oldham and Cummings 1996). But the system-based concept of fit (Drazin and Van de Ven 1985) takes a broader view on the success criteria of organizational interventions. For instance, the acceptance of an innovation program by management and staff, realistic expectations of the potential rewards by employees, and the organizational support for ideas are not only antecedents to successful innovation, but at the same time these factors should be considered as *affected by* the program. Vice versa, the effects from an incentive scheme feed back through their medium-term influence on corporate culture, e.g., if successful applications of new ideas positively affect management's attitude towards employee creativity or knowledge-sharing. Moreover, such a program may have unintended consequences or side-effects, to which a simplistic input–output-model might be insensitive. Based on their finding that innovation and productivity have a "dialectic" relationship, Bledow et al. (2009, p. 319) come to the following conclusion: "As innovation has emerged from contradictory organizational structures and cultures, 'one-best-way' recommendations for organizational innovation that do not take into account the particularities of a given organization are misguided and may even do more harm than add value." According to fit theory, it is impossible to design a good incentive program for innovation by deriving each parameter separately from an environment which is taken as independent and static.

I will now turn to the relevant dimensions of the fit between a reward scheme for innovation and its organizational context in a knowledge-intensive company.

Policies and procedures As an incentive system is an additional organizational policy, its inter-relation with existing policies and procedures must be taken into account. In a bureaucratic organization with rigid procedures, where strict adherence to detailed rules is expected (as this may be vital to the organization's performance), strong incentives to change the status quo may be misdirected since creativity and innovation tend to require deviating from existing practices, routines or even rules (Olin and Wickenberg 2001). On the other hand, if such an organization wants or needs change, particular care must be taken to offer incentives that can be expected to be effective in the right way within the given organizational context. Moreover, existing policies and procedures put constraints on the innovation process (Azoulay et al. 2011; Lei et al. 1999) and conversely, an incentive system may affect the agents' compliance with the policies.

Performance management and pay The monetary and non-monetary rewards offered by the program must be considered in relation to the compensation and benefits policies of an organization as well as its socio-demographic characteristics and the employees' expectations. The latter are of particular importance with regard to perceived effort-reward fairness (Ekvall 1971; Janssen 2000; Aguinis et al. 2013). As the rewards granted in the context of an incentive scheme for innovation are contingent upon certain forms of behavior and, most likely, additional success criteria, they constitute a form of performance management (Aguinis 2013) and should be at least compatible with existing performance management schemes.

Job-related tasks The relationship of the desired employee behavior with the existing work roles and job-related tasks must be considered. The implications can be both ways: On the one hand, employees must have sufficient "space and time" for creative thinking and the development of innovative ideas, on the other hand, there may be employees in knowledge-intensive organizations, e.g., in R&D units, who are already paid a salary or a bonus for the continuous development of new ideas, as this is their main job (Lerner and Wulf 2007). In this case additional incentives may be set to promote exploration and experiment and to encourage research in as yet unknown or remote areas (Manso 2011; Azoulay et al. 2011).

Knowledge management and team work While knowledge sharing is generally regarded as a determinant of innovation in knowledge-intensive contexts

(Swart and Kinnie 2003), knowledge-intensive companies may follow different strategies with regard to the generation, sharing and exploitation of knowledge (Edvardsson 2008; Scarbrough 2003). The prevalent forms of team work determine how knowledge is generated, shared and utilized practically in a particular organization (Sung and Choi 2012). An incentive scheme should leverage these social structures and foster practical ways of knowledge management which are compatible with the company's strategy (Earl 1994).

Practically, I suggest that the following four aspects should be considered when defining a scheme to incentivize innovation:

Motivational effect The fit between the program and the organizational context is bound to have an effect on the motivation of employees to be creative and to apply or put forward their ideas (Becker 1987; Ekvall 1971; Miron et al. 2004). While the immediate goal is to maximize the number of insights or ideas produced and forwarded by members of the organization, the value for the organization depends on the quality of the ideas.

Idea quality Therefore, the interaction effects of the program and the organizational environment on the adequacy of ideas with regard to the organization's goal are important (Ahmed and Shepherd 2010; Davila et al. 2006). For a knowledge-intensive organization, it should be clear whether the strategic focus is on generating new relevant knowledge or on new commercial applications of existing knowledge (Edvardsson 2008; Hansen et al. 1999).

Implementation likelihood The program needs to fit the organizational environment in such a way that it maximizes the implementation likelihood of good ideas. As the implementation of ideas often requires resources, the choice of the right ideas is important, and ideas should be prioritized by their expected value for the company (Burgelman and Sayles 1986). So again, this is not only a question of quantity, but also of quality and selection.

Side effects Any adverse side effects or unintended consequences of a program within the given organizational context should be foreseen and avoided, especially with a view on the program's potential to foster unethical behavior (Aguinis et al. 2013; Baucus et al. 2008) and its effect on the acquisition, sharing and diffusion of knowledge (Scarbrough 2003; Starbuck 1992; Swart and Kinnie 2003).

If there is a good fit of the incentive scheme for innovation and the organization the incentive scheme will be more effective compared to a bad fit. A good fit can be recognized by a general acceptance of the scheme, a sufficient number of insights produced and ideas suggested, an efficient and smooth evaluation and selection process, and the timely implementation of all valuable ideas. In case of a bad fit the program may have no effect at all, or, worse, unintended and undesired consequences, such as employees or managers developing an attitude of resistance against new ideas and innovation (Earl 1994; Starbuck 1992). A bad fit shows itself in a low number of ideas proposed, reluctant take-up and evaluation by the responsible experts and a low implementation ratio with long lead times. Based on the systems approach to fit (Drazin and Van de Ven 1985), the incentive scheme can be expected to be effective as an intervention even if it has a bad fit. Members of the organization will react to it in one way or another, and if the scheme does not have a good fit, these reactions, although rational and adequate from each actor's individual perspective, may be harmful to the organization as a whole.

5 Improving the Fit: Some Practical Aspects

Given these complex structural relations between the dimensions of an incentive scheme for innovation and aspects of the organizational context, what are the levers to achieve a good fit between the incentive program and the organizational context in a knowledge-incentive company? I propose four practical principles to guide the design and implementation of such a program.

Define the goal of the program First, it is important to clarify and formulate the goal of the innovation program and to ensure that it is aligned with the overall strategy of the organization (Davila et al. 2006), in particular with its strategic approach to knowledge management (Hansen et al. 1999). For instance, if the pronounced strategy of an organization is to grow regionally while focusing on core competencies and activities, then a program to create product ideas "out of the box" would be a mismatch. The goal of the program has two functions: It makes explicit what exactly the organization wants to achieve with the program and it gives employees guidance as to what kind of knowledge and ideas to generate and to spend time on. Ideally, the goal serves as a map for the organization's "innovation radar" so that its members have an indication of the areas which may be worthwhile exploring.

Check goal against organizational reality Second, the need for any incentivization (beyond existing compensation and benefits practices) should be derived from the gap between the innovation requirements and the organizational reality in terms of existing explorative and innovative efforts. Moreover, socio-demographic aspects of the workforce as well as product and market particularities of the organization should be considered. The following examples are among the practical questions to answer: Is any additional incentive required at all or do employees show sufficient intrinsic motivation to address the challenges creatively? What are employees' and managers preferences, e.g., with regard to risk? In particular, it is important not to kill existing intrinsic motivation by introducing external incentives which may be perceived as contracted-for rewards (Amabile et al. 1986; Amabile 2000). Are any additional organizational measures required to ensure that people have the ability to do what is asked from them (skills, time, resources...)? If such resources are missing, incentives will unlikely have the desired effect (Farr and Ford 1990).

It should also be checked if the organization is able to achieve the defined goal at all. How much support does the program have from top, middle and lower management? Are there any stakeholders in the organization who would tend to obstruct such changes? If so, can they be incentivized so that their expected payoff makes them support rather than oppose new ideas (Farr and Ford 1990)? One consequence of these analyses may be that an incentive program is the wrong approach and that, for instance, a new organizational structure or the change of key personnel is required (Earl 1994; Lei et al. 1999). If the goal of the program is incompatible with corporate reality, it has a bad fit and is bound to fail.

Align incentive program with existing policies and procedures Once the requirements for the incentive program have been identified based on the gap between the goals and the status quo, the levers of the incentive scheme can be set and calibrated. The two main points here are to offer rewards that are effective, in the given organizational situation, as incentives to explore, and to align the characteristics of the scheme with existing policies and procedures, especially in the compensation and benefits area. Both the expected performance and the rewards need to be well-defined (Aguinis et al. 2013). Which behavior exactly will be rewarded? Is it the generation or codification of knowledge, the submission of an idea which meets certain criteria, or only the commercially successful application of knowledge? If rewards are attached to the attainment of goals, these should be adequate to the challenge at hand. Davila et al. (2006) suggest using realistic goals to foster incremental innovation, but stretch goals when radical innovation is needed. The level of additional pay offered must be defined in relation to existing levels of compensation. It is unlikely that offering an extraordinarily high reward for an idea, which exceeds normal annual bonuses by an order of magnitude, will work any more efficiently than a reward which is in line with existing bonus practices (Van Djik and Van den Ende 2002). Rather, such blatantly inadequate offers may turn out to be harmful as people set their mind on the jackpot rather than the generation of relevant knowledge or ideas (Scarbrough 2003). Also, free-riding or even fraudulent behavior may be encouraged inadvertently (Aguinis et al. 2013; Baucus et al. 2008; Toubia 2006).

Define fair and transparent rules Also, the calculation basis on which monetary rewards are determined needs to be specified. In particular, when offering financial rewards their determination must be well-defined and transparent so that participants regard the scheme as fair (Aguinis et al. 2013; Janssen 2000). Basically, any figure available for financial controlling purposes can be used as the basis of calculation, e.g., incremental sales, additional earnings (e.g., EBIT) or a project's estimated net present value (NPV). Also, qualitative characteristics of a contribution such as its degree of novelty or whether it can be turned into a patent can be considered. In commercial contexts, employees are likely to compare their payoff to their contribution's utility for the company. If a new product based on an employee suggestion becomes profitable, the employee should feel that the company shares a fair part of the gain. Moreover, rewards should be given promptly so that temporal contiguity may support a positive feedback effect, thus increasing the employees' motivation to show further creative behavior. In the context of innovation this is often difficult as it may take a while until an idea has been put to practice, and it may take an even longer time until it has become clear to what extent it is successful (Farr and Ford 1990).

Non-monetary rewards also need careful consideration with regard to organizational fit. They should be designed in such a way that employees perceive them as appreciative and honest tokens of gratitude from the organization. If it is well known that the CEO does not at all enjoy social events with staff, it is not a good idea to offer a dinner party with the CEO as the prize of an idea competition. Conversely, if the CEO has lunch with varying members of staff every day in the cafeteria, a lunch date will not be regarded as something special. The cultural fit is critical here because while an employee-of-the-month scheme may be adequate for a fast-food restaurant, it is clearly not so for a management consulting firm (Edvardsson 2007; Starbuck 1992).

No matter which combination of monetary and non-monetary rewards is chosen, one form of recognition is vital to any kind of innovation program, namely appreciative feedback to the idea originator. With such feedback, the organization shows that it cares about the idea and that it recognizes the extra effort undertaken by the originator, especially if the idea is ultimately rejected (Van Djik and Van den Ende 2002). In addition, qualified feedback from supervisors or experts may help the originator to produce better ideas in the future (Manso 2011).

Many approaches and creative combinations of elements are feasible, and organizations should look out for creative incentives for creativity, for instance by offering development and training opportunities as rewards for creativity and innovation (Aguinis et al. 2013).

Communicate regularly and use feedback Although not an aspect of structural fit itself, regular, transparent and clear communication about the program's status, progress and results helps to keep the organization aligned toward the goal. Any successes such as the implementation of a good idea should be made public within the organization. Any feedback from employees or management should be attended to as it is the most helpful tool to check whether the program is effective along the way. If the program does not fit the organizational context in any respect, communication among the participants or towards the organizers (e.g., to the HR department) will probably give an indication of the problem and it should be taken as an opportunity to improve the fit based on this information.

Finally, the people responsible for administering the incentive scheme need to closely monitor its effect on the organization. It is not enough to ensure that rewards are given to employees for relevant contributions because these rewards may still be a waste of money (if the employees had shown their creative and supportive behavior anyway) or fail to realize further creative potentials of the organization (if the incentives are not effective in eliciting creative behavior across the organization). The best indication of a good fit is the combination of a high number of meaningful explorative contributions, their efficient and unagitated evaluation by the organization and the commercial application of the most valuable ones.

6 Summary and Discussion

New ideas and innovative ways of commercially applying knowledge are competitive assets for knowledge-intensive companies. They require exploration, research and the generation and acquisition of knowledge. Based on the existing literature

I have assumed that individual innovative behavior in knowledge-intensive work contexts can be positively influenced by innovation programs, in particular by offering monetary and non-monetary rewards for creative ideas and their application. Setting goals or incentives for innovation in knowledge-intensive companies is a highly specific case of performance management. The fit model of organizational structures implies that there is not a "one-size-fits-all" solution for an incentive system for innovation in knowledge-intensive companies. Most features of an incentive system can be beneficial in one particular circumstance but harmful in others. The following general recommendations have been derived as the key levers to achieve a good fit: clear definition of a program's goal, consistency with existing policies and procedures, transparent and fair rules and intensive use of communication and feedback. As fit is a function of the consistency of system-environment-relations, which imply feedback and adaptation effects, the exact definition of the parameters is a tentative, iterative process characterized by approximation and continuous improvement.

These findings have implications both for academic research and practical innovation management. While researchers agree that additional incentives may lead to more innovation in knowledge-intensive firms, the fit between incentive programs and their organizational context is critical for an innovation program's success. Beside personal and organizational characteristics, the fit between a reward program for innovation and the knowledge-intensive organizational context needs to be considered as an explanatory variable. The resulting hypothesis is that this fit is one of the factors affecting creativity, knowledge acquisition and the implementation of innovative business models in knowledge-intensive companies.

For the practitioner, this implies that in most cases, just offering money or trophies will not work in knowledge-intensive environments. Instead, the systemic interdependencies of the incentive program with its specific organizational context need to be considered. The incentive system needs to be calibrated by iterative improvements based on experience and feedback, and most likely additional changes to organizational structures or practices are required to improve the fit between incentive program and organizational context. An innovation program in knowledge-intensive companies should not only address the motivation to show creative work behavior, but also include measures to promote learning, training, knowledge-sharing and collaboration (Lei et al. 1999; Scarbrough 2003; Swart and Kinnie 2003). Although I have argued that most characteristics of an incentive scheme are neither good nor bad in themselves, some principles can be generalized based on the existing literature and the considerations above: Firstly, an innovation program and the accompanying incentive scheme in a knowledge-intensive environment need to be open and inclusive in order to generate a broad variety of truly creative approaches and to gain support throughout the organization. The latter is of particular importance for knowledge-intensive companies as new insights need this support in order to be applied commercially. Secondly, rewards should be adequate and they should reflect the economic value of any contribution to the firm, provided that this contribution would not be regarded as already covered by any other form of compensation. Economic efficiency of the scheme and reward fairness as perceived by the participants need to be balanced. Although non-monetary rewards are effective and should be used beside monetary ones (Davila et al. 2006; Aguinis et al. 2013), it is an imperative of both effectiveness and fairness to let the contributors of ideas and the supporters of innovation participate financially in any extra value that they helped to create.

A number of limitations of this study should be noted. The arguments presented here are conceptual and have not yet been tested empirically. Although the concept of fit explicitly acknowledges the particularities of a given organization, my approach does not differentiate between different types of knowledge-intensive companies on the conceptual level. Incentive systems may work differently e.g., in pharmaceutical companies, investment banks and IT consulting firms. Similarly, the results are only of limited use for non-commercial knowledge-intensive organizations such as government departments and universities. Although they share many organizational features with commercial entities, there are a number of important differences: It is not normally possible for such organizations to share economic profit with its members, there is usually less flexibility in pay and promotion, and their staff can be expected to be driven, at least partly, by motives other than financial gain. However, like companies, non-commercial knowledge-intensive organizations must ensure that any incentive scheme they consider introducing fits the specific organizational context, even though the relevant characteristics and available tools may be different.

Future research should put the hypothesized parameters of a good fit to an empirical test (both quantitatively and qualitatively). There is also an opportunity to better explain innovative work behavior if future empirical research confirms the hypothesis that the fit between incentive schemes and knowledge-intensive organizational contexts is a determinant of employee creativity.

Knowledge-intensive companies depend heavily on exploration, new ideas and innovation. Incentive schemes to foster creativity and innovation can be effective if they have the right fit to the organizational context. However, an incentive scheme is not an end in itself. Even if rewards are granted and formal recognition is given to innovators, the proof of its organizational fit lies in the generation of economic value through continuous and successful application of new knowledge and creative ideas throughout the organization.

References

Aguinis H (2013) Performance management, 3rd edn. Prentice Hall, Upper Saddle River

Aguinis H, Joo H, Gottfredson RK (2013) What monetary rewards can and cannot do: how to show employees the money. Bus Horiz 56(2):241–249

Ahmed PK (1998) Culture and climate for innovation. Eur J Innov Manag 1(1):30–43

Ahmed PK, Shepherd CD (2010) Innovation management: context, strategies, systems and processes. Financial Times Prentice Hall, Harlow

Amabile TM (1983) The social psychology of creativity: a componential conceptualization. J Pers Soc Psychol 45(2):357–376

Amabile TM (1996) Creativity in context: update to "The Social Psychology of Creativity". Westview Press, Boulder

Amabile TM (2000) Stimulate creativity by fueling passion. In: Locke EA (ed) The Blackwell handbook of principles of organizational behavior. Blackwell, Oxford, pp 331–341

Amabile TM, Hennessey BA, Grossman BS (1986) Social influences on creativity: the effects of contracted-for reward. J Pers Soc Psychol 50(1):14–23

Axtell CM, Holman DJ, Unsworth KL et al (2000) Shopfloor innovation: facilitating the suggestion and implementation of ideas. J Occup Organ Psychol 73(3):265–285

Azoulay P, Graff Zivin JS, Manso G (2011) Incentives and creativity: evidence from the academic life sciences. RAND J Econ 42(3):527–554

Baucus MS, Norton WI, Baucus DI et al (2008) Fostering creativity and innovation without encouraging unethical behavior. J Bus Ethics 81(1):97–115

Becker F (1987) Innovationsfördernde Anreizsysteme. Ein konzeptioneller Beitrag zu einem Innovationsmanagement. Zeitschrift für Personalforschung 1(1):29–60

Bledow R, Frese M, Anderson N et al (2009) A dialectic perspective on innovation: conflicting demands, multiple pathways, and ambidexterity. Ind Organ Psychol 2(3):305–337

Burgelman RA, Sayles LR (1986) Inside corporate innovation. Strategy, structure, and managerial skill. The Free Press, New York

Christensen CM (1997) The innovator's dilemma: when new technologies cause great firms to fail. Harvard Business School Press, Boston

Davila T, Epstein MJ, Shelton R (2006) Making innovation work. How to manage it, measure it, and profit from it. Wharton School Publishing, Upper Saddle River

Drazin R, Van de Ven AH (1985) Alternative forms of fit in contingency theory. Adm Sci Q 30 (4):514–539

Earl J (1994) Knowledge as strategy: reflections on Skandia international and Shorko films. In: Ciborra C, Jelassi T (eds) Strategic information systems: a European perspective. Wiley, New York, pp 53–69

Edvardsson IR (2008) HRM and knowledge management. Emp Relat 30(5):553–561

Egan TM (2005) Factors influencing individual creativity in the workplace: an examination of quantitative empirical research. Adv Develop Hum Res 7(2):160–181

Ekvall G (1971) Creativity at the place of work. Reklamito, Stockholm

Farr JL, Ford CM (1990) Individual innovation. In: West MA, Farr JL (eds) Innovation and creativity at work. Psychological and organizational strategies. Wiley, Chichester, pp 63–80

Gumusluoglu L, Ilsev A (2009) Transformational leadership, creativity, and organizational innovation. J Bus Res 62(4):461–473

Hansen MT, Nitin N, Tierney T (1999) What's your strategy for managing knowledge. Harv Bus Rev 77(2):106–116

Janssen O (2000) Job demands, perceptions of effort-reward fairness and innovative work behaviour. J Occup Organ Psychol 73(3):287–302

Kanter RM (1983) The change masters. Corporate entrepreneurs at work. Simon & Schuster, New York

Kanter RM (1988) When a thousand flowers bloom: structural, collective and social conditions for innovation in organizations. In: Staw BM, Cummings LL (eds) Research in organizational behavior, vol 10. J.A.I. Press, Greenwich, pp 169–211

Lei D, Slcoum JW, Pitts RA (1999) Designing organizations for competitive advantage: the power of unlearning and learning. Organ Dynam Winter:24–38

Lerner J, Wulf J (2007) Innovation and incentives: evidence from corporate R&D. Rev Econ Stat 89(4):634–644

Manso G (2011) Motivating innovation. J Financ 66(5):1823–1860

Miron E, Erez M, Naveh E (2004) Do personal characteristics and cultural values that promote innovation, quality and efficiency compete or complement each other? J Org Behav 25(2):175–199

Muller E, Doloreux D (2009) What we should know about knowledge-intensive business services. Technol Soc 31(3):64–72
Mumford MD, Gustafson SB (1988) Creativity syndrome: integration, application, and innovation. Psychol Bull 103(1):27–43
Oldham GR, Cummings A (1996) Employee creativity: personal and contextual factors at work. Acad Manag J 39(3):607–634
Olin T, Wickenberg J (2001) Rule breaking in new product development–crime or necessity? Creat Innov Manag 10(1):15–25
Scarbrough H (2003) Knowledge management, HRM and the innovation process. Int J Manpow 24 (5):501–516
Scott SG, Bruce RA (1994) Determinants of innovative behavior: a path model of individual innovation in the workplace. Acad Manag J 37(3):580–607
Shalley CE (1995) Effects of coaction, expected evaluation, and goal setting on creativity and productivity. Acad Manag J 38(2):483–503
Sheehan NT, Stabell CB (2007) Discovering new business models for knowledge intensive organizations. Strateg Leadersh 35(2):22–29
Starbuck WH (1992) Learning by knowledge-intensive firms. J Manag Stud 29(6):713–740
Sung Y, Choi JN (2012) Effects of team knowledge management on the creativity and financial performance of organizational teams. Organ Behav Hum Decis Process 118(1):4–13
Swart J, Kinnie N (2003) Sharing knowledge in knowledge-intensive firms. Hum Resour Manag J 13(2):60–75
Toubia O (2006) Idea generation, creativity, and incentives. Mark Sci 25(5):411–425
Unsworth K (2001) Unpacking creativity. Acad Manag Rev 26(2):289–297
Van Djik C, Van den Ende J (2002) Suggestion systems: transferring employee creativity into practicable ideas. R&D Manag 32(5):387–395
West MA (2002) Ideas are ten a penny: it's team implementation not idea generation that counts. Appl Psychol 51(3):411–424
West MA, Farr JL (1990) Innovation at work. In: West MA, Farr JL (eds) Innovation and creativity at work: psychological and organizational strategies. Wiley, Chichester, pp 3–13
Woodman RW, Sawyer JE, Griffin RW (1993) Toward a theory of organizational creativity. Acad Manag Rev 18(2):293–321
Zhou J, Shalley CE (2003) Research on employee creativity: a critical review and directions for future research. Res Pers Hum Resour Manag 22:165–217

Part VII
Applied Contributions, Country and Case Studies

Adoption and Use of Management Controls in Higher Education Institutions

Thomas W. Guenther and Ulrike Schmidt

Abstract Management control systems are formal, routine-based systems which help to maintain or alter organizational activities to increase efficiency and effectiveness. During the last decade the higher education sector faced crucial changes towards more autonomy and self-financing in most countries. Thus, also the management control systems were challenged to be adapted to these environmental changes. This article gives insight into the design of management control systems of 176 higher education institutions in Germany, Austria and Switzerland (response rate 40.9 %). We analyze both the adoption and use of budgeting, planning and reporting instruments, of financial control instruments and of instruments and systems for quality management.

1 The Spread of Management Control Systems in the Higher Education Sector

Management control systems are formal, routine-based systems which help to maintain or alter organizational activities (Simons 1995) and which enable an organization to effectively and efficiently accomplish its objectives by the optimal use of the organizational resources (Anthony 1965). Management control systems should be tailored to an organization's environment to support organizational variables in order to find the appropriate fit and enhance organizational performance (Langfield-Smith 1997, 2007; Otley 1980). Management control system research is focused on implementation and use (Malmi 2001; Speckbacher and Wentges 2012; Ittner and Larcker 2001) and on considering the fit with contingency factors (Chenhall 2003). Thus, for this paper we do not only look on the implementation, but also on the use of different management controls. In the last decade management control system research underlined the need to consider different controls existing in parallel and to analyze their interaction with each other. Thus,

T.W. Guenther (✉) • U. Schmidt
Technische Universität Dresden, Mommsenstraße 10, 01062 Dresden, Germany
e-mail: thomas.guenther@tu-dresden.de

I.M. Welpe et al. (eds.), *Incentives and Performance*,
DOI 10.1007/978-3-319-09785-5_22

comprehensive management control system frameworks with different controls, which mutually interact, have been developed (e.g., Simons 1995; Malmi and Brown 2008; Ferreira and Otley 2009; Merchant and van der Stede 2011).

Regarding the adoption of management control systems we refer to Malmi and Brown (2008) and use the structure of the management control as a package framework to analyze the adoption and suitability of a variety of management control instruments in the higher education sector.

Simons (1995) states in his levers of control (LoC) framework that all four control systems (beliefs system, boundary system, diagnostic control system and interactive control system) are intertwined and contribute together to the success of the organization. He distinguishes the four levers of control which demonstrate conflicting dimensions of control to implement an intended or emerging strategy for the organization. The beliefs system stands for the core values of the organization which can be formally implemented in a vision or mission statement or in the guidelines of an organization. The beliefs system helps to identify problems and to solve them. It inspires motivation of employees and supports them to find innovative solutions (Simons 1995). In contrast to the beliefs system, the boundary system defines limitations for the employees of the organization when they look for entrepreneurial opportunities. The purpose of a boundary system is to avoid risks for the organization. The boundary system is formally implemented in the code of conduct or in the compliance rules of an organization (Simons 1995). Thus, there is a tension created by pairing the beliefs and the boundary system which defines commitment, empowerment and freedom to operate (Simons 1995).

Whereas beliefs and boundary system describe the design of management controls, the diagnostic and interactive control systems are about the use of management control instruments in an organization. The diagnostic control system stands for the typical planning and feedback loop of management control systems. It compares actual performance measures with targets set before by the planning system of the organisation. Thus, analysing variances and drawing conclusions from that is the core of the diagnostic control system. Again, in contrast, the interactive control system is about strategic uncertainties and symbolizes the feedforward processes by communicating outcomes and results, by discussing consequences of failure and by finding innovative solutions, for example, in the form of emergent strategies to deal with strategic uncertainties of the organization. Also the diagnostic use and the interactive use establish a dynamic tension which is not per se bad for the organization but constructive to balance feedback and feedforward processes. Thus, all four levers of control form the framework of a management control system following the conceptualization of Simons (1995). In our study we use Simons' (1995) LoC framework not only to analyze what kind of management control systems are adopted, but to shift the focus on how they are used within the organizations of our setting.

In the last decade the higher education sector in Western countries faced crucial transformations (Brown and Brignall 2007; Parker 2011). Governments have reduced their financial support of higher education institutions and demand more competition and financial self-dependence of universities and colleges (Lawrence and Sharma 2002). Furthermore, under the headline of new public management instruments from the private sector have been transferred to the public sector to increase professionalism, efficiency and effectiveness (Hood 1995; Pollitt 2002). The intention was to change the focus from inputs to outputs. Thus, products as outputs of the public organization are defined. Costs of these outputs are calculated by the introduction of cost accounting to get artificial prices for the outputs on the one hand because efficient and competitive markets for the products often do not exist. On the other hand cost accounting is implemented to focus on efficiency, the relation of inputs and outputs, by calculating costs of public products and to inspire competition for efficient production of outputs. Furthermore, contracts between the government and the "producers" of public goods and services are concluded in order to emphasize accountability (Brown and Brignall 2007). With lowering hierarchical control more autonomy and more responsibility are given to heads of higher education institutions who in turn should hand down more autonomy for academic performance but also for financial and HR management to structural units such as departments and schools (Harman 2002). The counter-part of increased autonomy is increased reporting effort to satisfy information needs of higher education management and government authorities and to justify public funding with taxpayers' money. Furthermore, more autonomy within the higher education institutions results in a higher need for management control systems and management control instruments to use the higher autonomy on each hierarchy level of the higher education institution to enhance management. Granting autonomy also raises concerns of government heads and ministries how lower hierarchy levels in public administration will use or might misuse the gained autonomy. As a consequence more comprehensive reporting is introduced to counteract potential misuse. However, trust in well selected professional and administrative staff that live the core values of the organization, might avoid the need for extended reporting.

In Central Europe state universities still dominate the higher education sector. Nevertheless, the number of private schools and universities increases and these institutions gain market share in the higher education sector. This also results in higher competition in the higher education sector.

Considering the profound changes in the higher education sector it has to be stated that management control system research in this special part of the public sector is still limited. Research is mostly focusing on case studies (e.g., Arnaboldi and Azzone 2010; Modell 2006) or on single instruments (e.g., Lawrence and Sharma 2002; Modell 2003). Broader cross-sectional studies besides the study of Chung et al. (2009) on Australian universities are missing so far. Using survey data from 176 chancellors of universities from Germany, Austria, and Switzerland we show what types of management control systems are adopted by higher education institutions and how they are used.

2 Methods

2.1 Design and Sample of the Survey

Using the tailored design method by Dillman et al. (2009) we developed a standardized written questionnaire for data collection. All questions were derived and designed based on management control system theory. We use, if available, validated items and scales of previous studies. We report the sources of items when we present results on the specific aspect. The questionnaire was pretested with 11 experienced researchers and practitioners from higher education institutions to adjust items in wording to the higher education setting.

The population consists of all 478 officially recognized private or public higher education institutions in Germany, Austria, and Switzerland (full sample) (381 in Germany, 52 in Austria, 45 in Switzerland; 143 universities, 261 universities of applied sciences, 74 specialized universities for art, music and teacher education). The overall response rate in our study was 40.9 %. After correcting for questionnaires with missing data, 176 observations are used for the further analysis.

Respondents are primarily (vice) chancellors (39.2 %), (vice) presidents (20.4 %) or top management of the university (7.4 %), head of administration/finance (20.5 %) or head of controlling (6.8 %). Furthermore, on average respondents have been working in the current positions for 6.9 years and for 10.7 years at the responding higher education institution. In addition, on average they have 11.7 years of professional experience in higher education administration. For further information and details of the survey see Guenther et al. (2013).

2.2 Measures for Adoption and Use of Management Control Instruments

Following the theoretical framework of Simons (1995) and Malmi and Brown (2008) in our survey we differentiate between the adoption of management control systems and its use. Thus, we asked top management of higher education institutions differentiated questions first on the variety of management control instruments they adopted in their organization (adoption of management control) and second how they use the management control instruments (use of management control systems). For the adoption of management control systems we use the structure of the management control as a package framework of Malmi and Brown (2008) as the framework of Simons (1995) is not so specific and detailed about the design of management control instruments as he "only" differentiates between beliefs and boundary systems for the design and between interactive and diagnostic use for the use of management control systems. Thus, Simons' (1995) framework is used for the second question on how management control systems are used by higher education institutions.

To get insights in the adoption of management control instruments in the higher education sector, higher education management was asked which general management control instruments are adopted from the following three categories of control instruments and if they are adopted whether they are suitable for the special setting of higher education institutions. These questions help to assess the usability and the cost/benefit relation for the specific instrument. The three categories are:

(1) budgeting, planning and reporting instruments,
(2) financial control instruments and
(3) quality instruments and quality management systems.

This structure follows the categorization of Malmi and Brown (2008) who differentiate planning and cybernetic controls into long range planning, action planning, budgets, financial measurement systems, non-financial measurement systems (for the higher education sector especially quality instruments and quality management systems, but also performance measurement, risk and environmental management systems) and hybrid measurement systems (such as the balanced scorecard). We use a much more comprehensive list of instruments than mentioned in Malmi and Brown (2008) to be more detailed and more specific for the special characteristics of the higher education sector. We also categorize the management control instruments of Malmi and Brown (2008) differently. The categories of the management control instruments are addressed in the following section which covers the budgeting, planning and reporting process (Sect. 3.1) and in the financial control instruments (Sect. 3.2) which are the traditional source of management control in Continental Europe. Researchers have to be aware of using research instruments and items which are familiar to the management control culture of their respondents (Guenther 2013). Due to the importance of quality in higher education institutions we decided to have a separate battery of questions on this issue (Sect. 3.3). The items on adoption and suitability of management control instruments were taken from previous empirical surveys on management control instruments in German-speaking countries and adjusted to the special setting of higher education institutions (Guenther and Gonschorek 2011; Guenther and Heinicke 2013).

In addition to asking what instrument are adopted by higher education institutions and how they are perceived to be suitable for the higher education setting, in the last results section we report about the answers on the interactive and diagnostic use of management control systems in higher education institutions as in the framework of Simons (1995). The items were taken from previous empirical surveys on management control system research (Henri 2006; Widener 2007).

3 Results

3.1 *Adoption of Budgeting, Planning and Reporting Instruments*

Table 1 gives an overview of the adoption and suitability of budgeting, planning, and reporting instruments in the higher education institutions of our sample. Five major results can be derived. First, the budgeting process is used in almost all higher education institutions and additionally, it is seen to be helpful and suitable. Second, components of a strategic planning process (vision, mission statement, business plan) can be found in more than half of the higher education institutions and confirmation rates for suitability are also very high. Third, negotiating target agreements with structural units (e. g. institutes, chairs and professors) within a higher education institution can be found in 68.6 % of all institutions with a very high suitability rate of 90.0 %. Target agreements are typically connected with budgeting, planning and reporting processes as they incorporate specific plans for key performance indicators for these structural units and measure target achievement afterwards. Very often these target agreements are connected with budgets, i.e., input to enable target achievement and reward systems to incentivize target achievement afterwards. Thus, one central element of new public management for the higher education sector seems to be widely practiced and also accepted. Forth, higher education institutions are restrictive with adopting modern sophisticated and holistic management tools (e. g. performance measurement, risk management or environmental management systems) whereas confirmation rates of suitability are high. This may be caused by lacking personal and financial resources for the implementation of new techniques or by mental barriers against complex and innovative management tools. Finally, due to public financing and financial constraints instruments for mid-term financial and liquidity planning are widely spread and accepted.

3.2 *Adoption of Financial Control Instruments*

Historically, due to public funding cash accounting systems focusing on cash inflow and cash outflow were dominant. Driven by the introduction of new public management in the higher education sector double-entry bookkeeping has gained importance to better measure product costs for services in teaching and research (by also considering accruals, and depreciation and amortization of fixed assets) while the adoption of cash and extended cash accounting systems has decreased.

Table 2 shows that more than 50 % of the higher education institutions have introduced double-entry bookkeeping and financial accounting (i.e., balance sheet and income statement) recently. In addition, cost category and cost centre accounting is widely spread with high confirmation rates of suitability of around 80 %.

Table 1 Adoption of budgeting, planning, and reporting instruments in higher education institutions (greatest frequency for both adoption and suitability of the instruments in bold letters each)

Budgeting, planning and reporting instruments	Is the instrument adopted? Yes (%)	No (%)	Do not know/ No answer (%)	If adopted, is the instrument suitable? Suitable (%)	Partially suitable (%)	Not suitable (%)	So far not discussed/ No answer (%)
Budgeting/resource allocation system	**90.2**	7.5	2.3	**85.4**	12.7	0.0	1.9
Mission statement	**76.6**	18.3	5.1	**67.2**	29.9	2.2	0.7
Mid-term financial planning (1–4 years)	**72.6**	24.6	2.9	**76.2**	18.3	1.6	4.0
Target agreements within higher education institution	**68.6**	29.1	2.3	**90.0**	9.2	0.8	0.0
Strategic planning/ business plan	**64.6**	32.0	3.4	**77.9**	22.1	0.0	0.0
Liquidity planning	**63.4**	31.4	5.1	**82.0**	14.4	0.0	3.6
Process analyses (e.g., work flow, activity costing)	**57.7**	37.7	4.6	**76.2**	18.8	1.0	4.0
Budgeted balance sheet and income statement	**56.3**	38.5	5.2	**73.5**	21.4	2.0	3.1
Economic ratio analysis of financial statements	**56.0**	40.0	4.0	**70.4**	26.5	2.0	1.0
Vision statement	**51.1**	41.4	7.5	**66.3**	30.3	2.2	1.1
Performance measurement system	48.0	**48.6**	3.4	**65.1**	33.7	0.0	1.2
Risk management system (risk exposure, risk assessment, risk reporting)	36.8	**56.9**	6.3	**57.8**	39.1	1.6	1.6
Environmental management system (e.g., following EMAS)	16.6	**73.1**	10.3	**62.1**	20.7	3.4	13.8

Driven by EU regulation to show actual costs for research projects with third parties and to avoid public subsidizing the calculation of planned and actual costs for research projects is quite well adopted. However, cost unit accounting to calculate costs for teaching is not widely spread. Furthermore, assessing the efficiency of investment projects is also scarce as these are mostly financed by additional budgets granted from external sources such as state driven research foundations, EU, federal or state funds. Interestingly, the suitability for such calculations is seen to be very high. Thus, an implementation gap for some financial control instruments in higher

Table 2 Adoption of financial control instruments (greatest frequency for adoption and suitability of the instruments in bold)

Financial control instruments	Is the instrument adopted? Yes (%)	No (%)	Do not know/ No answer (%)	If adopted, is the instrument suitable? Suitable (%)	Partially suitable (%)	Not suitable (%)	So far not discussed/ No answer (%)
Cost centre accounting	**93.1**	3.4	3.4	**82.2**	14.7	0.0	3.1
Cost accounting in general	**88.0**	9.1	2.9	**75.0**	21.1	0.7	3.3
Cost category accounting	**85.7**	9.7	4.6	**79.9**	16.1	0.0	4.0
Calculation of planned costs for research projects with third parties	**80.6**	16.6	2.9	**81.6**	13.5	0.0	5.0
Product unit costing	**77.0**	18.4	4.6	**76.1**	20.1	0.7	3.0
Calculation of actual costs for research projects with third parties	**67.4**	28.0	4.6	**79.7**	12.7	0.8	6.8
Double-entry bookkeeping	**50.9**	41.1	8.0	**78.7**	19.1	0.0	2.2
Cash flow statements	38.3	**51.4**	10.3	**76.1**	20.9	3.0	0.0
Calculation of planned costs for teaching	36.0	**58.3**	5.7	**82.5**	14.3	0.0	3.2
Cash accounting	29.7	**58.3**	12.0	**42.3**	42.3	7.7	7.7
Calculation of actual costs for teaching	26.3	**65.7**	8.0	**91.1**	6.7	0.0	2.2
Marginal costing (e.g., fixed and variable costs per student)	22.9	**66.9**	10.3	**60.0**	35.0	0.0	5.0
Static investment appraisal for investment projects (e.g., comparison of costs or profit)	22.4	**67.2**	10.3	**66.7**	28.2	0.0	5.1
Extended cash accounting	20.0	**68.0**	12.0	**57.1**	37.1	2.9	2.9
Life cycle costing for investment projects	12.1	**75.3**	12.6	**76.2**	9.5	9.5	4.8

education practice can be derived (e. g., calculation of costs for teaching, life cycle costing and appraisal of investment projects).

3.3 Adoption of Quality Instruments and Quality Management Systems

Quality assurance of research and teaching becomes more and more important in the higher education sector. Competition for the most talented students and for scarce resources and funding opportunities for research projects results in a need for instruments to measure, control and manage quality in research and teaching. Table 3 demonstrates the adoption and suitability of quality instruments for our sample. It shows that a large variety of instruments is used by higher education institutions to control quality.

The evaluation of lectures by students and alumni surveys seem to be a must for universities which is confirmed by the very high adoption rates. The accreditation of programs or institutions is widely used (91.4 % of responding institutions), but the confirmation rate for suitability is lower with 70.3 %. Similar to quality benchmarking with other higher education institutions scepticism about suitability is quite high, which can be derived from a high share of respondents seeing "only" partial suitability. Furthermore, measuring quality by indicators and by awards for both research and teaching is less often adopted, however suitability is confirmed by two third of respondents. Altogether, a large variety of different quality instruments exists in parallel to each other in higher education institutions.

In contrast to single quality instruments which are generally widely spread in our sample (Table 3), comprehensive quality management systems are only rarely adopted in higher education institutions. Neither the European Foundation for Quality Management (EFQM) model, nor the International Organization for Standardization (ISO) 9000ff. standards, nor the Total Quality Management (TQM) approach are widely adopted and rates for suitability are, in relation to other instruments, relatively low with a confirmation rate of about 50 % (Table 4).

To sum up, the management control landscape has seen dramatic changes in the last decade. The introduction of new public management resulted in giving up traditional cash accounting systems in higher education institutions. Financial accounting and cost accounting have been highly promoted and introduced and are now widely spread and also widely accepted in the higher education sector. Another crucial component of new public management, concluding target agreements between higher education management and the structural units, is widely used and highly confirmed as an adequate control instrument by our respondents. Furthermore, we also see the impact of regulations on the use of management controls. This can be seen by the high spread of cost unit accounting for research projects with third parties due to existing EU regulations forbidding cross subsidizing of private firms via research funding. The same holds true for strategic planning, accreditation and target agreements which are demanded by many university laws in Germany, Austria, and Switzerland. Some of these politically supported control instruments are critically viewed by respondents as explained above. Furthermore, comprehensive management systems (for quality, risk, performance or environment) are only scarcely adopted by higher education institutions.

Table 3 Adoption of quality instruments (greatest frequency for adoption and suitability of the instruments in bold)

Quality instruments	Is the instrument adopted?			If adopted, is the instrument suitable?			
	Yes (%)	No (%)	Do not know/ No answer (%)	Suitable (%)	Partially suitable (%)	Not suitable (%)	So far not discussed/ No answer (%)
Evaluation of lectures by students	**96.0**	2.9	1.1	**85.5**	11.4	0.0	3.0
Accreditation	**91.4**	6.3	2.3	**70.3**	24.1	3.2	2.5
Student service offices	**88.6**	7.4	4.0	**86.4**	9.7	0.6	3.2
Alumni surveys	**88.0**	8.0	4.0	**90.3**	6.5	0.0	3.2
Student satisfaction surveys	**88.0**	5.7	6.3	**82.5**	14.3	0.0	3.2
Trainings in didactics for higher education institutions	**85.7**	10.9	3.4	**84.7**	10.7	0.0	4.7
Mentoring and tutoring programs	**85.1**	10.9	4.0	**85.2**	10.7	0.0	4.0
Scientific training programs, summer schools	**81.1**	13.1	5.7	**81.4**	14.3	0.0	4.3
Appointment rules for new faculty	**78.6**	14.5	6.9	**86.0**	11.8	0.0	2.2
Tuition/financial support for quality assurance	**69.7**	26.3	4.0	**85.2**	13.1	0.0	1.6
Benchmarking with other higher education institutions	**66.3**	30.3	3.4	**64.7**	31.9	0.9	2.6
Analysis of dropouts	**65.7**	28.0	6.3	**79.8**	17.5	0.0	2.6
Use of guidelines for good scientific practice (e.g., following the German Research Foundation (DFG))	**65.5**	24.1	10.3	**87.6**	8.8	0.0	3.5
Quality measurement using indicators	**62.9**	32.0	5.1	**69.1**	27.3	0.9	2.7
Peer reviews	**55.2**	35.1	9.8	**79.2**	16.7	0.0	4.2
Awards for excellent teaching	**49.4**	46.0	4.6	**75.6**	20.9	0.0	3.5
Awards for excellent research	43.1	**52.9**	4.0	**74.3**	23.0	0.0	2.7

Table 4 Adoption of quality management systems (greatest frequency in bold letters)

	Is the instrument adopted?			If adopted, is the instrument suitable?			
Quality management systems	Yes (%)	No (%)	Do not know/ No answer (%)	Suitable (%)	Partially suitable (%)	Not suitable (%)	So far not discussed/ No answer (%)
European Foundation for Quality Management Modell (EFQM)	21.1	**64.0**	14.9	**56.8**	37.8	0.0	5.4
ISO 9000ff. standards	18.3	**65.7**	16.0	**56.3**	25.0	3.1	15.6
Total Quality Management (TQM)	13.7	**71.4**	14.9	**58.3**	37.5	0.0	4.2

The same holds for cost calculation for teaching and for the evaluation of investment projects including life cycle assessments.

3.4 Intensity of Use of Management Control Instruments

Analyzing the adoption of management control instruments only gives insight into what instruments exist and which are used in higher education institutions. Applying Simons' (1995) framework for management control systems allows us to additionally explore how these instruments are used. Simons (1995) distinguishes between a diagnostic and an interactive use of management control systems, such as performance measurement systems (PMS) or budgets. The diagnostic use of a PMS is formally applied by management to monitor and reward organizational outcomes and correct deviations. The achievement of set targets and the comparison of actual vs. planned performance indicators underline the feedback loop within cybernetic management control systems. In the higher education sector the diagnostic use of a PMS is essential for reporting goal achievement of the structural units (e. g. departments, schools, institutes) to the higher education management. Management interactively uses a PMS to get personally involved in decision activities and in active dialogue throughout the organization. It inspires interaction between members of the organization and strategy development driven by measuring and subsequently discussing results. This is generally regarded as the feed forward loop of management control systems. For innovative industrial settings Henri (2006) and Widener (2007) show that management controls such as PMS are differently used within organizations. Referring to Simons (1995) we also collected data on the use of performance measures in higher education institutions distinguishing in diagnostic and interactive use. To measure diagnostic and interactive use of performance measurement systems we use items validated in previous empirical studies

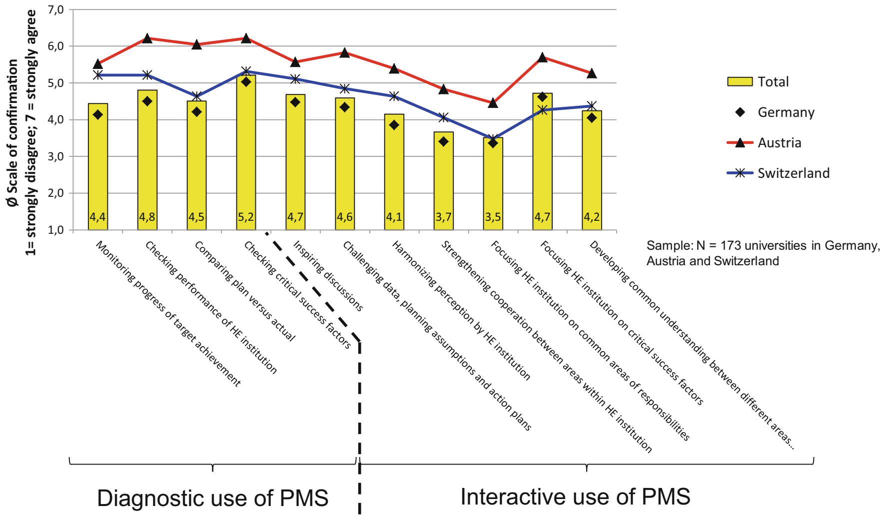


Fig. 1 Diagnostic and interactive use of performance measurement systems

on innovation settings by Vandenbosch (1999), Henri (2006) and Widener (2007). As we measure 7 point-Likert scales for all items (scales from 1 to 7 with 1 for very low usage and 7 for very high usage) we are able to measure the intensity of use of the performance measurement system of the higher education institution.

Figure 1 presents the results for the diagnostic and interactive use of PMS within the higher education institutions of our sample. Overall, the level of diagnostic and interactive use of the PMS is higher in Austrian and Swiss universities than in German universities. For Austria this might be driven by the regulation that Austrian universities have to deliver information to the administration by a so-called intellectual capital statement (*Wissensbilanz*) which is formally required by a regulation of the Austrian ministry of science and research (Wissensbilanzverordnung in BMWF 2010). Austrian universities seem to use and discuss the collected data of the PMS much more than German universities which confirms the intended benefits of the external regulation. In Switzerland higher interactive use of the PMS might by inspired by the fact that target agreements do not exist which might give room for more openly discussing outputs and outcomes of teaching and research.

The average of all items for the total sample in Fig. 1 is slightly above 4.0 which indicates only a medium level of diagnostic as well as interactive use of PMS within higher education institutions. Contrasting the use with the above results on the adoption of control instruments in Tables 1, 2, 3, and 4 one might get the impression that potentially driven by higher education sector reforms higher education institutions have implemented a variety of control instruments in the last decade, but the intensity of using them is only average across higher education institutions. This may be caused by either limited suitability of PMS or by reluctance of using the PMS instruments.

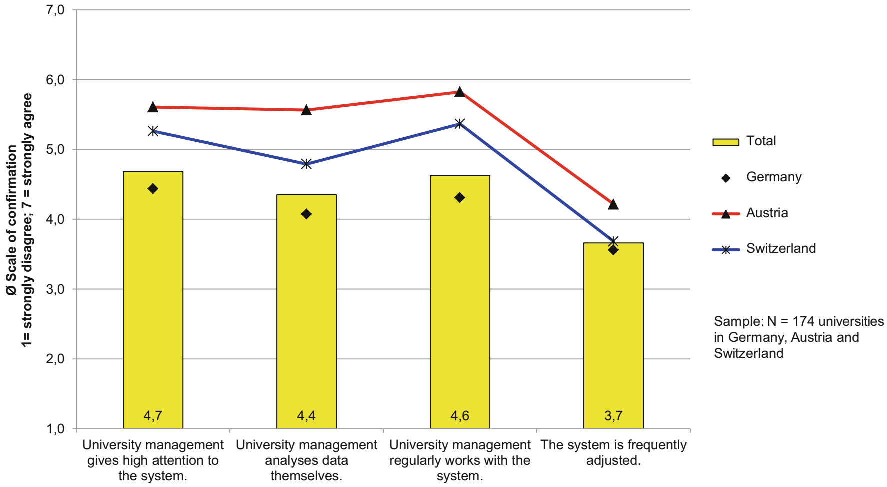


Fig. 2 Interactive use of performance measurement systems by heads of universities

Interestingly, for the items "Checking critical success factors" and "Focusing the higher education institution on critical success factors", German universities are much closer to Swiss universities and to the average than for other items. This might be explained by the stronger focus also driven by state governments on performance measures which are often part of the resource allocation model between the state and the universities. Additional pressure was created by the German Research Foundation (*Deutsche Forschungsgemeinschaft—DFG*) with the excellence program focusing on research excellence only. Furthermore, in general, there are no significant differences between levels of use when comparing diagnostic and interactive use of controls.

Figure 2 focuses on the role of the heads of the university (president, vice-president, chancellor) and how they are involved in using performance measures to manage their universities. Interactive use is understood as whether the top management really works with the data, discusses with subordinates, adjusts the PMS if data is missing or measurement problems occur etc. Again, the picture of the use of management controls is confirmed. The level of the use of PMS is only medium. It is noteworthy that for these questions most of the respondents are self-assessing their own behavior as they are part of the university management. This confirms the insufficient interactive use by management as an overestimating respondent bias is normally observed. Furthermore, the intensity of use is higher for Austrian and Swiss universities. In sum, management control systems are adopted, but university top management only uses performance measures interactively with medium intensity. In the article at hand we only present results for the use of PMS but the same holds also for the use of budgeting as a management control instrument. We report on the adoption of budgeting systems in Table 1 but not on the use of budgeting, data which also had been collected in our survey.

4 Implications for Practice and Research

The results we can derive from our survey for the adoption and use of management control instruments in the higher education sector allow conclusions for both practice and research. As different stakeholder groups are engaged in the higher education sector, we split the normative implications in addressing different stakeholder groups:

Implications for higher education institutions and their top management:

- Universities seem to be driven by higher education sector reforms to implement a large variety of new control instruments (double-entry bookkeeping, cost accounting, performance measures, strategic planning processes, target agreements, quality instruments etc.). However, the intensity of use of these instruments is, overall, "only" at a medium level. Higher education institutions are called to consider whether a management control instrument does make sense, also in the light of limited administrative personnel and financial capacity. What are the instruments which are most beneficial for the organization? Do benefits exceed costs in maintaining a given management control instrument?
- Just implementing a control instrument in the higher education sector does not seem to be sufficient for a higher education institution. Studies on the use of management control systems tell us that management controls can be used in different ways and that the type of use is crucial for the success of the organization. In innovative settings management controls are often intertwined and are mutually reinforcing themselves in order to enhance performance. Thus, the focus of practice in higher education institutions should shift from running control instruments to how to run and how to use management control systems. How does the roll-out of a new management control instrument look like? Who should use the instrument when and how frequently? What parties, what levels in the hierarchy and how many members should be involved in control processes? Do higher education institutions use management control instruments diagnostically or interactively or both?
- Furthermore, some of the analyzed management controls face only partial suitability in the higher education sector. Thus, the costs and the benefits of promising controls for the higher education institution should be carefully reflected (e.g., benchmarking, accreditation, financial ratio analysis). Do these instruments make sense at all for the specific setting of higher education institutions? Can we abandon some of these instruments if they are not beneficial? Can we adjust the instrument or reduce frequency to be beneficial for higher education institutions?
- We observe differences in the adoption and use of management control instruments between different countries. Thus, it might be interesting to explore why there are differences and also to search for other drivers explaining differences in the adoption and use of management controls. Are we willing to learn what management control instruments should be adopted and how to use them from

peers in the same country and especially from peers in other countries? Where do we find best in class adoption and usage? Where are our competitors? How do they run their universities? Are we willing and capable to visit and to learn from higher education institutions in other countries and other continents?

- We also see differences in unreported results in adoption and use between different types of higher education institutions, especially between universities on the one hand and universities of applied sciences and specialized universities for art, music and teacher education on the other hand. This raises further questions on the adoption and use of management control systems which have to be addressed by higher education management: What management control instruments can be used and how should they be used in special settings like an academy of art? Regarding the smaller sizes of these institutions how can management control instruments be downsized to their setting? How has their special setting been reflected in a different layout of management control systems?

Implications for funding and governing institutions in the higher education sector (research funding agencies such as the Deutsche Forschungsgemeinschaft (DFG) in Germany, ministries which are funding higher education institutions as an entity, governing agencies such as the science council (Wissenschaftsrat), higher education rector conference (Hochschulrektorenkonferenz) as examples for the German setting):

- Development in management control in higher education institutions has been driven by installing new instruments and by the tendency to give more autonomy to higher education institutions following the notion of new public management. This development was heavily supported by the above mentioned funding and governing institutions. Heads of these bodies should ask themselves whether they should shift from demanding higher education institutions to adopt management control instruments like cost accounting, strategic planning or quality management to focusing on whether and how these instruments are used and what the benefits of these instruments are. Is it reasonable to demand a list of management control instruments in university laws and regulations? Can the regulation of management control instruments, which should be implemented by higher education institutions, be reduced in favor of a stricter monitoring of outputs of an autonomous and self-controlling university? This might leave room for higher education institutions to decide themselves what management control instruments to apply and how to use them. Should evaluation of higher education institutions place particular emphasis on outcome and second on the actual use of management control systems instead of evaluating the existence of instruments in narratives?
- Funding and governing agencies should develop procedures and techniques on how to supervise and consult higher education institutions in finding the optimal design and use for the management control systems. Are these agencies willing to look on existing experiences of other supervising bodies on how to evaluate management control processes, on how to assess them and on how to give advice

(e.g., using professional and experienced controllers in governing bodies, having financial and controlling experts besides professors and academics in evaluation teams for excellence competitions or research funding)?

- Our study focuses on the management control systems of higher education institutions in German-speaking countries. However, higher education institutions are just the "producers" of research and teaching in the entire higher education system. Thus, also the management control systems of the governing and administrating bodies in the higher education sector (ministries, councils, research funding institutions etc.) should become part of the management control system perspective in the higher education sector. Are governing bodies capable of analyzing, discussing and diagnostically or interactively using reports of higher education institutions themselves? Are these bodies willing to learn from higher education systems in other countries (often one federal state in Germany just benchmarks with another federal state in Germany forgetting about the fact that competition in teaching and research is global)? When higher education institutions should have target agreements and strategic plans, are these derived from superior strategic plans for the higher education sector in the country or the federal state? When universities develop strategies for their higher education institution more or less autonomously who and how are these "divisional" plans coordinated for the whole country or the federal state to reach and fulfil public targets and to contribute to public welfare?

Implications for research in higher education institutions:

- Facing a multitude of management controls, it might be interesting to shed some light on what combinations of controls are more or less suitable and how these combinations are related to performance of higher education institutions in research and teaching. One question is here how an optimal configuration of a management control system looks like.
- Another interesting research avenue may be to explore the role of autonomy, strategy and other antecedents on the layout of management control systems in the higher education sector. What are antecedents and success factors of an effective management control system in the higher education sector?
- Since respondents of our surveys were the heads of university management, it would be interesting to explore the perceptions of other hierarchy levels of the university like deans or academic professionals (e. g., lecturers and professors). How do higher education institutions deal with management control systems on different levels of their hierarchy?
- There are other management systems (e. g., HR management, technology management) or controls (e. g., culture or administrative controls) which are not focused on in this article. Nevertheless, it might be interesting to also look into these controls to examine changes driven by higher education sector reforms.

Acknowledgment We are grateful for the financial support of the Saxonian Ministry of Science and Art (SMWK) and for technical support of Nancy Grossmann, Bianca Schoenherr and Sebastian John for collecting the data for the underlying survey.

References

Anthony RN (1965) Planning and control systems: a framework for analysis. Harvard Business School, Boston

Arnaboldi M, Azzone G (2010) Constructing performance measurement in the public sector. Crit Perspect Account 21(4):266–282

BMWF (2010) Verordnung der Bundesministerin für Wissenschaft und Forschung (Wissensbilanz-Verordnung 2010, WBV 2010). Regulation of the Minister of Science and Research. Intellectual capital statement regulation. Bundesgesetzblatt für die Republik Österreich, BGBLA_2010_II_216, https://www.ris.bka.gv.at/Dokumente/BgblAuth/BGBLA_2010_II_216/BGBLA_2010_II_216.pdf. Accessed 07 Mar 2014

Brown R, Brignall S (2007) Reflections on the use of a dual-methodology research design to evaluate accounting and management practice in UK university central administrative services. Manag Account Res 18:32–48

Chenhall RH (2003) Management control systems design within its organizational context: findings from contingency-based research and directions for the future. Account Organ Soc 28(2):127–168

Chung TKJ, Harrison GL, Reeve RC (2009) Interdependencies in organization design: a test in universities. J Manag Account Res 21:55–73

Dillman DA, Smyth JD, Christian LM (2009) Internet, mail, and mixed-mode surveys. The tailored design method, 3rd edn. Wiley, Hoboken

Ferreira A, Otley D (2009) The design and use of performance management systems: an extended framework for analysis. Manag Account Res 20(4):263–282

Guenther T (2013) Conceptualisations of 'controlling' in German-speaking countries: analysis and comparison with Anglo-American management control frameworks. J Man Contr 23 (4):269–290

Guenther T, Gonschorek T (2011) Wertorientierte Unternehmensführung im deutschen Mittelstand – Ergebnisse einer empirischen Studie, (Value based management in mid-sized firms - results of an empirical study). Controlling 23(1):18–27

Guenther T, Heinicke A (2013) Die Nutzung von Kennzahlen und Instrumenten zur Unternehmensführung im Mittelstand – Erste Ergebnisse einer empirischen Untersuchung (The use of performance measures and instruments for management control in mid-sized firms – first results of a survey), Dresdner Beiträge zur Betriebswirtschaftslehre, No. 167/2013

Guenther T, John S, Schmidt U (2013) Controllingsysteme der deutschsprachigen Hochschulen in Deutschland, Österreich und der Schweiz – Erste Ergebnisse einer empirischen Untersuchung (Management control systems in German speaking HE institutions in Germany, Austria and Switzerland – first results of an empirical investigation), Dresdner Beiträge zur Betriebswirtschaftslehre, No. 174/2013

Harman G (2002) Academic leaders or corporate managers: deans and heads in Australian higher education, 1977 to 1997. High Educ Manag Policy 14(2):53–70

Henri JF (2006) Management control systems and strategy: a resource-based perspective. Account Organ Soc 31(6):529–558

Hood C (1995) The "New Public Management" in the 1980s: variation on a theme. Account Organ Soc 20(2/3):93–109

Ittner CD, Larcker DF (2001) Assessing empirical research in managerial accounting: a value-based management perspective. J Account Econ 32:349–410

Langfield-Smith K (1997) Management control systems and strategy: a critical review. Account Organ Soc 22(2):207–232

Langfield-Smith K (2007) A review of quantitative research in management control systems and strategy. In: Chapman CS, Hopwood AG, Shields MD (eds) Handbook of management accounting research, vol 2. Elsevier, Amsterdam, pp 753–783

Lawrence S, Sharma U (2002) Commodification of education and academic labour, using the balanced scorecard in a university setting. Crit Perspect Account 13:661–677

Malmi T (2001) Balanced scorecards in Finnish companies: a research note. Manag Account Res 12(2):207–220
Malmi T, Brown DA (2008) Management control systems as a package. Opportunities challenges and research directions. Manag Account Res 19(4):287–300
Merchant KA, Van der Stede WA (2011) Management control systems, performance measurement, evaluation and incentives, 3rd edn. Person, Harlow
Modell S (2003) Goals versus institutions: the development of performance measurement in the Swedish university sector. Manag Account Res 14(4):333–359
Modell S (2006) Institutional and negotiated order perspectives on cost allocations: the case of the Swedish university sector. Eur Account Rev 15(2):219–251
Otley DT (1980) The contingency theory of management accounting: achievement and prognosis. Account Organ Soc 5(4):413–428
Parker L (2011) University corporatization: driving redefinition. Crit Perspect Account 22:434–450
Pollitt C (2002) Clarifying convergence: striking similarities and durable differences in public management reform. Public Manag Rev 4(1):471–492
Simons R (1995) Levers of control. Harvard Business School Press, Boston
Speckbacher G, Wentges P (2012) The impact of family control on the use of performance measures in strategic target setting and incentive compensation: a research note. Manag Account Res 23(1):34–46
Vandenbosch B (1999) An empirical analysis of the association between the use of executive support systems and perceived organizational competitiveness. Account Organ Soc 24 (1):77–92
Widener SK (2007) An empirical analysis of the levers of control framework. Account Organ Soc 32(7/8):757–788

The Suitability of Intellectual Capital Reports for the Quantitative Measurement of Overall University Performance

Otto A. Altenburger and Michaela M. Schaffhauser-Linzatti

Abstract Universities are exposed to significant pressures from public and private stakeholders to steadily increase their academic performance in research, teaching, and industrial cooperation. The Intellectual Capital Report has been regarded as a suitable reporting tool to measure, publish, and evaluate their efforts and success. It allows for integrating non-monetary values on intangible assets such as human, structural, and relational capital. In 2002, Austria has been the first country worldwide to introduce the annual publication of an Intellectual Capital Report for public universities obligatorily. In spite of its broad design and innovative, modern characteristics it is to question whether this instrument is suitable not only to verbally describe university performance but also to provide a basis for quantitative measurement and comparison. Such quantitative approaches comprise multiple statistical instruments, Data Envelopment Analysis, and — maybe somewhat aside — fuzzy logic approaches. We reveal that the Intellectual Capital Report applied for Austrian universities cannot provide an adequate data base for these methods so far. Consequently, we suggest that it should either be adapted to the necessary formal requirements or undergo a deep discussion on its general usefulness and relevance.

1 Introduction

Management of private enterprises as well as of public organizations might induce the ongoing development of specific reporting tools to meet the current information needs relevant to their decision-making processes. Reflecting management information systems from a different angle, not only management creates new reporting instruments, but the introduction of such new reporting instruments also conversely leads to changes in the information policy and in the management behavior (Rauner and Schaffhauser-Linzatti 2002). Consequently, the introduction of innovative reporting tools needs to undergo a critical discussion from the very beginning in

O.A. Altenburger • M.M. Schaffhauser-Linzatti (✉)
University of Vienna, Vienna, Austria
e-mail: michaela.linzatti@univie.ac.at

I.M. Welpe et al. (eds.), *Incentives and Performance*,
DOI 10.1007/978-3-319-09785-5_23

which way they might contribute to better stakeholder information, improved decision making, and finally increased efficiency and quality of the institutions concerned.

The last decades faced some hypes about reporting tools that reflected the spirit of the times and the social development mainly of the industrialized economies. Among others, very heterogeneous innovations such as social accounting (Ungureanu 2012), environmental accounting (Artene et al. 2012), and Intellectual Capital Reports (Edvinsson and Malone 1997) emerged. The social and environmental balance sheets never really succeeded and nearly vanished; they left behind some residual information in already existing tools or still undergo a discussion whether accounting or simple reporting should be preferred. By contrast, some ideas such as the Balanced Scorecard (Kaplan and Norton 1992), which indeed pours old wine into new skins, began its sweep through the world of operational business (Schaffhauser-Linzatti and Faisst 2003).

We are now living in the age of the so-called knowledge society (Drucker 1993). The increasing relevance of intellectual values (Stewart 1998) led to the development of a new reporting movement which emerged under the denomination of Intellectual Capital Reporting. Intellectual Capital Reports comprise quantitative as well as qualitative approaches which reflect the disciplines and perspectives of their proponents. Austria has been the first country in which an Intellectual Capital Report has been demanded by law. It must be published annually by all Austrian public universities and might serve as a role model for further developments worldwide.

Universities produce intellectual values which have been made public since ever. Due to their intangible characteristics, these performances, however, cannot be standardized and evaluated by means of regular accounting or finance procedures. That is why universities serve as an excellent field of implementing alternative reporting instruments to present their results in academic research and teaching. From a management perspective, the Intellectual Capital Reports intend to deliver the necessary information to support universities' decision makers, to increase performance, and to reallocate their tight resources. Hence, these reports might be an impetus to reconsider academic management behavior and strategies. To reflect the general application, structure, and future relevance of Intellectual Capital Reports the Austrian Intellectual Capital Reports will be subsequently used as an example to derive strengths and weaknesses of this instrument and to draw general conclusions.

The article is structured as follows: Sect. 2 presents a short overview over the development and the different approaches of Intellectual Capital Reports and relates their characteristics to the main requirements of university reporting. In Sect. 3, we analyze how the Austrian Intellectual Capital Reports and their predecessors have entered quantitative approaches. We then derive pitfalls to avoid when installing universities' Intellectual Capital Reports. Section 4 confirms the Austrian concerns and adds international critics. Section 5 concludes.

2 Approaches of Intellectual Capital Reports for Universities

In a very broad definition, Intellectual Capital Reports are applied to collect, measure, and present information on Intellectual Capital for documentation, management, and controlling purposes. They address internal and external stakeholders (Schaffhauser-Linzatti 2003).

Internally, public institutions such as universities may apply Intellectual Capital Reports for a better resource allocation, efficiency measurement, or personnel deployment planning and regard them as a supplement to the already established instruments of cost accounting and controlling which may replace Balanced Scorecards (Schaffhauser-Linzatti and Faisst 2003). However, the main emphasis of the Intellectual Capital Reports is information for external stakeholders. Universities try to demonstrate their performance by listing past activities; their main products are the creation and the reproduction of knowledge.

As can be easily proven, annual reports derived from financial accounting are not sufficient to reflect the intellectual assets and intellectual output of universities (Schaffhauser-Linzatti 2005). According to the narrow regulations of the current law (Maul 2003), extended since then in Germany, but not in Austria, only acquired patents and licenses are included in financial statements. In contrast, Intellectual Capital Reports also aim at communicating intellectual achievement potentials and intellectual performances and allow for a flexible publication of the universities' individual emphases and characteristics. However, these intended targets and advancements did not lead to one single, generally accepted approach so far. This implies that no satisfying and overall applicable concept has been finally agreed upon by academics and practitioners, which is further reflected by the fact that Intellectual Capital Reports have not entered business and governmental reporting on a broad basis. The following rough classification into two diverse directions proves the heterogeneity of the Intellectual Capital Report models.

First, the older approaches — also referred to as Human Resource Accounting — are built upon monetary values to evaluate intellectual assets and to put them on the balance sheet. For an overview see Cañibano et al. (2000). So far, none of these approaches have been successfully implemented due to the high need of information and uncertain information quality.

Second, the more recent approaches reject the idea of monetary evaluation. They try to visualize Intellectual Capital and performances in a quantitative and qualitative way. These models have in common that they are more or less structured like ratio systems or — even simpler — as loosely related statistical information. That is why they are regarded as an applicable solution for universities. The new era of non-monetary Intellectual Capital Reports started with the Skandia Navigator (Edvinsson 1997), followed by the Intellectual Assets Monitor (Sveiby 1997), the Intellectual Capital Index (Roos and Roos 1997), the Calculated Intangible Value (Stewart 1998), the Market Value Added (Lev and Sougiannis 1996), and the approach developed by the Austrian Research Centers, the so-called ARC-model (Leitner 2005).

The ARC-model meets the requirements of New Public Management (Thom and Ritz 2000) by integrating outputs and goals of a non-profit, public organization to enable performance-oriented controlling of its processes. It was the role model for the Intellectual Capital Report entering Section 13 (6) of the Austrian Universities Act 2002, (UG 2002). This Intellectual Capital Report has, as a minimum, to present: (1) the sphere of action, social goals and self-imposed objectives and strategies of the university; (2) its Intellectual Capital, broken down into human, structural and relational capital; (3) the performance processes set out in the performance agreement, including their outputs and impacts. Structure and design of the Intellectual Capital Reports are regulated by a decree called Wissensbilanz-verordnung of the responsible Federal Minister. Such a decree was published first in 2006, followed by a new decree in 2010, slightly amended during the last years (Altenburger et al. 2012).

3 Performance Studies for Austrian Universities

3.1 Overview over the Studies

Apart from the qualitative, verbal information that can also be given in any other form of publication, Intellectual Capital Reports must allow for quantitative analyses in order to operationalize future steps. To reveal whether the obligatory Intellectual Capital Reports meet such criteria and whether the data base has been improved, we conduct a qualitative meta-comparison of studies applying two different reporting tools: first, the so-called ABIV-studies based on the ABIV (work report of the head of department) before the implementation of the Universities Act 2002 until 2005; and second, the so-called ICR-studies based on the universities' Intellectual Capital Report introduced 2006. Table 1 summarizes the references of the included ABIV-studies and ICR-studies and their methodological approaches.

The ABIV did not represent any form of Intellectual Capital Report, but was a simple questionnaire with 48 indicators which had to be filled out manually by each researcher; the results were accumulated by the departments and forwarded to the dean. The ABIV-studies refer to these data of the years 2000 and 2001. On the one hand, an aim of the analyses was to evaluate universities and departments according to their efficiency. Statistical analyses (see Sect. 3.2) were applied to filter out appropriate ratios which were entered into a Data Envelopment Analysis (DEA; see Sect. 3.3). On the other hand, the results of the ABIV-studies revealed indicators which are indeed meaningful and do not just add to any administrative data graveyards. These indicators could have been identified as useful data base for the Intellectual Capital Report, however, they did not significantly enter the final

Table 1 Studies and approaches

ABIV—Arbeitsbericht des Institutsvorstandes	ICR—Intellectual Capital Report
Studies included in this comparison	
Leitner et al. (2005, 2007)	Altenburger and Schaffhauser-Linzatti (2007, 2009), Altenburger et al. (2009, 2012)
	Schaffhauser-Linzatti (2003, 2005, 2008, 2010, 2011)
	Veltri et al. (2009)
Quantitative approaches applied by these studies	
Correlation analyses (Pearson, Kendall τ)	Correlation analyses (Pearson, Kendall τ)
Regression analyses	Regression analyses
Factor analyses	Factor analyses
–	Stochastic Frontier Analysis
DEA-analyses	–
–	Fuzzy logic expert system

Wissensbilanzverordnung. Now, the Intellectual Capital Report demands 28 ratio categories (except for medical universities) with up to 44 stratification factors which can be further subdivided and which is consequently different at each university. As an example, field of studies is regarded as one stratification factor — imagine the University of Vienna with about 180 study programs!

The ICR-studies tried to retrace the ABIV-studies with the new data material to figure out which of the proven approaches are still realizable. As it soon became obvious that the new data base did not allow for many methods such as DEA, the ICR-studies added new performance measurement approaches such as fuzzy logic (see Sect. 3.3).

Table 2 contrasts the ratios and ratio categories that have been used in the ABIV-studies and the ICR-studies in general after adjusting for correlations. The listing of the data used in the ICR-studies follows the structure of the Wissensbilanz-verordnung without the detailed and voluminous stratification factors.

The following subsections describe the quantitative approaches applied and summarize main results to prove their applicability. For any detailed results we must refer the reader to the original papers.

3.2 *Statistical Analyses*

The statistical analyses discussed here were applied both in the ABIV-studies and the ICR-studies. Correlation analyses measure linear relations between two variables without revealing the direction of cause and effect; regression analyses measure whether and how variables are related without revealing the direction of

Table 2 Indicators applied

ABIV—Arbeitsbericht des Institutsvorstandes	ICR—Intellectual Capital Report
Overview of indicators and indicator categories used	
48 indicators in total	28 ratio categories (+5 for medical universities)
	Up to 44 stratification categories per ratio category
–	1. Intellectual Capital
–	A. Human Capital
Staff (input variable)	Staff
–	Number of habilitations
–	Number of appointments
–	Percentage of women
–	Gender pay gap
–	B. Relational Capital
–	Outgoing staff (at least 5 days)
–	Incoming staff (at least 5 days)
–	C. Structural Capital
–	Number of cooperations with other institutions
Financial funds provided by third parties (industry)	R&D project income
Finished projects ad personam (industry)	–
Finished projects of the department (research)	–
–	2. Core processes
–	A. Teaching and further education
–	Teaching load of staff
–	Number of studies
–	Average duration of studies
–	Applicants with special admission conditions
–	Number of students (bachelor, master)
–	Number of enrolled studies
–	Number of outgoing students in exchange programmes
–	Number of incoming students in exchange programmes
–	Success rate of students
–	B. Research and arts
–	Research staff
–	PhD students employed at the university
–	3. Output
–	A. Teaching and further education
–	Number of degrees
–	Number of degrees within tolerance period
–	Number of degrees of students with studies abroad

(continued)

Table 2 (continued)

ABIV—Arbeitsbericht des Institutsvorstandes	ICR—Intellectual Capital Report
–	B. Research and arts
Monographs (research)	Number of publications
Journal papers (research)	–
Project reports (research)	–
Other publications (research)	–
Presentations (research)	Number of presentations
Space (in m^2, input variable)	–
Examinations (teaching)	–
Finished supervised diploma theses (teaching)	–
Finished supervised PhD theses (research)	–
Patents (research)	(To be reported outside ICR)

this relationship (e.g., Chatterjee and Hadi 2012). Among others, Firer and Williams (2003) reveal the relation of Intellectual Capital and business performance, Antonella and Veltri (2011) or Bonaccorsi and Daraio (2002) the relation of Intellectual Capital and university performance.

In the ABIV-studies, all then requested indicators were correlated applying Pearson and Kendall τ b correlations due to different scales of the data to reduce redundant variables on a university and on a faculty level. The remaining indicators were then grouped according to possible departmental emphases, teaching, research, and industrial cooperation. The ratios presented in Table 2 refer to the sub-studies on the 133 technical and natural science departments in seven public Austrian universities. The results fulfilled most of the expectations experienced in daily university life. Roughly summarized, we identified the following correlations: among all categories of staff (e.g., professors, assistants); between staff and output; between staff and PhD-students; between staff and research output; between staff and industrial cooperation; and between research and teaching. Interestingly, no correlations were found between departmental size and teaching; departmental size and research performance; and departmental size and third-party funds. The frequently used ordinary least square regression has also been applied to explain differences in indicator variances. In addition to the two input variables staff and space (m^2), a subsequent factor analysis verified the attribution of the output variables to the three emphases research, teaching, and industrial cooperation (see Table 2, allocation in brackets) which also corresponds to international findings (Salerno 2003).

In the ICR-studies correlation analyses to reveal redundant variables did not lead to any tangible results as over all disciplines at each university nearly all ratios were correlated. This fact proves that a differentiation among fields of research and study is necessary but overall not possible due to differently structuring stratification factors of the Intellectual Capital Report (Altenburger et al. 2012). Standardized methods comprising simple ratio analyses, linear regression, factor analyses, and Stochastic Frontier Analysis were chosen to highlight single topics of interest in order to reveal application possibilities of the Intellectual Capital Report, e.g., gender influences on university performance (Altenburger et al. 2009; Schaffhauser-Linzatti 2011).

3.3 Data Envelopment Analysis

A DEA could only be performed by the ABIV-studies (see Table 1). Structural problems with ICR data did not allow for calculating DEA in the ICR-studies.

DEA has been developed as a non-parametric model based on the method of fractional linear programming. Being a management and consultancy tool it evaluates and benchmarks the efficiency of different, independent decision-making units acting under the same technology. Based on estimated differences between observed values and an ideal production frontier a decision-making unit is called efficient as long as no other unit exists that reaches the same output with less input or produces more output with the same input, respectively (Charnes et al. 1978). DEA allows for ranking decision-making units and for specifying improvement potentials in regard to changes in specific inputs, outputs, and their combinations. That's why many industries apply DEA successfully; among others banks (Wheelock and Wilson 1995) or start-up companies (Fuertes Callen et al. 2005). Being independent of prices, data units, and predefined weights of inputs and outputs, DEA is especially valuable for non-profit organizations (Leitner 2005; Ahn and Seiford 1990); applications for universities are run by, e.g., Bonaccorsi and Daraio (2002), Fandel (2003), Sinuany-Stern et al. (1994) or Wagner (2002).

DEA enables to enter qualitative and quantitative information and variables of different scales into the same calculation. It also allows for multiple input and output ratios in contrast to simple performance ratios; the Golden DEA rule requires that the number of inputs and the number of outputs should be smaller or equal the number of decision-making units divided by three. As the independence and number of ratios and decision-making units selected by the statistical analyses before (see Sect. 3.2 and Table 2) fulfilled all information requirements, the ABIV-studies could perform DEA successfully.

The results of DEA are presented as percentages with 100 % for a totally efficient decision-making unit. Therefore, DEA enables the ABIV-studies to rank the departments according to their overall efficiency and to single scores in research, teaching, and industrial cooperation. The voluminous results of the ABIV-studies are summarized in Leitner et al. 2007. Overall, about half of the departments included into the study were efficient with an efficiency score of at least 85 %. Besides a simple ranking of the efficiency of the Austrian universities, four additional categories of differences were identified. First, significant regional differences within an identical legal framework; second, varying differences in the defined fields of study between the periods under observation (2001 and 2002); third, the influence of the department size, where small and large departments overperform (e.g., Gander 1995; Johnes and Johnes 1993); and at last unexpected differences among the specialization on industry, teaching, research, or a wide range of university tasks. To prove the robustness of the results, different DEA approaches by Charnes et al. (1978) and Banker et al. (1984) and a subsequent sensitivity analysis were calculated. A Malmquist-Index for long-term observations was not applied due to the short observation period, see Fare et al. 1994.

3.4 Fuzzy Logic

The fuzzy logic approach has only been applied as part of the ICR-studies (Veltri 2007; Veltri et al. 2009) and was never used for the ABIV-studies. The fuzzy expert system approach has been chosen because of its flexibility to adapt the quantitative as well as qualitative indicators in an individual approach.

This cognitive framework reproduces experts' perceptions on complex problems and formalizes non-quantitative concepts in such a way that they can be processed by computers. Experts enter knowledge into a knowledge base in symbolic form, mainly by rule blocks. An inferential engine derives numerical values out of this knowledge base by using logic and heuristics (Magni et al. 2006). To do so, first "if-then" rules are derived by multiple experts acting under incomplete and imperfect information with reduced complexity (Zadeh 1978), e.g., "*The more X is Y, the more certain A is B*". The fuzzy logic approach transfers these assumptions into a formal model and delivers a numerical value (Dubois and Prade 1980). Finally, the results have to be tested for their reasonability, if required adapted, and validated by sensibility analyses. The application of fuzzy systems in economics is steadily increasing and comprises, among others, finance and investment (Magni et al. 2004), accounting (Zebda 1991), economics, or insurance (Facchinetti et al. 2001).

The fuzzy logic expert system by Veltri (2007; Veltri et al. 2009) evaluates the Intellectual Capital of universities (for further studies on this focus see also Bozbura et al. 2007 and Tai and Chen 2009). The basic structure of the model has been developed as a general approach to develop a fuzzy expert system for universities' Intellectual Capital. To prove its applicability and reliability, it was empirically applied to the Intellectual Capital Reports of the Austrian public universities in 2006. The selected indicators follow the structure of the Austrian universities' Intellectual Capital Reports (see Table 2). The model is structured as a decision tree which runs from branches to the trunk. It measures the value-creation power of the university's Intellectual Capital. The final output is determined by human, structural, and relational capital which again depend on further variables, cascading backwards until the final input variables, the value drivers, are reached. These value drivers are combined by "if-then" rule blocks, by modularization and by creating intermediate variables; finally the IC Index with a single value between 0 and 100 is derived.

The results of the comprehensive IC Index deliver a ranking of Austrian universities. Further disassembled indices split up the index into the three main input categories human, structural, and relational capital at the next level of the decision tree, and such illustrate how the overall performance is compound and where strengths and weaknesses occur.

3.5 Lessons Learnt from the Austrian Studies

This subsection derives the most critical pitfalls and weaknesses of the Intellectual Capital Report by referring to the experience of the ABIV-studies and the ICR-studies. It serves as a basis for Sect. 4 to add to general data requirements and problems revealed in the literature.

As shown in Table 2, correlation analyses reduced the original 48 ABIV-indicators to 12 independent indicators which were comparable among clearly defined fields of studies and research. These indicators, being clear and manageable, allowed for applying the most important quantitative instruments. The ABIV-successor, the current Intellectual Capital Report, includes only 28 ratios (except for medical universities) which are, however, further deepened by stratification factors such as gender, nation, or field of studies. Hence, the final number of indicators to be reported depends on the university and its range of tasks; an extrapolation of the first Wissensbilanzverordnung (see Sect. 2) came up to a maximum of 32,000 single data per year (Schaffhauser-Linzatti 2010) which have been reduced by further amendments.

The simple reduction of indicators and stratification factors, however, did not improve the structure of the Intellectual Capital Report substantially. The summary of the ICR-studies above showed that in spite of a voluminous data collection the most meaningful quantitative approaches cannot be performed anymore. Among others, simple statistical analyses show that relations between research and teaching could not be calculated due to not-matching classifications of fields of studies and fields of research. Moreover, the Golden DEA rule is not fulfilled any more. Therefore, the efficiency of departments and universities cannot be evaluated by sophisticated instruments. The most problematic weaknesses of the Intellectual Capital Report in comparison to the ABIV are:

- Choice of ratios: The demanded indicators are significantly correlated and not independent which prohibits the application of regression analyses. A reduction to the most significant indicators would help. Also, most indicators are of quantitative, but non-financial nature, although Intellectual Capital Reports should keep the balance between numerical and verbal information. As experienced during work on the fuzzy logic model, soft elements such as employees' satisfaction or leadership abilities are missing. Apart from calculations, the relation between teaching and research indicators is not balanced. As the Intellectual Capital Report is used for resource allocation, strategic engagement towards the most profitable indicators might lead to a biased university performance in the long run.

- Definition and stratification factors: In spite of the numerous stratification factors, comparable and additional important information is missing. For example, correlation analyses indicated that not only number of staff, but also qualification of employees matter, which is not demanded by the Intellectual Capital Report. Nationality of students and staff is not given consistently. Missing data depth does not allow for many regressions or DEA because of the Golden DEA rule. Most of all, the classifications of scientific disciplines and of studies do not match. Consequently, any comparison of teaching and research or assignment of academic staff, e.g., correlations between students and teachers, is made impossible. Some classifications are not even harmonized with other university-related statistics.
- Time series of data: Due to frequent adjustments of the required data, long-term studies or comparative studies are not possible.

4 Lessons to Be Learnt

4.1 Overview

The Austrian experience shows that quality of data must beat quantity, that the cost-benefit ratio improvements by introducing new approaches must exceed the drawbacks of former systems, and that data applications and analyses have to be planned jointly with the design of the new tools. When leaving aside the applicability of the indicators in different quantitative approaches and going a step backwards to the required data, a comprehensive literature review revealed that exactly such topics are not addressed internationally on a broad theoretical basis and gives room for two questions. First, vast and steadily increasing literature concentrates on structure, applicability and sense or absurdity of evaluations and rankings of universities (Krull 2011 suggests to regard them as part of the entertainment industry), but still, only few publications can be found on a general approach how to identify adequate indicators (e.g., Bonaccorsi and Daraio 2007); the majority of related publications only describes how ratios are used in specific institutions or for specific purposes (for a comprehensive overview see Slunder 2008). Second, no publication has been found about auditing, i.e., quality assurance, of such data.

4.2 Identification of Adequate Indicators

In the rapidly changing socio-political context of universities, indicators are mainly classified according to the three categories: teaching, research, and industrial cooperation as connection between university research and external economies (European Commission 2005); as an exception, ABET (2006) adds administrative levels. García-Aracil and Palomares-Montero (2009) demonstrate the missing consensus about indicators and classification criteria in the OECD countries (see, e.g., OECD 2007; ENQA 2007; CIHE 2007). According to Sizer et al. (1992), indicators enter five primary uses: monitoring, evaluation, dialogue, rationalization, and allocation of resources. However, most of the indicator systems do not address the target they are implemented for and consequently steer the discussion into the wrong direction.

The general classification as well as most indicators proposed in the literature match with the Austrian findings in the ABIV-studies (Leitner et al. 2007). But one of the main problems is the allocation of indicators to input or output categories as well as their allocation to teaching or research. Input is in general defined as staff and, partially, infrastructure, output always comes back to different forms of expenditure, publications, and licenses (for measuring the performance of research see, e.g., Berghoff et al. 2007 or an alternative approach by Hofmeister 2011; for industrial cooperation Gulbrandsen and Smeby 2005). However, these classifications depend on the fundamental view whether teaching and research are complementary functions or rivalling academic activities (e.g., Sarrico et al. 1997).

Still unsolved is the problem of combining qualitative and quantitative indicators and when to use absolute and relative values. Also, multidimensional and single ratios lead to diverging results, as proven among others by Turner (2005). Here, the CHE approach strikes out on new paths (Berghoff et al. 2008). Only few analyses include external indicators such as employability, knowing about the pitfalls and data problems (Diem and Wolter 2011).

Consequently, we learn that, apart from their application area, indicators literally only indicate, but do not exactly measure performances. Hence, we cannot expect precision of information, only approximations of a stated target.

4.3 Auditing

Every quantitative result is only as good as the data basis used. Although intra-university control mechanisms have been installed, many data, also resource-relevant ones, are not externally audited which induces two implications:

First, it may lead university managers into temptation to present the own unit in a favorable way by forwarding generally correct, but sugar-coated information. On the other hand, the ambiguity of a strict auditing or other forms of performance control like ranking procedures may foster a biased reactivity by strategic selection of conducted performances (Leeuw 1996). This so-called performance paradox has been revealed not only in for-profit organizations (e.g., Kaplan and Norton 1992), but also in the course of New Public Management studies (e.g., Van Thiel and Leeuw 2002), among others for educational institutions (Espeland and Sauder 2007).

Second, data may be incorrect due to either unintentional errors or intentional fraud. The auditing profession has developed and still improves instruments to reveal such irregularities, e.g., the "ten commandments" to address fraud during the audit (IDW 2012). These tools (e.g., Terlinde 2005; Sancar 2012) can be easily adapted for universities for internal as well as external audits.

Consequently, we learn that the data for such reports should be audited by external institutions. At least, an obligation for detailed internal audits should be implemented. Without auditing of data any evaluation performance measurement must be questioned.

On a more general level, the efficiency of implementing sophisticated performance indicator systems could be subjected to the judgmental scrutiny of the scientific community (for the discussion in business see Gaither 1994), if the search for the identification of valid indicators does not seem to lead to satisfying results. Moreover, their usefulness for conducting meaningful quantitative approaches has to be kept in mind simultaneously. This step has not been focused so far.

Summary and Conclusions

This paper derived requirements for reporting instruments on Intellectual Capital focusing on indicators to evaluate universities' performances. To do so, we compared two different series of studies on performance measurement of Austrian universities. The results of the single studies remain outside this research (see Table 1 for the series and their single studies).

The two series were based on two different reporting tools: the former ABIV was replaced by the Intellectual Capital Report according to the Universities Act 2002. Both series first tried to filter out representative indicators for the fields of teaching, research, and industrial cooperation. Then, these indicators entered sophisticated quantitative instruments to evaluate these three fields individually and combined. The ICR-studies were originally designed to reproduce the results of the ABIV-studies to enlarge the long-term view of the universities' development. It soon became clear that the new data structure did not allow for retracing the former studies. Hence, alternative, but less powerful tools were applied to gain information on universities' performances.

The impossibility to apply useful and meaningful approaches gave the impetus to this qualitative meta-analysis which adds to the literature in two aspects. We revealed which indicators could be successfully used for Austrian universities. On the one hand, we found out that the, at first glance, few

(continued)

indicators listed in Table 2 in the ABIV-column are sufficient to install sophisticated quantitative instruments which lead to significant and applicable results. On the other hand, the highly praised Intellectual Capital Report turned out as not useful for deep quantitative analyses, but only for statistical reasons. Moreover, it demands too many indicators and enforces excessive bureaucracy. A reasonable cost-benefit ratio cannot be certified.

The main findings of this meta-analysis point to apparent results which are simple and already well-known, but often neglected. Well-structured, clear, and yet simple statistical instruments which may be wrongly blamed for being old-fashioned can be very effective and efficient whereas complicated models in spite of their theoretical appropriateness and international reputation may follow the contemporary trend of alternative reporting tools, but do not work in the intended way. Most of all the structure of each reporting tool must be consistent in its definitions and stratification factors.

Limitations of our main findings result from the naturally small sample of Austrian universities and the legally predefined data structure. The missing comparability of the studies reveals weaknesses and pitfalls of the Intellectual Capital Report, but does not allow for quantitative analyses on the differences to other reporting tools which tend to lead to more scientific explanatory power. In its existing structure, no more in-depths analyses to gain further results can be conducted. Consequently, further research shall include statistical and comparative qualitative content analyses of reporting tools outside Austria to lay a broader base for filtering out adequate indicators for measuring university performance. Practical implications of this study are ongoing discussions with the legislative bodies to restructure, standardize, and simplify the Intellectual Capital Report in order to display its inherent, but still concealed strengths.

We are aware of the fact that there does not exist one single, perfect solution. However, scientists and university managers steadily have to strive for improved solutions by conducting qualitative and quantitative evaluations and analyses on past evaluations. Lessons to be learnt by standard setters in Austria and beyond are, first, that one should not skip well-tried strategies in favor of non-tested experiments without any need. Second, according to “structure follows strategy” the instalment of a data collecting and presenting tool should offensively address the question of future interpretation and analyzing possibilities.

References

ABET (2006) Accreditation policy and procedure manual. Effective for evaluations during the 2007–2008 accreditation cycle. Available via ABET. http://www.abet.org/

Ahn T, Seiford LM (1990) Sensitivity of DEA to models and variables sets in a hypothesis test setting: the efficiency of university operations. In: Ijiri Y (ed) Creative and innovative approaches to the science of management. Quorum Books, Westport, pp 191–208

Altenburger OA, Schaffhauser-Linzatti MM (2007) Instrumente zur Steuerung von intellektuellem Kapital – eine Analyse anhand oesterreichischer Universitaeten. In: 9. Workshop der Kommission "Hochschulmanagement" im Verband der Hochschullehrer fuer Betriebswirtschaft e.V. (VHB), Muenster, 23–24 Feb 2007

Altenburger OA, Schaffhauser-Linzatti MM (2009) Ratios on intangibles as an instrument for budget allocation: conceptual thoughts about improvements for Austrian Public Universities. In: 4th international forum on knowledge asset dynamics and intellectual asset centre congress, University of Glasgow, Glasgow, 17–18 Feb 2009

Altenburger OA, Podda P, Schaffhauser-Linzatti MM (2009) Gender aspects in the Austrian Universities' Intellectual Capital Reports. In: 5th EIASM workshop on visualising, measuring, and managing intangibles and intellectual capital, Dresden, 8–9 Oct 2009

Altenburger OA, Schaffhauser-Linzatti MM, Pablik E, Schmidt J (2012) Zehn Jahre Verankerung der Wissensbilanz im oesterreichischen Universitaetsgesetz 2002, Fachtagung Innovation, Leistungsmessung und Anreizsysteme in Wissenschaft und Wirtschaft – Governance wissensintensiver Organisationen, Technische Universitaet Muenchen, Munich, 27–28 Nov 2012

Antonella S, Veltri S (2011) The intellectual capital report within universities: comparing experiences. The Annals of University of Oradea, Economic Sciences, 2nd issue, pp 626–663

Artene AE, Domil AE, Bostan I, Anale CR (2012) Implementing environmental balance sheet within European private companies (SPE). Seria Ştiinţe Econ XVIII:710

Banker RD, Charnes A, Cooper WW (1984) Some models for estimating technical and scale inefficiences in data envelopment analysis. Manag Sci 30:1078–1092

Berghoff S, Federkeil G, Giebisch P, Hachmeister C-D, Hennings M, Mueller-Boeling D (2007) CHE_ForschungsRanking_2007_AP102. CHE, Guetersloh

Berghoff S, Federkeil G, Giebisch P, Hachmeister C-D, Hennings M, Mueller-Boeling D, Roessler I (2008) CHE-HochschulRanking – Vorgehensweise und Indikatoren, Arbeitspapiere 88. CHE, Guetersloh

Bonaccorsi A, Daraio C (2002) The organization of science: size, agglomeration and age effects in scientific productivity. In: SPRU NPR-net conference: rethinking science policy: analytical frameworks for evidence-based policy. SPRU, University of Sussex, Brighton, 21–23 Mar 2002

Bonaccorsi A, Daraio C (2007) Universities as strategic knowledge creators: some preliminary evidence. In: Universities and strategic knowledge creation. Edward Elgar, Northampton, pp 31–81

Bozbura FT, Beskese A, Kaharaman C (2007) Prioritization of human capital measurement indicators using fuzzy AHP. Exp Syst Appl 32:1100–1112

Cañibano L, Garcia-Ayuso M, Sanchez P (2000) Accounting for intangibles: a literature review. J Account Lit 19:102–130

Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. Eur J Oper Res 2:429–444

Chatterjee S, Hadi A (2012) Regression analysis by example. Wiley, New Jersey

CIHE (2007) Standards for accreditation. Commission on institutions of higher education. New England Association of Schools and Colleges. http://www.neasc.org/cihe/standards_for_accreditation_2005.pdf

Diem A, Wolter SC (2011) Beurteilung von Hochschulen anhand der Arbeitsmarktfaehigkeit ihrer Absolventen und Absolventinnen – Eine Analyse der Schweizer Universitaeten. Schweizerische Koordinationsstelle fuer Bildungsforschung, Aarau

Drucker P (1993) Die postkapitalistische Gesellschaft. Econ Verlag, Duesseldorf

Dubois D, Prade H (1980) Fuzzy sets and systems: theory and applications. In: Fuzzy logic implementation and applications. Academic Press, New York

Edvinsson L (1997) Developing intellectual capital at Skandia. Long Range Plann 30:366–373

Edvinsson L, Malone MS (1997) Intellectual capital: realizing your company's true value by finding its hidden brainpower. HarperBusiness, New York

ENQA (2007) Standards and guidelines for quality assurance in the European Higher Education Area. ENQA. http://www.enqa.eu/files/ESG_v03.pdf

Espeland WN, Sauder M (2007) Rankings and reactivity: how public measures recreate social worlds. Am J Soc 113:1–40

European Commission (2005) European universities: enhancing Europe's research base. European Commission, Brussels

Facchinetti G, Cosma S, Mastroleo G, Ferretti R (2001) A fuzzy credit rating approach for small firm bank creditworthiness. An Italian case. In: Proceedings of CIMA 2001, international ICSC-NAISO congress on computational intelligence: methods & applications, Bangor, Wales, 19–22 June 2001

Fandel G (2003) Zur Leistung nordrhein-westfaelischer Universitaeten. Gegenueberstellung einer Verteilungsloesung und der Effizienzmaße einer Data Envelopment Analysis. In: Backes-Gellner U, Schmidtke C (eds) Hochschuloekonomie – Analysen interner Steuerungsprobleme und gesamtwirtschaftlicher Effekte. Duncker & Humblot, Berlin, pp 33–50

Fare R, Grosskopf S, Knox-Lovell CA (1994) Production Frontiers. Cambridge University Press, London

Federal Ministry of Education, Science and Culture UG (2002) Bundesgesetz ueber die Organisation der Universitaeten und ihre Studien (Universitaetsgesetz 2002 – UG). BGBl. I Nr. 120/2002

Federal Ministry of Education, Science and Culture (2010) Verordnung ueber die Wissensbilanz (Wissensbilanz-Verordnung 2010 – WBV 2010). BGBl. II Nr. 216/2010

Firer S, Williams M (2003) Intellectual capital and traditional measures of corporate performance. J Intell Cap 4:348–360

Fuertes Callen Y, Mar MC, Serrano CC (2005) Measuring DEA efficiency in internet companies. Decis Support Syst 38:557–573

Gaither G (1994) Measuring up: the promises and fitfalls of performance indicators in higher education. ASHE-ERIC, Higher education report, vol 5. George Washington University, Washington, DC

Gander JP (1995) Academic research and teaching productivities: a case study. Technol Forecast Soc Change 49:311–319

García-Aracil A, Palomares-Montero D (2009) Examining benchmark indicator systems for the evaluation of higher education institutions. High Educ 60:217–234

Gulbrandsen M, Smeby J-C (2005) Industry funding and university professors' research performance. Res Policy 34:932–950

Hofmeister R (2011) Measuring the value of research: a generational accounting approach. University of Konstanz, Konstanz

IDW Pruefungsstandard (2012) Zur Aufdeckung von Unregelmaeßigkeiten im Rahmen der Abschlusspruefung (IDW PS 210). Institut der Wirtschaftspruefer in Deutschland, Duesseldorf

Johnes G, Johnes J (1993) Measuring the research performance of UK economics departments: an application of data envelopment analysis. Oxf Econ Pap 45:332–347

Kaplan RS, Norton DP (1992) Balanced scorecard – measures that drive performance. Harv Bus Rev, Jan-Feb:70–79

Krull W (2011) Bewertung, Begutachtung und Evaluation in Wissenschaft und Forschung. In: Evaluation: new balance of power, iFQ-Working Paper No 9, Nov 2011, pp 15–25

Leeuw FL (1996) Performance auditing, new public management and performance improvement: questions and answers. Account Audit Account J 9:92–102
Leitner K-H (2005) Managing and reporting intangible assets in research technology organisations. R&D Manag 35:125–136
Leitner K-H, Schaffhauser-Linzatti MM, Stowasser R, Wagner K (2005) Data envelopment analysis as method for evaluating intellectual capital. J Intell Cap 6:528–543
Leitner K-H, Schaffhauser-Linzatti MM, Stowasser R, Wagner K (2007) The impact of size and specialisation on universities' department performance: A DEA analysis applied to Austrian Universities. High Educ 53:517–538
Lev B, Sougiannis T (1996) The capitalization, amortization and value-relevance of R&D. J Account Econ 21:107–138
Magni CA, Mastroleo G, Vignola M, Facchinetti G (2004) Strategic options and expert systems: a fruitful marriage. Soft Comput 8:179–192
Magni CA, Malagoli S, Mastroleo G (2006) An alternative approach to firm's evaluation: expert systems and fuzzy logic. Int J Inf Technol 5:195–225
Maul KH (2003) Moeglichkeiten und Grenzen der Quantifizierung von intellektuellem Vermoegen in Wissensbilanzen. In: Oesterreichische Rektorenkonferenz (ed) Wissensbilanz: Bilanz des Wissens? Oesterreichische Rektorenkonferenz, Vienna, pp 36–53
OECD (2007) Education at a glance. OECD indicators. OECD, Paris
Rauner MS, Schaffhauser-Linzatti MM (2002) Impact of the new Austrian in-patient payment strategy on hospital behaviour: a system-dynamics-model. Socio Econ Plan Sci 36:161–182
Roos G, Roos J (1997) Measuring your company's intellectual performance. Long Range Plan 30:413–426
Salerno C (2003) What we know about the efficiency of higher education institutions: the best evidence. Beleidsgerichte studies hoger onderwijs en wetenschappelijkonderzoek. University of Twente, Twente
Sancar B (2012) Instrumente zur Aufdeckung und Praevention von Wirtschaftskriminalitaet (doloser Handlungen) in Unternehmen. Grin Verlag GmbH, Norderstedt
Sarrico CS, Hogan SM, Dyson RG, Athanassopoulos AD (1997) Data envelopment analysis and university selection. J Oper Res Soc 48:1163–1177
Schaffhauser-Linzatti MM (2003) Zusammenfassung Wissensbilanz: Bilanz des Wissens? Die Wissensbilanz fuer Universitaeten im UG 2002. In: Oesterreichische Rektorenkonferenz (ed) Wissensbilanz: Bilanz des Wissens? Oesterreichische Rektorenkonferenz, Vienna, pp 78–86
Schaffhauser-Linzatti MM (2005) Welches Interesse verbindet Sie mit der universitaeren Wissensbilanz? In: Finanzwesen und Controlling der Universitaet Wien (ed) Wissensbilanzierung – Beitraege der Universitaet Wien. University of Vienna, pp 46–60
Schaffhauser-Linzatti MM (2008) Analyse von Indikatoren mit Hilfe der Data Envelopment Analysis am Beispiel von Universitaeten. In: 10 Jahre Wissensbilanz in Oesterreich. Austrian Academy of Sciences, Vienna
Schaffhauser-Linzatti MM (2010) Immaterielle Werte sind ewige Sorgenkinder des Bilanzrechts. Habilitationcolloquium, University of Vienna, 8 Oct 2010
Schaffhauser-Linzatti MM (2011) Die Wissensbilanz als innovatives Managementinstrument im oeffentlichen Bereich. In: Schauer R, Thom N, Holgers D (eds) Innovative Verwaltungen, Innovationsmanagement als Instrument von Verwaltungsreformen. Internationales Forschungscolloquium Public Management (PUMA-Forschungscolloquium), Eine Dokumentation. Trauner, Linz, pp 461–476
Schaffhauser-Linzatti MM, Faisst M (2003) Possible application of the balanced scorecard at the University of Vienna. In: Montanheiro L, Kuźnik F, Ochojski A (eds) Public and private sector partnerships: sustainable success. Sacred Heart University Press, Sheffield, pp 383–402
Sinuany-Stern Z, Mehrez A, Barboy A (1994) Academic departments efficiency DEA. Comput Oper Res 21:543–556

Sizer J, Spee A, Bormans R (1992) The role of performance indicators in higher education. High Educ 24:133–155

Slunder S (2008) Kennzahlenbasierte Leistungsmessung und darauf aufbauende Mittelallokation in Universitaeten – eine empirische und implikationstheoretische Analyse. PhD Thesis, University of Duisburg-Essen

Stewart TA (1998) Der vierte Produktionsfaktor – Wachstum und Wettbewerbsvorteile durch Wissensmanagement. Hanser Fachbuch, Munich

Sveiby K-E (1997) The new organizational wealth – managing and measuring knowledge-based assets. Berrett-Koehler, San Francisco

Tai W, Chen C (2009) A new evaluation model for intellectual capital based on computing with linguistic variables. Exp Syst Appl 36:3483–3488

Terlinde C (2005) Aufdeckung von Bilanzmanipulationen in der deutschen Pruefungspraxis, Ergebnisse einer empirischen Studie. Deutscher Universitäts-Verlag, Wiesbaden

Thom N, Ritz A (2000) Public management. Innovative Konzepte zur Fuehrung im oeffentlichen Sektor. Gabler, Wiesbaden

Turner D (2005) Benchmarking in universities: league tables revisited. Oxf Rev Educ 31:353–371

Ungureanu M (2012) Social accounting – factor for developing the enterprise's internal and external environment. Bull Trans Univ Brasov 5:169

Van Thiel S, Leeuw FL (2002) The performance paradox in the public sector. Pub Perform Manag Rev 25:267–281

Veltri S (2007) The relationship between intellectual capital management and firm performance. Paper presented at the international forum on knowledge asset dynamics (IFKAD) 2007, 2nd edn, Matera, 22–23 June 2007

Veltri S, Schaffhauser-Linzatti MM, Giovanni M (2009) The development of a corporate intellectual capital index in the Austrian University sector using a fuzzy logic approach. In: 5th EIASM Workshop on visualising, measuring, and managing intangibles and intellectual capital, Dresden, 8–9 Oct 2009

Wagner K (2002) Evaluierung der Forschungsleistungen an der Technischen Universitaet Wien. Analyse verschiedener Modelle zur Effizienzmessung. University of Technology, Vienna

Wheelock C, Wilson PW (1995) Evaluating the efficiency of commercial banks: does our view of what banks do matter? In: Review of the Federal Reserve Bank of St. Louis, 77, July-Aug: 39–52

Zadeh LA (1978) Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets Syst 1:3–28

Zebda A (1991) The problem of ambiguity and vagueness in accounting. Behav Res Account 3:117–145

The Elusive Effectiveness of Performance Measurement in Science: Insights from a German University

Christoph Biester and Tim Flink

Abstract Over the last decades, procedures of compiling, measuring and evaluating academic performance have made incursions into the realm of universities. The result is an effect on employment negotiations, increased competitive performance measurement in funding allocation and, in the context of salary distribution, impact on how individual and collective achievements of academic staff are compared among each other or between different research institutions. Despite the pervasiveness of this type of systematic performance measurement impinging upon nearly all university activities, we still know little about whether these systems matter for researchers under evaluation. Based on empirical insights from an evaluation at a large German university, we discuss perceptions of professors exposed to one university performance measurement system. That exposure seems to trigger, in particular, worrisome attitudes of ambivalence towards the university and the academic value system.

1 Introduction

In the course of an emerging Audit Society (Power 1999), the public sector has taken to extensive evaluations of nearly all social activities, not sparing higher education and the system of scholarship. Universities, traditionally regarded as the loci of hitherto academic freedom and muse (Stichweh 2003), have been forced to adopt output-oriented management systems, albeit at different periods of time Walker and Sharp 1991. In Germany's case, universities were confronted with a mixture of more or less pronounced goals. The so-called New Steering Model (Kommunale Gemeinschaftsstelle für Verwaltungsmangement 1993), following New Public Management leitmotifs of accountability and competition (Dunleavy

The authors thank Jacob Watson for his great editorial support, Tim Seitz for immediate but thorough research assistance and their anonymous reviewers for encouraging a substantial revision of the chapter's original version. Tim Flink owes gratitude to Chia-Yu Kou for her advice on dealing with insights from organizational psychology.

C. Biester (✉) • T. Flink
Wissenschaftszentrum Berlin für Sozialforschung gGmbH, Berlin, Germany
e-mail: christoph.biester@wzb.eu

I.M. Welpe et al. (eds.), *Incentives and Performance*,
DOI 10.1007/978-3-319-09785-5_24

and Hood 1994) entailed the implementation of predominantly output-oriented, i.e., indicator-based management systems for higher-education institutions (Bogumil et al. 2013).[1] In consequence, universities are now requested to report on the impact of conducted research, teaching, knowledge transfer and other activities in order to demonstrate their relevance to society (Wissenschaftsrat 2006). However, as federalism and constitutionally guaranteed academic freedom in Germany (Hüther et al. 2011) have set the playing field, differences have increasingly become apparent in both how universities of the 16 federal states have implemented output governance systems to reap basic and third-party funds (Jansen 2010) and how they differ from each other in adopting *individual* performance measuring systems vis-à-vis the individual federal state principals. Not least, since ordinary professors in Germany are public servants to their states and thus enjoy the highest degrees of autonomy in international comparison (Buchholz et al. 2009), universities' management levels can only but resort to indirect steering activities. Established in 2002, the W-salary[2] directive (W-Besoldung) (ProfBesReformG 2002) exemplifies one such indirect activity, which is nonetheless a powerful instrument of performance measuring and rewarding and arguably unfolds veritable governance properties (Biester 2010, 2013): For the first time in post-war German history, performance measurement directly impacts the *personal* salary of university professors. In light of these developments, it may become pervasive as a governance technique. This is because decisions pertaining to appointments and salaries are not permitted to be based solely on the output and *measured* impact of professors' research, but must take into account, for instance, the assessment of their acquired research grants —a social construct which tends to get translated as actual research achievements (Flink and Simon 2014). Not least, professors are required to show a track record of teaching courses, however, an evaluation of the content seems to be neglected (Wissenschaftsrat 2005).

Hence, it seems that a one-way road is paved: The more political instances put pressure on universities' management levels to govern their institutions autonomously (Guston and Keniston 1994) and to justify their output in a yearly rhythm,

---

[1] Arguably, NPM reforms pertaining to German higher education and science institutions were heralded by the German Council for Council of Science and Humanities (Wissenschaftsrat) suggesting in its recommendations on *Competition of the German Higher Education System* (1985) that evaluations should enhance "achievements by means of a public comparison" and bring about "transparency in achievement" (WR 1985). One goal was to counteract an increasing loss of trust in the capability of the scientific community to regulate itself, exhibited both by the political administration and the general public (Krücken and Maier 2007), while other goals, e.g., to make Germany's science system internationally more competitive (WR 1992), ensure gender equality (WR 1998), better and structured training of young researchers (WR 2002) were consecutively linked up to this reform.

[2] The W-salary directive regulates the salaries of professors in the German higher education system consisting of a basic pay (lower in comparison to the former C-salary) and a top-up pay remitted to the individual professor depending on his/her performance in research and teaching as well as on tasks and responsibilities of academic self-organization (e.g., university presidencies, faculty responsibilities).

the more university employees are bound to bear the consequences of their activities. Scholars have criticized the social consequences of academic performance measurement, predicting difficult times for scholarship and the higher education system as a whole: They warn that the freedom of research and teaching is in danger, that the intrinsic motivation to produce new knowledge for reasons of pure research interest gets suppressed (Osterloh and Frey 2008; Kieser 2010), that unintended effects, such as opportunism and misconduct, arise (Sieweke 2010) and that scholars increasingly find favor with research to the disadvantage of teaching (Ronge 1998), as well as the possibility that narrower subjects might die out (Hartmann 2006). Unfortunately, such scholarly output addressing the realm of higher education institutions is hardly borne by rigor of empirical evidence (except Sauder and Espeland 2009). In cases where empirical data are employed, they serve mostly to deliver a bird's-eye view (Paradeise and Thoenig 2013) but neglect in-depth examinations of individual universities, non-university research institutes and their staff members. Even worse, sociology and political science hardly attend to the cognition of economics and psychology that have, inter alia, investigated the question as to whether intrinsic motivation is triggered by extrinsic motivators.[3] This empirical lacuna in the social sciences is astonishing given the genuine dismay academics have with the seemingly increasing trend towards organizational accountability at the price of self-determination and peer group recognition (Flink and Simon 2014). This critique may also apply to the social studies of science, as performance measuring systems for higher education and scholarship would indeed offer a wellspring of relevant research puzzles, e.g., the stability of one of their classical hypotheses: that science is borne by a self-referential normative structure (Merton 1990) and thus needs no further extrinsic motivators. This leads us to our main objective for this chapter: If academics primarily follow their research interests[4] co-defined by a globally spread community, how do they perceive organizational performance measurement techniques embellished with financial rewards? We will shed light on this question by presenting empirical findings from an evaluation of a performance measurement system at a large German university.

Our findings are structured as follows. First, we introduce some important properties of the evaluated performance measurement system. Second, we introduce the research methods applied, a combination of an online survey and

[3] These include the works of Deci et al. (1989), Jensen/Murphy (1990), Gneezy and Rustichini (2000), Osterloh/Frey (2000), Ryan/Deci (2000), Kunz and Pfaff (2002), Bock et al. (2005), Ederer and Manso (2012), a great deal of which discusses the gulf of economic or behavioral management premises of rational, i.e., selfish and extrinsically motivated human beings versus psychologically informed premises of self-determined human beings, who might be motivated by extrinsic rewards indirectly, while experiencing gratification if they perceive their activities as enjoyable, challenging and purposeful (for a brilliant literature review, see Weibel et al. 2010).

[4] Flink et al. (2012) surveyed professors with high performance in research at German universities (N = 2,538), the findings illustrated that professors care much less about their university profile but most about their own research interests and their peer community.

individual professor interviews. Third, we discuss the evaluation of the data to conclude, fourth, that performance measurement at this particular university has a disciplining rather than a motivating effect on professors: While most professors do indeed participate, and some financially profit from the system, their perceptions reflect inner conflicts towards performance measurement, which is best described by the term ambivalence. The results lead us to the conclusion that while performance measuring in scholarship may prove successful in observing and partly triggering individual activities, its negative social consequences —, most notably embittered attitudes towards the academic system as a whole — make it worth reconsidering.

2 The Case of Performance Measurement System at a Large German University

In 2000, the university management of our case study followed ministerial demands to implement an output-oriented performance measuring system. It is worth noting no other university has installed a system of such quantitative rigor under the auspices of the same federal state or any other of the federal states in Germany. While only some parts of the input stemming from this performance measuring are actually put to use by the university management to illustrate output and thus obtain funding from the state, professors are requested to systematically enter all of their own and their staffs' achievements into an online performance recording system. Here, professors run across 30 criteria[5] that roughly follow the categories of research, teaching, knowledge transfer and activities of academic self-administration. The system converts their achievements into points[6] attached to which are two conversion keys.[7] Addressing both professors under the old C-salary directive and the new W-salary directive, one conversion key translates achievements into points dedicated to equip their faculties with resources, i.e., research staff, instruments and facility resources.[8] A second conversion key applies *only* to professors falling under the W-salary directive and translates their points into sums for performance-based top-up funding. The system disrespects disciplinary or subject-specific differences. For example, while professors in art history report

[5] The exact number of criteria is not specified here to ensure case anonymity.

[6] To give you some examples how achievements are translated into points: supervision and assessment of dissertations/habilitations = 40 points; book publication = 16 points; peer-reviewed article = 8 points; student thesis supervision and/or exam assessment = 2 points; reviewing a research proposal = 2 points; student exam = 1 point.

[7] The achievements are controlled by the administration. As an example, publications need to be referenced with international standard book numbers, or assessed exams require the adding of student identification numbers.

[8] All faculties developed their individual allocation model to receive funds from the university management.

that they are not used to publishing in peer-reviewed journals, an achievement that the system rewards with a considerable amount of points, and while their degree courses show many fewer modules and students than, for instance, engineering, they are not offered any alternatives to receive funding for their achievements.[9]

According to the sum of accumulated points, professors under the W-salary directive can receive bonus payments adding up to their monthly salary, which the system ranks in a five-tier model,[10] with each tier carrying a different financial weight: rank 1: 0–100 points = €0.00; rank 2: 100–200 points = €90.00; rank 3: 200–300 points = €280.00; rank 4: 200–300 points = €370.00; rank 5: over 400 points = €500.00. Apparently, this model offers cumulative advantages for those professors classified into the higher ranks (Merton 1988): If, for instance, you reach level 3, one achieved point will be worth €2.80, while at level 4 it is already worth €3.70. In light of the fact that professors might need years to achieve enough points in order to climb up to higher ranks, equal achievements (e.g., a peer reviewed journal article handed in by a level 2 professor in comparison with a level 4 professor) performed by lower-ranking professors effectively result in less money. It must also be noted that performance-related payments are carried out incrementally; a professor able to demonstrate an increase in his or her achievements during the reporting period, but who does not cross the threshold to reach the next level, will remain at the current level. In contrast, anyone demonstrating poor performance over two consecutive reporting periods (6 years in total) will be downgraded a rank level.[11] It is crucial to know that by starting or continuing their academic career at this university, professors have the opportunity to be classified into higher ranks from day 1; whereas, for professors coming from elsewhere, i.e., industry or academia, this depends on assessment of previous activity and, most certainly, on their negotiation skills. Moreover, professors cannot be forced by the university management to participate in this system. While a loss of potential bonus payments for those falling under the W-salary directive might be bearable,[12] all the professors can nonetheless expect to experience moral pressure from their faculty management, as defiance results in reduction their faculty's resources.

[9] Reviewed annually by the university management, the input of achievements can be carried out continually over the course of the year. An advisory council, consisting of professors from different faculties, acts as a switchboard between the professors, who can send in questions or criticism or suggest amendments, and the university management, which decides about how to deal with professorial requests.

[10] Professors can only rise up to a higher rank, if their performance is sustained for a period of 3 years.

[11] If professors demonstrate steady performance over a long period of time, an application for the removal of the time limit for the performance-related bonus can be accepted by the university management.

[12] At least, this is what professors reported to us.

3 Data and Methods

Our evaluation of the university performance measurement system followed a multi-stage process, combining qualitative (expert interviews) and quantitative (survey) methods. In the first stage, four main thematic areas (user evaluation, governance effects, perceptions of performance measurement criteria and attitudes towards bonus payments in the W-salary directive) were agreed upon with the university management. After a pretest of a semi-structured questionnaire (Meuser and Nagel 2009), interviews were conducted with all of the status groups involved in the performance measurement system, both those who manage it and those who are exposed to it.[13] The interviewees' sample comprised of representatives from the university presidency and administration (n = 3), deans (n = 3), the chairman and one member of the measurement system's advisory board (n = 2), as well as professors (n = 10) from a range of disciplines, departments and different salary grades (C, W2, W3), hence showing the greatest possible contrast. A content analysis (Mayring 2010) of interviews[14] served as the basis for planning an online survey questionnaire.

Pretested with three professors of this university, the online survey[15] served as a quantitative method to ask all of the relevant[16] professors a total of 53 questions, divided into six thematic areas. First, we asked questions on the four main thematic areas pertaining to the evaluation of the performance measurement system and, second, questions on work satisfaction and employment relationships. The survey's response rate amounts to 59 %, an exceptionally good result. Altogether, the survey provided a total of 276 variables[17] and 139 cases for data analyses. Moreover, individual survey answers could be correlated with each person's data extracted from the measurement system, including their age and gender, their academic subject, their salary status, either falling under the C- or W-salary directive and, in particular, their individual performance points.[18] This allowed us to shed light on questions such as whether low performing professors perceive performance

[13] We pursued two distinct goals by carrying out expert interviews: (1) gain technical knowledge, i.e., about administrative competences and specialized knowledge pertaining to the measurement system, (2) to look out for process-related knowledge about potentially interesting interactions, decisions and organizational dynamics and (3) detect interpretive schemas borne by everyday knowledge that the interviewees have generated in grappling with the measurement system (see Bogner et al. 2009).

[14] All interviews were conducted in German. The sequences displayed in this chapter were translated by the authors.

[15] The online survey was developed and implemented using the software *limesurvey*.

[16] We neither addressed professors on the salary grade W1, nor guest professors, special professors and acting professors, as they do not take part in the university performance measurement system.

[17] The high number of variables in comparison to the number of questions is due to the differentiated sampling of varying circumstances for professors on the C- and W-salary scales.

[18] It is important to note that all personal information was immediately coded and that correlations were made on such aggregate level that it guarantees full anonymity to surveyed staff.

measuring more critically than high performers or if differences in their response behavior depend on their subject or academic discipline, their age, gender and on their salary status. Combining qualitative expert interviews with survey data, we can determine whether performance measuring impacts the professors' behavior and, not least, in their perceptions of work- and status-related concepts, such as autonomy, self-determination and gratification. However, other than presenting a colorful picture, we cannot deduce any causal links between survey results and interview statements, with the first representing attitudes and the latter only but offering *potential* explanations for these attitudes. And yet, the latter represent more than just individuals' answers, but are to be read as interpretive schemas bridging the schism of structure and agency (Sewell 1992).

4 Results

4.1 Perceptions on Performance Measurement and Steering

How do we explain that all professors addressed by the university performance measurement system do actively participate in it, but about 45 % rate this particular system as very unfair, and another 13 % as rather unfair?[19]

Our first question is if the professors oppose the performance measurement system in general. In the online survey, professors were asked whether or not they agreed with the statement. "Steering via the performance measurement system is necessary." Our hypothesis predicted that a clear majority would not agree with this statement. We were surprised, however, that the opposite was the case: just under a quarter strongly agreed with the statement, while 15.1 % agreed somewhat (Fig. 1).

When analyzed according to discipline, it turns out that the humanities and social science researchers are not the ones to least agree with the statement, but rather the natural scientists, however, the differences are not statistically significant. Furthermore, the agreement or disagreement with the statement was independent of salary grade and individual research achievements (Fig. 2).

As the amounts for performance-related bonuses in the context of the W-salary directive are determined by means of the evaluation system, this subject was also included in the questionnaire. The professors paid W-salaries were asked to assess the suitability of the system on a scale from fundamentally appropriate (2) to fundamentally unsound (−2), on the assumption that it would be perceived as

[19] The survey question to be answered was "The procedure of the [name of the] system assessing and rating my performance is overall fair." The highest rejection (64 %) comes from engineering scientists, followed by the social sciences/humanities (62 %), life sciences (45 %) and natural sciences (44 %), without any statistically significant differences according to professors' salary status, gender and age.

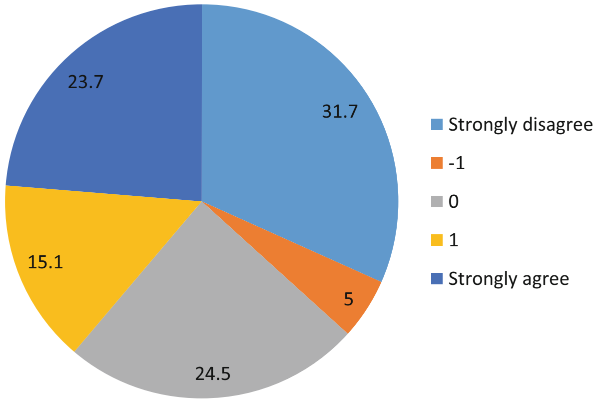


Fig. 1 Management with the performance measurement system is necessary (N = 139)

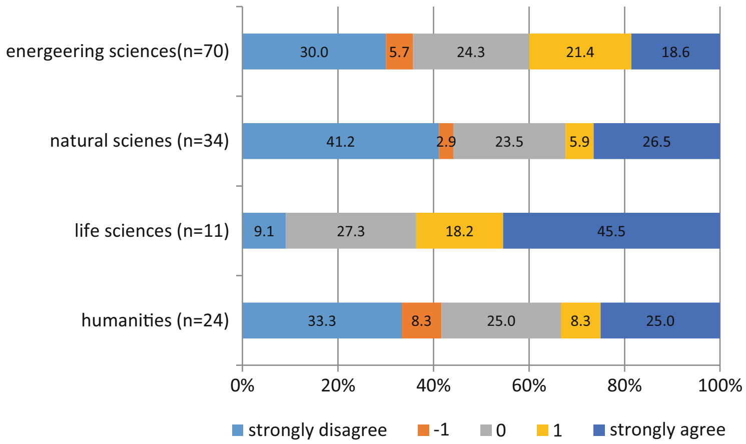


Fig. 2 Management with the performance measurement system is necessary, according to disciplines (N = 139)

fundamentally unsound by the majority. However, of the professors surveyed, 51.4 % considered the W-salary directive to be fundamentally appropriate. Only under a fifth (18.9 %) considered it to be fundamentally unsound.[20]

While we are still struggling with the question as to why the steering via the system is widely accepted, but the system as such is held to be unfair, further responses to questionnaire items might shed light on this puzzle (Fig. 3).

[20] The high level of support for performance-related bonuses in academic pay was positively correlated (Pearson $r = 0.389$, $p < 0.01$) with the incremental value obtained for research. This means professors who perform highly in research are those fundamentally in favor of the W salary directive. With regard to teaching, this correlation was however not determined.

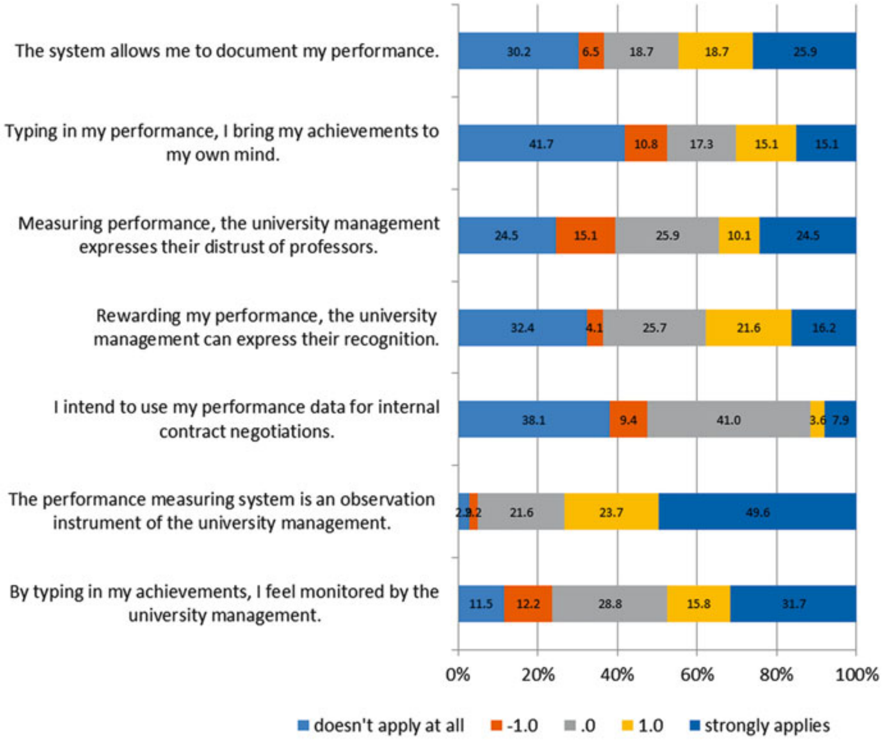


Fig. 3 Use of performance input and monitoring experience (N = 139)

The answer as to whether the system is suitable for documenting individual performances shows rather equal distributions on both negative and positive ends. A fairly negative response is provided to our suggestion that (1) entering performances would stimulate the respondents' individual self-reflections on their individual achievements. Moreover, professors report (2) that they are neither in need of using their performance data for internal contract negotiations (disagreed with by overall 47 %), nor do they seem to have a clear opinion about this possibility (41 %). When it comes to questions pertaining to interactions with the university management, i.e., the presidency and its staff[21] in charge of the measurement system, answers reflect clearer and rather negative perceptions. Here, performance measurement seems to reveal a conflict between the individual researcher's autonomy on the one hand and organizational control on the other hand: About half of all responding professors *strongly* agree with the statement (3) "The measurement system is *all but about observing* our work," and another fourth of all respondents

[21] Questions pertaining to control experience were actor-specifically differentiated. Over 70 % strongly impute steering interests to the university presidency (another 10 % rather confirm this), while deans and faculty administrations were only selected by 40 % (strong confirmation), and each by further 18 % (weaker confirmation).

merely support this opinion. That professors also experience (4) the sensation that *management attempts to monitor* their activities is strongly confirmed by nearly one third, and rather confirmed by a further 16 % of the respondents. Even one third of all respondents believe that (5) “The reason for measuring their performance is *distrust* by the university management.” While 40 % do not share this opinion, it is nonetheless worrisome to see such a robust negative attitude.

The analysis of the expert interviews presents a differentiated picture of the perception of performance measurement in the university. First of all, most statements show sensitivity in that academic activities must be held accountable to society. Whether performance measurement is something feasible is answered in a differentiated way.

> I consider it plausible to require performance in research and teaching to be assessed, and also that it is the right thing to do in the sense of having an incentive effect. As in many other contexts, there is always the question, ‘How do I measure it?’ And that is naturally not quite trivial in research and teaching, as well as in many other cases. (Interview 3)

The basic agreement with the measurement and evaluation of performance is associated with an incentive effect, with the fundamental question of the feasibility of measuring and evaluating research and teaching achievements then pointed out.

> But quite fundamentally I find it right, because I do see that in academia there is a high degree of freedom, which is good, and there are certainly many people employed here, who are highly motivated. But despite this, from my point of view, particular achievements should be noticeable, and those that produce poor performance should also be noticed, so that the institute can make sure that things are improved. (Interview 3)

The basic agreement with performance measurement is associated with a system of recognition that “outstanding achievements should be noticeable,” which goes beyond the recognition of achievement by the relevant scholarly community, which in this context was not mentioned once. At the same time, poor performance should also be visible to enable institutes to react, with no suggestion being offered as to how the institute could and should improve matters.

For another interviewee, there was no doubt that research and teaching performance is subject to criteria that can be used to evaluate performance. However, this could not take place in a solely quantitative manner due to the complexity of the subject matter. The transfer of qualitative performance measurement, provided in a statement, into a figure is regarded as a problem:

> Well, generally both achievements in teaching as well as research are subject to criteria with which one must comply. These are qualitative and quantitative criteria. You rightly said that we judge, for example, publications and research applications. But when I do that, a statement and no figure comes out at the end. And that is exactly the big problem that I see here in this performance measurement system: we are illustrating an extremely complex matter with one figure. (Interview 2)

While professors maintain that research and teaching achievements are basically quantifiable, performance measurement and evaluation are regarded to be a form of recognition. Not least, they are held to be plausible and necessary in the sense that they can act as an incentive system. This agreement is however only in exceptional cases, expressed explicitly and without reservation, as in the following quote:

> Ultimately, one has to justify what one does to oneself as well as to others who provide the money and also entrust one with the responsibility for students, trainees and research funding. When we demand that our students undergo an evaluation, we should also do the same for professors or university departments. (Interview 7)

Yet, in relation to what criteria should be used for performance measurement, numerous discipline-specific issues were raised across all of the interviews. Given in merely one typical example of the interviews, this statement establishes a connection with a question asked in the online survey on the inclusion and evaluation of qualitative criteria in the performance measurement system:

> I'll tell you quite honestly, I don't think much of it; in my opinion it probably isn't really possible to make these evaluation criteria suitable for all scientific disciplines [...] A research paper could be two pages or 50 pages long. It could appear in a high-impact journal or in a run-of-the-mill journal. This all counts as roughly the same in, for example, the performance measurement system, without particular weighting being given to the impact factor of the journal. (Interview 4)

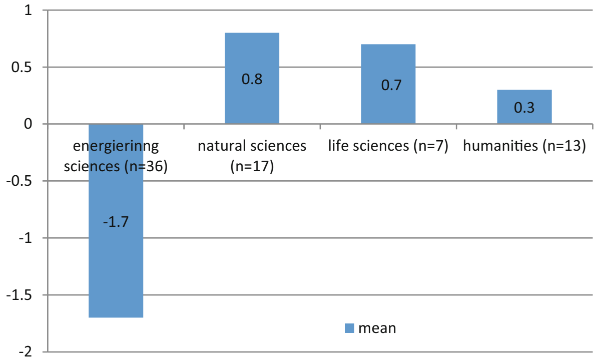
The exclusively quantitative evaluation of research and teaching performance was viewed critically by many of the interview partners. In the online survey, a good half (54 %) of professors had a favorable opinion towards the inclusion of qualitative indicators in the system; whereas, one quarter was against and a further quarter selected the option "Do not know." In a second stage, those professors in favor of the use of qualitative indicators were presented with a selection of various parameters to evaluate, including the use of impact factors for the evaluation of publications. The inclusion of impact factors in the performance measurement system for the evaluation of publications was considered by a good quarter of the academics to be very appropriate (27.4 %); whereas, a good fifth (21.9 %) viewed it as not at all appropriate. The following chart shows how this evaluation differed between academic areas (natural sciences, life sciences, engineering sciences as well as the humanities and social sciences) (Fig. 4).

Whereas the average evaluation of the impact factor indicator was positive in both the humanities and the social sciences, as well as in the natural and life sciences, academics in the engineering sciences considered this indicator to be only of limited appropriateness.

4.2 *The Effectiveness of Performance Measurement and Evaluation*

The analysis of the interviews and results from the online survey indicate that performance measurement and evaluation are indeed effective at this university. However, the question is: *how* do they affect professors? For the analysis presented here, it is assumed that an effect exists when a change in conduct in connection with the performance measurement and evaluation system is detectable, as evidenced by the following quote:

Fig. 4 Differentiation according to journal quality using impact factors (N = 73)



> Regarding this performance measurement system, I know what gets you points, then you do pay a bit more attention to that, so that you organize your daily routine accordingly, so that you will work on tasks that will bring points with greater urgency or in greater number, or so that you are carrying them out mainly in order to be able to actually register them in this performance measurement system. [. . .] It has happened that I made the direct offer to colleagues, 'when you need a reviewer for your doctoral students, I am available.' Then you go on the offensive rather than remaining passive. (Interview 4)

Such clear statements are not typical of the interviews. Quite openly admitting that certain tasks are performed with more urgency in order to receive points is an exception, but this phenomenon suggests that performance measurement and evaluation are effective in the sense of steering work towards fulfilling predetermined criteria. Another indication that the effect of performance measurement and evaluation does indeed manifest itself lies in the fact that the statement dealing with the effect of the system tend to be emotional; that is, people feel personally affected. The system itself is a source of offence, but seeing that one's individual achievement is visible in the system and brings something to the department is a source of pride and results in a positive feeling:

> There is something that I had personally thought was not possible. There is something insulting about it. I consider myself to be a very engaged professor. I do it with dedication, love, I find this job fantastic, and I know where I achieve a lot. But the fact that when I am not watching out for the points, not doing the important things, not publishing in peer-reviewed journals, writing articles or supervising doctoral students — the big cash cows of the performance measurement system — when I am not keeping an eye on these things, I had thought that this does not exert control over me, because I say that I will not let such a system dictate how I develop my professorship. But if I don't do these things, I realize that it will insult me. (Interview 5)

> But I also observe within my department that, naturally, when people have the feeling that I have, for example, shown my colleagues this points table and do observe that people are very proud when they see they have done something that brings something to the department. (Interview 8)

> [W]ord gets round about who is a good researcher. And that is accidentally naturally correlated with the performance measurement system. We live for this recognition. People

> do want recognition. You can call it vanity, but it feels good if you face the world with a good standing in the performance measurement system. (Interview 9)

In each of the three quoted passages emotions are expressed. In the first passage, a conflict between individuals' attitudes to work and the system is clear. The experience of a disparity between how one's duties as a professor, according to one's own opinion, ought to be realized and developed and that what the criteria within the system convey as important has caused ambivalence, i.e., in this case offence. It is not the case that the measurement and the evaluation of performance have zero impact on the organizational structure of the professorship or of the profession, in the sense that differing rationales or frames of reference meeting on the conflict line of scholarship and governance are proving to be ineffective. Rather, the system develops a steering effect that may not reach far into the profession, but nevertheless has reached individuals perceiving it as an affront. Conversely, if performance within the system is rendered visible, and if the consequences are tangible, namely so that the individual achievement even brings something to the discipline, this becomes a source of pride and is thus perceived as a form of recognition, which feels good.

Even though the question of the effectiveness of performance measurement and evaluation is difficult to answer due to an absence of clearly defined terms, it is obvious from the cited passages that the performance measurement and evaluation system affects the behavior of professors, causes offence and pride as well as conferring local recognition.

4.3 Unintended Effects and Structural Problems

The fact that the evaluation also showed that clear indications of unintended effects and undesired steering as well as associated structural problems exist must not be concealed, although this is only briefly described here. It is clear from the interviews that the measurement and evaluation of teaching performance is particularly problematic due to the variation in module structures used in degree courses.

In the evaluation of research performance, a similar structural problem with regard to doctorates and post-doctorates is apparent. In a narrow field of study, fewer degrees will be awarded than in a larger one. This leads to a serious problem, as described in one case: Professors may be faced with the dilemma of deciding whether to supervise a post-doctoral project that does not fit thematically into their discipline or specialism but which would clearly improve the personal points ranking. In other words, will a professor decide according to disciplinary quality standards or will she/he leer at achievement points?

A further effect, which is without a doubt to be assigned to the unintended category, arises from the practice that the points for co-authored publications are divided between the authors. In one case, not only was it stated that competition between professors within the university increased as a consequence of this practice, but it was also noted that this led to research disciplines becoming more thematically inward-looking despite the fact that increased interdisciplinarity is widely called for. To be fair, it must be noted that the established practice of dividing points has probably caused the number of publications with multiple authors to diminish.

5 Summary and Further Research Questions

The evaluation of a performance measurement system at a large German university illustrates that performance measurement affects the behaviors of exposed professors by causing a dichotomy between agreement and disagreement or criticism and not least by provoking high degrees of ambivalence, i.e., simultaneous, conflicting reactions towards being measured. The fundamental agreement with performance measurement and evaluation points to a positively-viewed system of incentives and recognition and the fundamental measurability and evaluability of research and teaching achievements in general, e.g., for being accountable to the society or for serving one's own faculty. In contrast, doubts are cast upon the extent to which measurement and evaluation are possible at all, especially if evaluating all disciplines' performance through a single set of criteria can be viable.

Moreover, the emotions expressed in the interviews and the specific descriptions of activities show that performance measurement yield stress in terms of decision-making, but more importantly, it has personal effects: When an individual's achievements are not recognized, this is experienced as a slight affecting even the self-regard of those being measured. However, if individual achievements become visible, by means that are obviously intrinsic to the logic of the academic system, this produces pride and is perceived as recognition. Hence, it does not appear to be the case that the measurement and evaluation of individual achievements by the university management system has zero impact on the profession in the sense that differing frames of reference meeting on the line of conflict between scholarship and organizational governance are proving to be ineffective. Further evidence for the effect of performance measurement and evaluation are the descriptions of what are termed unintended effects. Some of our findings point to the effect of individuals partly engaging in activities of shirking or deception, whilst exposed to performance measurement, rather than simply refusing participation by reasons of criticism.

Altogether, the question remains whether we need better, i.e., either clear-cut or wide definitions of the term effectiveness, in particular when it comes to individual or collective exposure to performance measuring. In this, for further research we encourage to bridge a seemingly disciplinary gap between quantitative and qualitative empirical social research and behavioral economics and psychology. Here, the realm of higher education and scholarship offers a wellspring of exciting phenomena, following the observation that academics are socialized to pursue their activities with highly intrinsic motivations, i.e., they see their job as a unique form of 'calling', whilst facing techniques of organizational accountability.

References

Biester C (2010) Change management in higher education: the introduction of performance oriented payment in German universities. In: Papanikos G, Pappas N (eds) Problems and prospects in higher education. ATINER, Athen

Biester C (2013) Leistungsorientierte Vergütung in der Wissenschaft. Eine theoretische und empirische Analyse der neuen Professorenbesoldung. Beltz, Weinheim

Bock GW, Zmud RW, Kim YG, Lee JN (2005) Behavioral intention formation in knowledge sharing: examining the roles of extrinsic motivators, social-psychological forces, and organizational climate. MIS Q 29(1):87–111, Special Issue on Information Technologies and Knowledge Management

Bogner A, Littig B, Menz W (eds) (2009) Interviewing experts. Methodology and practice. Macmillan, Basingstoke

Bogumil J, Burgi M, Heinze RG et al (2013) Modernisierung der Universitäten. Umsetzungsstand und Wirkungen neuer Steuerungsinstrumente. Edition Sigma, Berlin

Buchholz K et al. (2009) Attraktivität von Arbeitsbedingungen in der Wissenschaft im internationalen Vergleich. Wie erfolgreich sind die eingeleiteten wissenschaftspolitischen Initiativen und Programme? WZB, Berlin. http://www.e-fi.de/fileadmin/Studien/StuDIS2009/EFI-Bericht-WZB_rev_Febr09.pdf. Accessed 12 May 2014

Deci EL, Connell JP, Ryan RM (1989) Self-determination in a work organization. J Appl Psychol 74(4):580–590

Dunleavy P, Hood C (1994) From old public administration to new public management. Pub Money Manag 14(3):9–16

Ederer F, Manso G (2012) Is pay-for-performance detrimental to innovation? Manag Sci 59 (7):1496–1513

Flink T, Simon D (2014) Erfolg in der Wissenschaft: Von der Ambivalenz klassischer Anerkennung und neuer Leistungsmessung. In: Hänzi D, Matthies H, Simon D (eds) Erfolg. Konstellationen und Paradoxien einer gesellschaftlichen Leitorientierung, vol 29, Leviathan Sonderband. Nomos, Baden-Baden, pp 123–144

Flink T, Rogge JC, Roßmann S, Simon D (2012) Angleichung statt Vielfalt. Deutsche Universitäten auf der Suche nach Profil. WZBrief Bildung 22:1–7

Gneezy U, Rustichini A (2000) Pay enough or don't pay at all. Q J Econ 115(3):791–810

Guston DH, Keniston K (1994) The fragile contract: university science and the federal government. MIT, Cambridge

Hartmann M (2006) Die Exzellenzinitiative — ein Paradigmenwechsel in der deutschen Hochschulpolitik. Leviathan. Berliner Zeitschrift für Sozialwissenschaft 34(4):447–465

Hüther O, Jacob AK, Seidler HH, Wilke K (2011) Hochschulautonomie in Gesetz und Praxis. Eine Umsetzungsanalyse vordem Hintergrund der Förderprogramme des Stifterverbandes und der Heinz Nixdorf Stiftung, Speyer. http://www.stifterverband.info/wissenschaft_und_hochschule/hochschulen_im_wettbewerb/archiv/deregulierte_hochschule/hochschulautonomie_in_gesetz_und_praxis_kurzversion.pdf. Accesed 20 May 2014

Jensen MC, Murphy KJ (1990) Performance pay and top-management incentives. J Polit Econ 98 (2):225–264

Jansen D (2010) Von der Steuerung zur Governance: Wandel der Staatlichkeit? In: Simon D, Knie A, Hornbostel S (eds) Handbuch Wissenschaftspolitik. Wiesbaden, VS Verlag, pp 39–50

KGSt, Kommunale Gemeinschaftsstelle für Verwaltungsvereinfachung (1993) Das Neue Steuerungsmodell. Begründung Konturen Anwendung. KGSt-Bericht 5, Köln

Kieser A (2010) Unternehmen Wissenschaft? Leviathan 38(3):347–367
Krücken G, Maier F (2007) Turning the university into an organizational actor. In: Drori GS, Meyer J, Hwang H (eds) Globalization and organization. Oxford University Press, Oxford, pp 241–257
Kunz AH, Pfaff D (2002) Agency theory, performance evaluation, and the hypothetical construct of intrinsic motivation. Acc Organ Soc 27(3):275–295
Mayring P (2010) Qualitative inhaltsanalyse. Beltz, Weinheim
Merton RK (1988) The Matthew effect in science II. Cumulative advantage and the symbolism of intellectual property. ISIS 79(4):606–623
Merton RK (1990[1942]) The normative structure of science. In: Alexander J, Seidman S (ed) Culture and society. Cambridge University Press, Cambridge, pp 67–74
Meuser M, Nagel U (2009) Das Experteninterview – konzeptionelle Grundlagen und methodische Anlage. In: Pickel S, Pickel G, Lauth HJ, Jahn D (eds) Methoden der vergleichenden Politik- und Sozialwissenschaft. Neue Entwicklungen und Anwendungen. VS Verlag für Sozialwissenschaften, Wiesbaden, pp 465–479
Osterloh M, Frey B (2000) Motivation, knowledge transfer, and organizational forms. Organ Sci 11(5):538–550
Osterloh M, Frey BS (2008) Anreize im Wissenschaftssystem. Arbeitspapier, Zürich. https://www.unizh.ch/iou/orga/ssl-\-dir/wiki/uploads/Main/Anreize_final_12.9.08.pdf. Accessed 20 May 2014
Paradeise C, Thoenig JC (2013) Academic institutions in search of quality: local orders and global standards. Organ Stud 34(2):189–218
Power M (1999) The audit society: rituals of verification. Oxford University Press, Oxford
ProfBesReformG (2002) Gesetz zur Reform der Professorenbesoldung – ProfBesReformG. Bundesgesetzblatt 11:686–692
Ronge V (1998) Leistungsentlohnung von Professoren? Kritische Anmerkungen zur aktuellen Diskussion. Forschung und Lehre 8:412–413
Ryan RM, Deci EL (2000) Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 55(1):68–78
Sauder M, Espeland W (2009) The discipline of rankings: tight coupling and organizational change. Am Sociol Rev 74(1):63–82
Sewell WH (1992) A theory of structure: duality, agency, and transformation. Am J Sociol 98 (1):1–29
Sieweke S (2010) Strukturelle Effekte der W-Besoldung. Zum Erfordernis der Disziplinenunterscheidung bei leistungsorientierter Bezahlung. Wissenschaftsmanagement 3:40–44
Stichweh R (2003) Genese des globalen Wissenschaftssystems. Soziale Systeme 9(1):3–26
Walker W, Sharp M (1991) Thatcherism and technical advance. Reform without process? Part I: the historical background. Polit Q 62(2):262–272
Weibel A, Rost K, Osterloh M (2010) Pay for performance in the public sector – benefits and (hidden) costs. J Public Adm Res Theory 20(2):387–412
Wissenschaftsrat (1985) Empfehlungen zum Wettbewerb im deutschen Hochschulsystem. Wissenschaftsrat, Köln
Wissenschaftsrat (1992) Empfehlungen zur Internationalisierung der Wissenschaftsbeziehungen. Wissenschaftsrat, Köln
Wissenschaftsrat (1998) Empfehlungen zur Chancengleichheit von Frauen in Wissenschaft und Forschung, Mainz
Wissenschaftsrat (2002) Empfehlungen zur Doktorandenausbildung, Saarbrücken
Wissenschaftsrat (2005) Empfehlungen zur Ausgestaltung von Berufungsverfahren. Wissenschaftsrat, Köln
Wissenschaftsrat (2006) Empfehlungen zur künftigen Rolle der Universitäten im Wissenschaftssystem. Wissenschaftsrat, Köln

Performance Management Systems and their Influence on the Governance Structures of Portuguese Universities: A Case Study

Ana I. Melo and Cláudia S. Sarrico

Abstract The aim of this chapter is to understand how performance is being measured and managed in universities and the way Performance Management Systems (PMS) have affected the roles and influences of key actors in the governance of universities. Results from a high performing Portuguese university show the inexistence of a fully developed PMS. In fact, even though there has been a substantial increase in the measurement of performance in most areas, there seems to be a lack of action, especially regarding individual performance. In terms of governance, the analysis of the case study through a new governance framework shows that external pressures to implement PMS (mainly coming from the state and from European policies) are the most influential ones. Moreover, results show that there has been a centralization of authority and a change in the roles of key actors. The number of external members in the main governing bodies grew and academics' bureaucratic work increased enormously. Nevertheless, academics are still the most powerful group in terms of decision-making. The governance reforms that took place in the Portuguese higher education system enable the implementation and functioning of PMS. But other factors should also be considered, namely the level of communication and the level of stakeholder involvement. These factors help to overcome resistances and to build trust, the most difficult piece of the performance management framework.

A.I. Melo (✉)
School of Technology and Management (ESTGA), University of Aveiro, Aveiro, Portugal

CIPES – Centre for Research in Higher Education Policies, Matosinhos, Portugal
e-mail: ana.melo@ua.pt

C.S. Sarrico
ISEG – Lisboa School of Economics and Management, Universidade de Lisboa, Lisbon, Portugal

CIPES – Centre for Research in Higher Education Policies, Matosinhos, Portugal
e-mail: cssarrico@iseg.ulisboa.pt

I.M. Welpe et al. (eds.), *Incentives and Performance*,
DOI 10.1007/978-3-319-09785-5_25

1 Introduction

The nature of public services has changed over the years, mainly induced by new government orientations and social, economic and technological changes. In many of these services, a new form of management, more concerned with the organization and coordination of services towards an increased efficiency in service delivery, has been implemented (Bleiklie et al. 2000; Mwita 2000; Pollitt 2003). In order to classify and explain the reforms that took place in many countries, some authors came up with concepts such as 'managerialism' (Aucoin 1990; Pollitt 1990) or 'New Public Management (NPM)' (Hood 1991).

Similarly to what happened in many public organizations, universities also faced increasing pressures to change their 'traditional' nature (Amaral and Magalhães 2002). According to the existing literature, several exogenous forces have contributed to the urge to reform these institutions. Among these are the following: first, the shift from being 'Ivory Towers', inhabited by scholars with the liberty to pursue knowledge in a rigorous and critical way, enjoying the independence of mind that came from autonomy and intellectual freedom (Barry et al. 2001; Czarniawska and Genell 2002), to being deliverers of mass higher education (Halsey 1995); second, the increasing difficulty of exclusively financing the institutions with public funds; third, European policies; and finally, the emergence of new approaches to public policy, such as NPM (Hood 1991; Shattock 1999; Chevaillier 2002; Salter and Tapper 2002).

Growing demands to become more efficient, effective and accountable, led to an increased interest in introducing control mechanisms aimed at assessing organizational performance. As a result, Performance Management Systems (PMS) have been implemented in some universities and many of these institutions have started to rethink their forms of organization, governance and management (Vilalta 2001).

Even though many universities claim that they have implemented PMS and that they are now more accountable to their stakeholders (Melo et al. 2010), it is neither clear how performance is being measured and managed nor what the real effect of these new managerial arrangements on the governance structures of these institutions has been, particularly concerning the roles of key actors. Therefore, the central focus of this chapter is first, to understand the way performance information is being collected and used and second, to understand the extent to which the roles and influences of the main actors in the governance structures of universities have been affected by PMS.

Data from a Portuguese university is presented, one that considers itself innovative and entrepreneurial, being, in fact, the only Portuguese university that belongs to the European Consortium of Innovative Universities (ECIU).[1] As

[1] The ECIU was founded in 1997 by ten European universities, being its goal to create a European network, where participating institutions can exchange experience and best practice of projects within education, research and regional development.

such, one would expect that such an institution would have implemented adequate systems to measure, report and manage performance.

The chapter is structured in the following way: first, a systems view of performance management in higher education is introduced; second, an analytical framework incorporating the main actors in the governance structures of universities is displayed and the research questions are outlined; third, the research design, methods and setting are introduced; fourth, findings are presented according to the framework developed; and finally, results are discussed and conclusions are drawn.

2 Theoretical Background

2.1 Performance Management in Higher Education: A Systems View

For the purpose of this chapter, performance management is defined as an integrated system where performance information is closely linked to strategic steering. It consists of three stages: the first is the *measurement stage*, which involves measuring the input, output, level of activity or outcome of organizations, people and programs, thereby gathering performance information (Radnor and Barnes 2007; Askim 2008); the second is the *reporting stage*, which entails communicating performance information to decision-makers; and the third is the *management stage*, which consists of using the information and acting upon it, aiming at improvements in behavior, motivation and processes (Bouckaert and van Dooren 2003; Radnor and Barnes 2007).

Given the importance of linking the measurement process with strategic planning and the need to look at several levels of performance, it is considered adequate to use an 'input–output model' to look at the performance of universities (Fig. 1). This model, based on systems-theory, provides tools for a dynamic and systemic thinking, since it acknowledges the existence of a closed loop between the actions of performance measuring, taking corrective action and achieving outcome response (Boland and Fowler 2000). It comprises four main components: inputs, processes, outputs and outcomes.

According to this model, higher education is seen as the *process* of transforming *inputs* (e.g., students' and academics' time) into *outputs*, which can be broadly classified as relating to the three areas of every university's mission: teaching, research and third mission. *Outcomes* are the products of a university in the long run, and include, for instance, building a well-educated society (Boland and Fowler 2000). All this process is monitored and controlled. At the end, the output and the outcome are measured against pre-established targets and, if there is a difference between these and the actual outputs/outcomes, corrective action occurs.

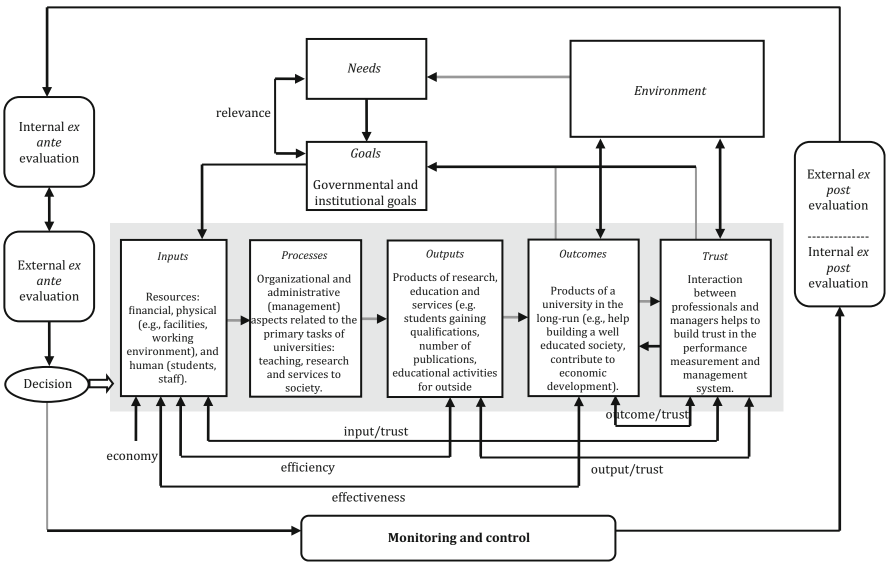


Fig. 1 A systems view of performance management in universities [Adapted from Bouckaert and Halligan (2008, p. 33) and Dochy et al. (1990, p. 145)]

Since performance measurement should not be used as an end in itself, but to provide staff with feedback designed to enable them to develop and improve their practice, generating ownership and building *trust* is essential for a well-succeeded performance management system (Bouckaert and Halligan 2008). This means that professionals ought to be invited to say what constitutes a good service and what they want to be assessed on (De Bruijn 2007). Plus, there should be a clear identification of the functions of performance measurement as well as the intended forums for dealing with the performance measurement results. In this way, the manager and professional can trust that any deviation from it will demand consultation.

In the last stage of the system's view there is an ex post audit and/or evaluation, comprising both an internal and an external dimension. In higher education these tasks can be performed by an accreditation agency. Ideally, this feeds forward to the next cycle.

In order to assess performance, some criteria are usually used. These normally relate to the three "e"s: *economy*, concerned with ensuring the lowest possible cost; *efficiency*, concerned with how much output is achieved for a given level of input at a specified level of volume and quality; and *effectiveness*, concerned with the extent to which services confer the benefits which they are intended to confer (Holloway 1999). The three "e"s relate to the more rational, hard type of control mechanism in performance management. Another fundamental dimension is trust, which represents a softer type of control mechanism of performance. Trust, when present, acts as a facilitator of the more rational dimension. Trust-based control systems rely on traditions, on professions, on standard practice. A key challenge is to keep equilibrium between these two systems, and indeed make them work in unison (Bouckaert and Halligan 2008).

If working well, a PMS should provide information on important matters, promote appropriate behavior, provide mechanisms for accountability and control, and create a mechanism for intervention and learning (Haas and Kleingeld 1998; Neely 1998); that is, it should be used for improvement purposes (Radnor and Barnes 2007). But that seldom happens. In fact, some authors (e.g., Radnor and McGuire 2004; Hood 2006; Laegreid et al. 2008) argued that the focus for many public service organizations is on measurement, leading to an excessive amount of data collected with little action.

But, how are these systems working inside universities? And, how have they influenced key actors in the governance of universities?

In order to answer these questions it is essential to understand the way universities are governed.

2.2 Governance Structures in Higher Education

De Boer (2002: 44) regards governance structures as a 'set of rules concerning authority and power related to the performance of a university's activities directed towards a set of common goals'. It reflects the way an organization divides and integrates responsibility and authority.

As Fig. 2 illustrates, governance structures can be conceptualized by an 'inner ring' and an 'outer ring'. The 'inner ring' represents the internal coordination mechanisms, and is composed of the members of the university's governing bodies—the 'four Estates'. These are: students, academics, non-academic staff and external representatives.

The 'outer ring' embodies the external coordination mechanisms and is composed of the state, Europe (understood as European policies) and the market.

This model extends Clark's (1983) 'triangle of coordination' to other internal stakeholders of the university, revisiting the concept of the university's Estates proposed by Neave and Rhoades (1987). According to Neave (2009), the characteristic of an Estate is the central part played by prescribed and formal status. To Neave (2009) there has been a move from what Clark (1983) called 'academic oligarchy' to an extended constituency in which all three Estates—Academic, Student and Administrative—have their formal elected place. To the three Estates, this research adds a new one—the 'External Representatives Estate', since these members have become increasingly important in the governance and management of universities.

Therefore, Fig. 2 is proposed as the analytical framework that will help to understand how the introduction of the performance management system described above has affected governance structures and the roles, influences and accountabilities of the key actors (the 'inner ring').

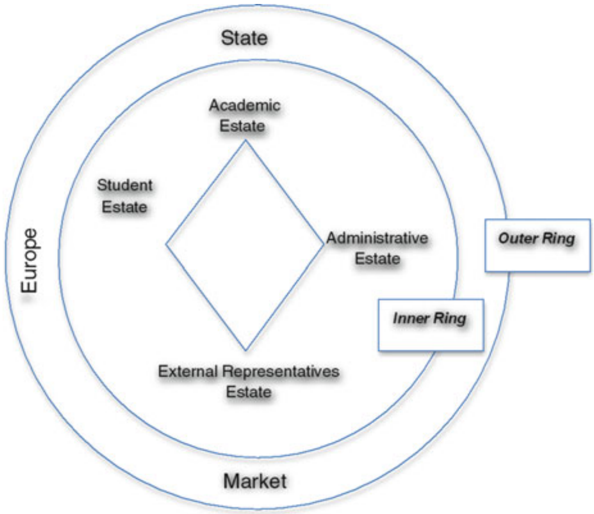


Fig. 2 Governance structures in higher education

3 The Case Study

3.1 Research Design and Methods

To illustrate what is happening, it was decided to study the case of a Portuguese university (PU) in-depth. The Portuguese higher education system was chosen because it has recently gone through major reforms, which aim at putting progressively in place mechanisms of control, in order to fulfill demands for increased accountability. PU was selected because it is an entrepreneurial organization, recognized for its good performance, and it was one of the few that has decided to become a public foundation subjected to private law, which means that it has to raise at least 50 % of its revenue. It would be expected that such a university would have implemented mechanisms to measure and manage performance.

In relation to performance, it has been regarded as a system, using two main dimensions: the *collection* and *use* of performance information data. The key actors studied are the four Estates that compose the ‘inner ring’ (see Fig. 2): academics, non-academic staff, students and lay members.

Mixed methods were used to assemble data. These comprised documentary analysis and interviews. The documents analyzed included policy and strategic documents, minutes of meetings, the results of internal surveys, and statistical data collected from secondary sources. The semi-structured interviews were conducted to 39 members of the four groups that sit in the governing bodies of a university: academics ($n = 24$), non-academic staff ($n = 9$), students ($n = 4$) and external representatives ($n = 2$). The number of interviews conducted to each group was related to their weight inside the existing governance structures. The interviews were all recorded and transcribed, totalizing 42 h of recordings. This meant that each interview lasted, on average, 1 h and 7 min.

The interview schedule integrated three main sets of questions. The first group of questions was centered on performance measurement and management practices, at all levels. In order to help the interviewees answer these questions, a prompt card was shown to them, integrating: the main activities that compose the mission of a university (teaching, research and third mission), the employees of a university (academic and non-academic staff), ‘customers’ (students), services and finance. The second set of questions focused on the pressures (both internal and external) felt to measure and manage performance. The last set of questions was mainly related to the influences exerted on decision-making by each one of the Estates; and the impact the introduction of measurement and management practices might have had on these groups.

All the quotations used were coded in order to ensure confidentiality. S refers to students, L to lay members, NA to non-academic staff and A to academics.

3.2 Research Setting

The chosen university—PU—was established in the early 1970s. It is divided into 17 different departments, and has around 13,500 students. It employs nearly 1,500 members of staff, which comprise, approximately, 1,100 academics.

At the time the interviews took place (between January and June 2009), this institution had three main decision-making bodies at the central level. First, there was the University Assembly, which was composed of a large number of members (approximately 110, comprising academics, students and non-academic staff). This body only held formal meetings on occasions such as the election of the Rector or the approval of the University’s Statutes and their alteration. Second, there was the University Senate, with nearly 50 members, integrating academics, non-academic staff, students and lay members. This was the most important collective decision-making body, since it decided not only on academic matters, but also on the approval of the budget, annual plans and strategic plans. Finally, there was the Rector, who presided over the Senate. The Rector, who was elected, appointed high-level institutional officers (Vice-Rectors and Pro-Rectors).

3.3 Performance Measurement and Management Practices

3.3.1 How is Performance Measured?

Results show that the degree of measurement and the way performance information is used vary considerably according to the area.

Teaching and learning is mainly assessed through students' feedback. At the end of each semester, the university asks students to fill in a questionnaire about the content of the course and about academics. Then, the Office of Information Management summarizes data from the questionnaires in a graphical form and a score from one to nine is given.

Professional bodies also used to visit the university periodically, in order to measure the effectiveness of teaching and learning and accredit degrees. Since its creation, the Agency for Assessment and Accreditation of Higher Education has suspended all the accreditation procedures performed by professional bodies.

Internally, there were two Vice-Rectors responsible for education, one for undergraduate degrees and the other for post-graduate degrees, and three institutes (IFIU, IFPG and IFSP) responsible for coordinating teaching and learning within the university. Some of the interviewees felt that sometimes there was an overlap between the activities developed by IFIU and the ones carried out by the Pedagogic Council. Additionally, the Office of Information Management also collects data on teaching and learning (e.g., success rates and retention rates).

Since 2005, there has been no evaluation of programs or degrees at a national level. There used to be a global coordinating body of the evaluation system called the National Council for the Evaluation of Higher Education. This body had to assist and assure the credibility of the process of higher education, and to review and report on the quality assurance procedures. However, according to ENQA (2006), follow-up of the assessments was inexistent and, in many cases, the reports failed to provide consistent, clear and sufficient information to the stakeholders. Most significant was the general perception that the evaluation results had no consequences, since there were no plans of action drawn up to overcome or attenuate weaknesses or reinforce strengths.

Following the recommendations of the European Association for Quality Assurance in Higher Education (ENQA), the Ministry Science, Technology and Higher Education created the Agency for Assessment and Accreditation of Higher Education (Decree-Law 369/07). The Agency has recently started a preliminary process of accreditation.

Research and scholarship is reviewed mainly because of an external audit performed by the Foundation for Science and Technology to Research Units. To an academic 'this is the most evaluated area' (A67). The evaluation system on which it is based comprises a periodic evaluation by panels of international experts, which include direct contact with the researchers through visits to all units. This process culminates with the panel attributing a qualitative grade, which determines funding. The parameters that are looked at are: size of research contracts and quality of research outputs, translated by the number of publications, citations and the impact factor of journals.

Internally, there was one Vice-Rector looking after research and there is a Research Institute, which was responsible for coordinating research activities within the university. The Office of Information Management also collects statistics on research, but only on demand. To gather that information it usually uses secondary sources, such as Citation Indexes.

Regarding *third mission*, it is consensual that no measures are used, apart from financial measures (e.g., income generated by services provided). 'It is merely impressionistic' (A42). In governance terms, there is one Vice-Rector responsible for it.

In practice, *academics*' performance is measured in terms of research, providing information on publications, supervision of postgraduate students, coordination of projects and research grants, and in terms of teaching, through the feedback obtained from students at the end of each semester. Historically, research has been the most important issue, not only because it is the most easily measured, but also because it is the one that contributes the most to career progression. The Office of Information Management and the Office of Human Resources also collect statistics on academics, such as numbers, categories and salaries, sorted by department. Nevertheless, most interviewees agree 'the evaluation methodology is not adequate' (L82).

Students are very closely monitored in terms of numbers, degree results, completion rates and retention rates, even though there is no follow-up of students after graduation. The Office of Information Management and the Academic Office gather that information, mainly due to the need to send it to Ministry on a yearly basis.

Non-academic staff is effectively measured through a national system developed by Government to assess public servants (the Integrated System for the Evaluation of Performance within Public Administration—SIADAP). Within central administration, each member of staff discusses objectives with the Director of his or her service and ways to reach them, being their performance then compared with pre-established objectives. Within departments, non-academic staff agrees their objectives with the head of department, being then assessed by him or her at the end of each year. The performance of each member of staff can be considered 'Non-Adequate' (1–1,999), 'Adequate' (2–3,999) or 'Relevant' (4–5). Only 25 % of the workers can be awarded a grade between four and five, and, only a maximum of 5 % can be considered 'Excellent' (5). The Council responsible for coordinating the evaluation process within the university receives all the results and revises them. All the interviewees consider SIADAP an extremely bureaucratic system that is 'too time consuming for what you get from it' (A69). The Office of Information Management and the Office of Human Resources also collect statistics on non-academic staff, such as numbers, categories and salaries.

Support services are assessed at a central level though the evaluation of the Director of each service by the Administrator. At a departmental level, the head of department assesses each service. Apart from that, there are two support services that launch satisfaction surveys on an annual basis. That is done voluntarily.

In relation to the performance of support services, the university is starting to implement the Evaluation and Accountability Framework. This system, integrated in SIADAP, was developed by Government to assess the performance of public services, at a national level. The Office for Quality, Evaluation and Procedures is coordinating its development inside the university.

It is consensual that *finance* is clearly measured both at university and departmental levels. The key performance indicators used by the Finance Office are developed by that office and have been agreed upon by the Administrator. These have to comply with the ones defined by the law that regulates public spending. The Consolidated Results were then presented to Senate and ratified by this body, after being certified by an external auditor. The Office of Information Management gathers all financial information. Departments have their own budgets and have some autonomy in running those budgets, within pre-established rules.

3.3.2 How is Performance Information Used? Who Uses that Information?

Data on *research* is closely analyzed, especially due to the need to prepare for external evaluation exercises. Results obtained on this area were generally looked at by the Research Institute and used by research units to improve their performance.

Although some data is collected on *teaching and learning and academic staff* (through students' feedback), it seems consensual that departments do very little with it. Even though some heads of department have taken some action concerning data from the questionnaires, such as, changing the courses lectured by a particular teacher, this seldom happens. If academics have tenure, it is still not very clear what can be done, since it depends on a clearer definition of the Statutes that regulate the academic profession. In addition, there seems to be 'a lack of legitimacy to act upon the data collected from the questionnaires, especially if heads of department are not full professors' (A47). This is particularly problematic in relatively new departments, which do not have a huge number of full professors.

Moreover, the majority of the interviewees question the validity of the questionnaires, since the way they are administered does not guarantee the representativeness of the sample and its response rates are quite low. According to a student 'there is not an evaluation culture in the university. . . [and] there has not been enough time to find the best way to explain each student that there may be direct consequences from their participation' (S46). In fact, many of the interviewees feel that students' will to participate in the academic life has decreased over the years. Several explanations have been advanced to account for these changes: first, the shorter period students spend at the university, after the implementation of the Bologna Process; second, the lack of information provided to incoming students concerning the mission, development plan and governance structures of the university; third, the increasing competition for jobs; and fourth, the pressure exerted by families for students to finish their degrees as soon as possible.

In addition, the majority of the interviewees regard the questionnaires as unfit for purpose. They are considered too long, discouraging students from filling them in. Plus, some people believe there should be complementary tools to assess the quality of teaching and learning. An interviewee stated: 'The questionnaires are not enough. . . a different system must be used to validate results' (A39).

Therefore, results on academics' performance seem to be only used in terms of career progression. In fact, their curriculum is thoroughly evaluated when they apply for a position, especially their research activities. Pedagogic activities and management activities account very little for career progression, which may lead to a perverse effect: academics may feel tempted to focus on research and spend less time preparing their courses or performing management duties.

A lot of data is collected on *students*, even though not much is done with it.

In relation to poor-performing *non-academic staff*, bad evaluations in the Integrated System for the Evaluation of Performance within Public Administration have a direct impact on their promotion.

Data concerning *support services* and *finance* is analyzed and used for improvement purposes by the central administration.

Third mission is not subjected to a lot of measurement, and the scarce data that gets collected seems to be neglected.

3.3.3 Pressures to Measure and Changes in the Roles, Influences and Accountabilities of Key Actors

The role of measurement in universities has changed considerably, over the last years, especially with the creation of new laws, which demand more efficiency, effectiveness and accountability from universities, and with the introduction of external audits.

Externally, pressures come from the state, Europe and the market (the 'outer' ring of Fig. 2). Internally, 'there is a general perception in the university, mainly among those linked to the management of the institution, that the implementation of a performance management system is needed to facilitate decision-making' (A38). In fact, this university has willingly asked to be evaluated by the European University Association (EUA 2007).

External Pressures

State

Most interviewees feel that the main pressures to measure come from Government and are imposed on universities by law: 'Law imposed the big changes that are happening, otherwise no one would move!' (NA50)

In fact, a lot of legislation came out since 2007, creating a new evaluation framework for Portuguese universities; the Agency for Assessment and Accreditation of Higher Education, believed to be the main pressure; a new juridical regime for universities; and a new Statute for Academic Careers. These laws led to vast reforms in the Portuguese higher education system.

Europe

The European Commission published a modernization agenda for universities, which was welcomed by the member states and the main stakeholders in higher education. The main fields of reform were: curricular reform (also promoted through the Bologna Process); governance reform, accomplished through more university autonomy, strategic partnerships (e.g., with businesses) and quality assurance; and funding reform, which means finding diversified sources of income better linked to performance. It became clear that the implementation of these reforms needed to be assessed, demanding increased measurement. As an interviewee put it: 'There have been international pressures (. . .) mainly European [ones]' (A47).

Market

Even though some interviewees stated the market had not a strong influence in Portugal and that it had not influenced what happened inside the university, almost all agreed that the competition between universities increased, and some even argued that this competition has 'benefited universities and other sectors of the Portuguese economy' (NA55). To the interviewees: 'the number of students is decreasing and the students that exist are just those. (. . .) We are competing for the same universe' (A67). Actually, some departments struggled to get students. That was why a number of interviewees mentioned that attracting students became an important issue and that universities felt the pressure to be better, which meant 'reinforcing their marketing' (A43) and 'image' (A80).

Additionally, society started to demand more accountability from public institutions, especially educational ones: 'We have to be accountable to taxpayers, to justify the money spent' (S75).

But how have these pressures changed the roles of the key actors included in the 'inner ring' of governance?

Changes in the Roles of the Estates

Academic Estate

Several interviewees argued that the role of being an academic changed. They stated that academics are now held more accountable for their actions, meaning their students increasingly assess them. To an interviewee this does not mean they have lost their autonomy:

> I do not think they have lost their autonomy.. . . What they have now realized is that they cannot have the same future that academics had 20 or 30 years ago. . .. Today, an Assistant Professor does not know what chances he or she has to progress in his or her career. (A45)

This uncertainty has, to a number of interviewees, increased competition between academics.

Moreover, academics felt they were increasingly asked to perform bureaucratic tasks, including work related to performance management, and expected to perform other roles (e.g., management roles), which, according to them, deviated their attention from research and teaching, and for which they were not adequately rewarded in terms of career progression.

Although most interviewees agreed that academics were now more assessed (especially due to the introduction of the questionnaire about the content of the course and about academics), there seemed to be, according to them, little consequences for poor performers, rather than some not frequent 'internal reengineering' (e.g., attributing courses to other academics).

Additionally, academics also worried about having to 'share' their decision-making power with external members, whose role increased inside the university's governance structures. Nevertheless, they were still believed to be the most powerful group: 'It is obvious that at the end of the day, the power lies with academics' (A49).

Administrative Estate

With the introduction of the Integrated System for the Evaluation of Performance within Public Administration, non-academic members of staff became more assessed than before. The introduction of this system raised competition and created a bad environment within services and departments, according to some interviewees. This group felt their efforts were not recognized and felt disappointed and discouraged. Moreover, similarly to academics, they felt the workload increased enormously and many of them highlighted the need to increase the number of non-academics in the university, especially when compared to the number of academics (three times more). Additionally, their representativeness in university governing bodies was reduced to one member.

Student Estate

Students' roles and influences have changed over the years, especially since they started to be seen as 'consumers' of higher education. Therefore, in the last years, attracting students and maintaining them satisfied became a concern of every university.

In fact, and even though students always participated very actively in decision-making within this university, they became more influential in terms of assessment with the introduction of the questionnaire to assess courses and academics. Actually, their opinions became the main tools used to assess teaching and learning (even though most interviewees questioned the validity of this tool, as explained before). Nevertheless, they still were not believed to be very influential in terms of strategic management.

External Representatives Estate

The number of lay members in governing bodies grew with the introduction of new governance structures and their role has been enhanced within the university. They

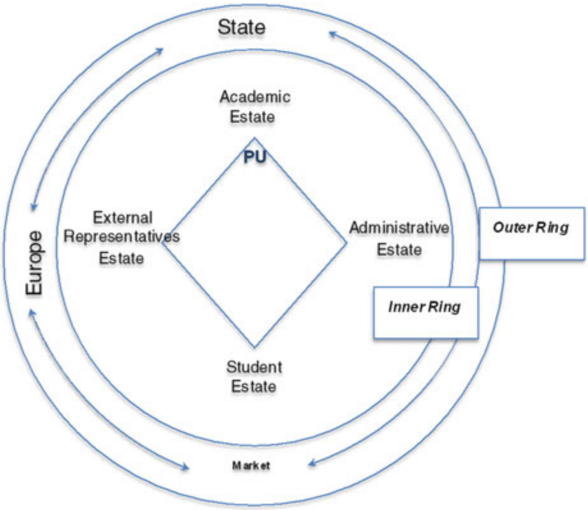


Fig. 3 The positioning of PU in the framework of governance structures in higher education

now represent almost 30 % of the members that sit on the General Council, the ultimate decision-making body, and one of them chairs it.

They have been co-opted by the university, being most of them from the region where the university is located. They were chosen mainly because of their prestige and connections.

They are considered important to the university by most interviewees, especially since they bring in completely different perspectives from insiders: 'It is important for the university to be connected to the surrounding environment' (S51).

The results obtained thus confirm the suitability of the new governance framework proposed in the chapter to study the governance arrangements of universities. In fact, after analyzing the external coordination mechanisms and the role and influence of the four Estates in decision-making, this university (PU) can be placed in that framework, as shown in Fig. 3.

The state and Europe are the main external coordination mechanisms, with the role of the market starting now to emerge.

The positioning of PU against the academic estates' corner shows that even though the number of external members increased in the main governing body, the Academic Estate is clearly the dominant one.

Discussion and Conclusion

Through a systems view of performance management and the presentation of a new governance framework for universities, this study contributes to the research on the way performance is measured and managed in universities and to the research on the effect performance management systems might have had on the roles and influences of key actors in the governance of universities.

(continued)

Data analysis reinforces the findings of some authors (e.g., Vilalta 2001), which state that there has been a substantial increase in the measurement of performance in the university over the years. In fact, data showed that more areas are now assessed (third mission was clearly an exception), albeit many interviewees agreed that better measures could be in place in some of them.

The increased level of measurement was greatly influenced by the external environment, resulting mainly from European policy, namely the Bologna Declaration; and from the state, which published a lot of legislation in the last years. Additionally, the role of the market, even though minimal, started to show, as competition for students became tougher. Therefore, it could be argued that the main pressures to measure came mainly from Europe and from the state, being little influenced by the market (see Fig. 3). Internally, a new Contract-Program, which integrated some objectives, indicators and targets, led to a different attitude in terms of the need to measure performance. Moreover, interviewees also expected the external members of the General Council to push more towards the introduction of control mechanisms. Thus, similarly to what van Dooren (2006) found out for public services, performance management in this university became more systematic and institutionalized.

Concerning the management of performance, many of the interviewees mentioned the lack of use of performance data, especially regarding the individual performance of both academic and non-academic staff. Thus, the closed loop between measurement and taking corrective action, acknowledged by Boland and Fowler (2000), does not exist. The reasons presented were mainly related to the legal framework, which was considered very protective. These findings are consistent with some literature on the public sector (e.g., Radnor and McGuire 2004 and Hood 2006), which suggests an excessive focus on measurement with little action, and can arguably be extrapolated to other universities. In fact, if individual performance information is not used very much in a university that is an 'extreme case' (Yin 2003), one might expect that it will be used even less in universities that are less entrepreneurial.

In terms of the components of the performance management system (see Fig. 1), findings indicate a concern mainly with outputs, with several areas being measured (with the exception of third mission). Data shows little preoccupation with inputs and processes; outcomes are also not measured, given the difficulties in doing so.

Given these findings, it can be stated that performance is not managed in a systemic way in our case study, as presented in Fig. 1.

In relation to governance, it was apparent that the introduction of control mechanisms led to some changes in the governance of the university, following the general trend towards the centralization of authority in the

(continued)

institution-level governing structures (the Rector, for example, has now more power than before). These results reinforce Vilalta's (2001) findings, which state that with the introduction of control mechanisms, many universities started to rethink their forms of organization, governance and management.

Even if it is still early to understand the real impact of the new structure (imposed on universities by law in 2007 and implemented in 2009), it was generally regarded by the interviewees as more efficient, given the decrease in the number of committees and in their membership. Moreover, the leadership structure was considered clearer and the participation of the outside world bigger. The lighter, more centralized, and more externally participated structure was thought to enable more strategic decision-making and provide increased strategic coherence, which were considered fundamental for the introduction and functioning of a performance management system.

Although, there were considerable changes in the university's structure, essentially driven by European and national interests, the introduction of measurement and management practices also led to changes in the roles played by the Estates involved in the governance and management of the university, with the exception of the Student Estate, who rarely changed its role in terms of decision-making.

Concerning the Academic Estate, the bureaucratic work demanded from academics increased a lot. Academics were increasingly expected to perform other roles (e.g., management roles). In their view this did not necessarily lead to an increase in the quality of their teaching and research, since it left them less time to focus on these tasks. Some academics also mentioned the possibility of a decline in the 'academic-voice' in institutional decision-making. Nevertheless, it was noted that they still have the most active voice in the university and that the 'collegial type' of coordination still persists at this institution (shown by the positioning of PU against the academic estates' corner in Fig. 3).

Concerning the Administrative Estate, and although non-academic staff were never very powerful inside the university, now they are even less represented in governing bodies.

In relation to the External Representatives Estate, the presence of external members increased significantly, even though they felt they did not participate very much in strategic decision-making.

Although it is acknowledged that the governance reforms that took place in many higher education systems—more institutional autonomy from the state, increased centralization of decision-making inside the institutions, stronger leadership at the top, increased accountability and wider participation of external members—are enablers for the implementation and good functioning of performance management systems, there are still other variables to take into consideration. Two important variables seem to be the level

(continued)

of communication and the level of stakeholder involvement. In fact, a good level of communication between bodies, between units and between individuals, and the involvement of different actors in the development of such a system will arguably overcome resistances and build trust. Trust is the most difficult piece to develop of the performance management framework, as devised in Fig. 1, but arguably, as discussed previously, a crucial one. As Thomas (2004) argued, an ideal performance management system should be embedded in the organization, stable and widely understood and supported.

The work presented contributes to a better understanding of performance management practices in universities. It was developed in the Portuguese context and complements previous work in a British context (Melo et al. 2010). However, it is based on a single case study, albeit in-depth. As future work, we envisage a more extensive research project, using survey methods, of a more representative sample of European universities.

References

Amaral A, Magalhães A (2002) The emergent role of external stakeholders in European higher education governance. In: Amaral A, Meek VL, Larsen IM (eds) Governing higher education: national perspectives on institutional governance. Kluwer, Dordrecht, pp 1–21

Askim J (2008) Determinants of performance information utilization in political decision making. In: van Dooren W, van de Walle S (eds) Performance information in the public sector: how it is used. Macmillan, Houndmills, pp 125–139

Aucoin P (1990) Administrative reform in public management: paradigms, principles, paradoxes and pendulums. Governance 3(2):115–137

Barry J, Chandler J, Clark H (2001) Between the ivory tower and the academic assembly line. J Manag Stud 38(1):87–101

Bleiklie I, Hostaker R, Vabo A (2000) Policy and practice in higher education: reforming Norwergian universities. Jessica Kingsley, London

Boland T, Fowler A (2000) A systems perspective of performance management in public sector organizations. Int J Pub Sector Manag 13(5):417–446

Bouckaert G, Halligan J (2008) Managing performance: international comparisons. Routledge, New York

Bouckaert G, van Dooren W (2003) Performance management on public sector organizations. In: Löffler E, Bovaird T (eds) Public management and governance. Routledge, London, pp 127–136

Chevaillier T (2002) University governance and finance: the impact of changes in resource allocation on decision making structures. In: Amaral A, Jones G (eds) Governing higher education: national perspectives on institutional governance. Kluwer, Dordrecht, pp 87–98

Clark BR (1983) The higher education system: academic organization in cross-national perspective. University of California Press, Berkeley

Czarniawska B, Genell K (2002) Gone shopping? Universities on their way to the market. Scand J Manag 18:455–474

De Boer H (2002) Trust, the essence of governance? In: Amaral A, Glen AJ, Karseth B (eds) Governing higher education: national perspectives on institutional governance. Kluwer, Dordrecht, pp 43–61

De Bruijn H (2007) Managing performance in the public sector. Routledge, London

Dochy FJRC, Segers MSR, Wijnen WHFW (1990) Selecting performance indicators: a proposal as a result of research. In: Goedegebuure L, Maasen P, Westerheijden DF (eds) Peer review and performance indicators. Lemma, Utrecht, pp 135–153

ENQA (2006) Quality assurance of higher education in Portugal: an assessment of the existing system and recommendations for a future system. ENQA, Helsinki

EUA (2007) EUA evaluation report: university of Aveiro. EUA, Brussels

Haas MD, Kleingeld A (1998) Multilevel design of performance measurement systems: enhancing strategic dialogue through the organization. Manag Account Res 10:223–261

Halsey AH (1995) Decline of donnish domination. Claredon Press, Oxford

Holloway J (1999) Managing performance. In: Rose A, Lawton A (eds) Public services management. Pearson, Harlow, pp 238–259

Hood C (1991) A public management for all seasons. Public Adm 69(1):3–19

Hood C (2006) Gaming in the targetworld: the targets approach to managing british public services. Public Adm Rev 66(4):515–521

Laegreid P, Roness PG, Rubecksen K (2008) Performance information and performance steering: integrated system or loose coupling. In: van Dooren W, van de Walle S (eds) Performance information in the public sector: how it is used. Macmillan, Houndmills, pp 42–57

Melo AI, Sarrico CS, Radnor Z (2010) The influence of performance management systems on key actors in universities: the case of an english university. Pub Manag Rev 12(2):233–254

Mwita JI (2000) Performance management model: a systems-based approach to public service quality. Int J Pub Sector Manag 13(1):19–37

Neave GR (2009) The academic state revisited: reflections on academia's rapid progress from the Capitoline Hill to the Tarpeian Rock. In: Enders J, de Weert E (eds) The changing face of academic life: analytical and comparative perspectives. Macmillan, Basingstoke, pp 15–35

Neave GR, Rhoades G (1987) The academic estate in Western Europe. In: Clark B (ed) The academic profession: national, disciplinary and institutional settings. University of California Press, Los Angeles, pp 211–270

Neely A (1998) Measurement business performance – why, what and how. Economist Books, London

Pollitt C (1990) Managerialism and the public services: the anglo-american experience. Blackwell, Oxford

Pollitt C (2003) The essential public manager. Open University Press, Maidenhead

Radnor Z, Barnes D (2007) Historical analysis of performance measurement and management in operations management. Int J Product Perform Manag 56(5/6):384–396

Radnor Z, McGuire M (2004) Performance management in the public sector: fact or fiction? Int J Product Perform Manag 53(3):245–260

Salter B, Tapper T (2002) The external pressures on the internal governance of universities. High Educ Q 56(3):245–256

Shattock M (1999) Governance and management in universities: the way we live now. J Edu Policy 14(3):271–282

Thomas P (2004) Performance measurement, reporting and accountability: recent trends and future directions. The Saskatchewan Institute of Public Policy, Regina

van Dooren W (2006) Performance measurement in the Flemish public sector: a supply and demand approach. Ph.D. thesis, K.U. Leuven, Leuven

Vilalta J (2001) University policy and coordination systems between governments and universities: the experience of the Catalan university system. Tert Educ Manag 7:9–22

Yin RK (2003) Case study research, design and methods. Sage, Newbury Park

Performance Management and Disciplinary Efficiency Comparison

Matthias Klumpp

Abstract Performance measurement and management in universities in a disciplinary perspective is a newly established perspective due to increasing interest in line with international university rankings as well as public management concepts. From the perspectives of methodology as well as data access and quality the disciplinary performance and productivity measurement poses several severe hurdles to research and practice. This chapter outlines conceptual views as well as a data envelopment analysis for four science disciplines in German universities as well as universities of applied sciences.

1 Introduction

The higher education management question of *performance management* is crucial and difficult at the same time—mainly due to a high diversity of objectives and performance fields (Dundar and Lewis 1995; Lindsay 1982; Sarrico 2010). For example, typically different inputs (researchers, students, information, laboratories and other resources) as well as different—and even in terms of input time contradicting—outputs (e.g., publications, graduates, citations, patents, transfer co-operations, community services) are produced in order to align with expectations of stakeholders within society.

This *complexity* is increasing due to a rising number of stakeholders as well as increasing expectations towards universities as knowledge and innovation hubs from societies and political institutions. Therefore, besides the traditional fields of research, teaching and third mission (dissemination and transfer), new objectives and performance fields are becoming important (Korhonen et al. 2001)—as highlighted, for example, by the inclusion of gender indicators into university indicator based budgeting schemes in Germany.

M. Klumpp (✉)
University of Duisburg-Essen, Campus Essen (PIM) & FOM University of Applied Sciences Essen (ild), Essen, Germany
e-mail: matthias.klumpp@pim.uni-due.de

I.M. Welpe et al. (eds.), *Incentives and Performance*,
DOI 10.1007/978-3-319-09785-5_26

The question of university efficiency is even more difficult on a comparative *faculty* i.e., *disciplinary* level, e.g., regarding university budget allocation decisions among different faculties, schools and academic disciplines as their objectives and output indicators differ even more. This contribution therefore suggests the use of the non-parametric efficiency analysis method, data envelopment analysis (DEA; Charnes et al. 1978; Lewin and Morey 1981; Kocher et al. 2006), in order to compare distinctive output and performance indicators across different disciplines without an external or given weighting among specific indicators. This is important because a given weighting say, for example, of book versus journal publications or third party funding versus graduates may strongly affect the final efficiency analysis due to the discipline-specific differences in these output fields. Using example university data it is shown that the DEA method may be very promising for university as well as faculty management in order to provide a "fair" performance measurement facing different output categories and indicators across disciplines.

2 University and Performance Measurement: The Disciplinary Gap

Methodologies used in measuring the efficiency of university operations have been manifold (Glass et al. 1998; Kao and Hung 2008; Zangoueinezhad and Moshabaki 2011)—and have interestingly many similarities to ranking endeavors in the output field. In a theoretical overview there are four distinctive areas (A to D) for *performance and productivity measurement*:

(A) Simple one-dimensional outputs as *performance measurements* with just one output indicator are quite often used in higher education management and policies, e.g., for comparing universities (or departments thereof) regarding their number of graduates per year; or universities, faculties and even research groups regarding the number of publications, patent registrations or citations per year. For third mission activities, indicators such as number or turnover of spin-offs or the total number of their employees are used to measure performance on a university or faculty level.

(B) Usually, most university and even faculty ratings use a number of output *indicators combined* in relation to the specific objective of the ranking (i.e. Van Vaught and Ziegele 2012). For a ranking of teaching quality a combination of teacher-student-ratio, student satisfaction, international orientation and expert reputation might be used; for a research ranking a combination of industry income (third party funding), publications, citations and peer reputation could be applied (Shin and Totkoushian 2012). The most often used method to calculate the overall score for such combined indicator rankings is a weighted scoring system.

(C) *Simple productivity metrics* usually operate with a relation between one output indicator (e.g., number of publications) and one input indicator (e.g., one researcher per budget amount). Essential for the distinction between performance and productivity measurement (efficiency) is the inclusion of an input indicator, commonly addressed as the 'size question' (as usually performance indicators favor larger institutions or units which more easily reach higher output numbers, for example, in terms of graduates or publication numbers). Though the division of output number by input numbers is used quite often, theoretically also the division of inputs by outputs is feasible and may sometimes also yield interesting insights.

(D) For the inclusion of *multiple input and multiple output indicators,* a number of methods are available in order to calculate a single measurement result. The two mostly used are stochastic frontier analysis (SFA) and data envelopment analysis (DEA):

  (i) SFA: The *stochastic frontier analysis* uses a given production function in order to calculate productivity measures from input and output data (Kumbhakar and Lovell 2000). If such a production function is known this is a feasible method, as it indicates clearly the improvement potential for all non-efficient units (Jacobs 2001; Stevens 2005). But if there is no known production function for inputs and outputs this is less valuable though assumptions may be made (Coelli 1995).

  (ii) DEA: The *data envelopment analysis* was proposed in 1978 as a non-parametric multi-criteria efficiency measurement method (Charnes et al. 1978, 1991; Seiford 1996; Thannasoulis et al. 2008; Zhu and Cook 2007). It is commonly used in multi-dimensional output industries such as service industries (i.e. health care: Butler and Li 2005) and also education and higher education (i.e. McMillan and Chan 2006; McMillan and Datta 1998; Ng and Li 2000; Sarrico et al. 1997; Taylor and Harris 2004).

In discussions regarding the different fields of measurement, it is usually acknowledged that *single* output indicators cannot quite depict the complex task of a university, especially since they do not take into account the distinction between the objective areas of research, teaching and third mission, neglecting the Humboldt Principle of an unity of these areas within universities (Sarrico 2010). Additionally, with just one *output* measurement the size of the higher education institution is crucial—larger universities may have an advantage in this perspective as we experience in ranking evaluation and the ensuing rush for mergers in order to reach world-class universities. Therefore it is obvious that in developing adequate measurement and comparison systems should focus on category D with *multiple input and multiple output measurements*.

Whereas for the university *institutional level* many performance and efficiency analyses are existing for several countries in an international perspective (e.g., Abbott and Doucouliagos 2003; Beasley 1995; Worthington and Higgs 2011; with one of the first being Ahn et al. 1989), the *disciplinary perspective* is seldom to be encountered in quantitative performance and efficiency measurement. Especially *international comparative* studies use only institutional data. Access to data on more detailed levels (disciplines, schools, faculties) is not available (i.e. Agasisti and Johnes 2009). The first step in this direction was taken in the framework of international rankings such as the Shanghai ARWU ranking (Academic Ranking of World Universities)—this league table started in 2007 and 2009 to establish also “field” and “subject” rankings (ARWU 2014; Liu and Cheng 2005) in order to subdivide the institutional level into the science disciplines, in this case the *fields* Natural Sciences and Mathematics (SCI), Engineering/Technology and Computer Sciences (ENG), Life and Agriculture Sciences (LIFE), Clinical Medicine and Pharmacy (MED) and Social Science (SOC). A further high-esteemed international ranking, the *CWTS Leiden Ranking 2013* of Leiden University in the Netherlands also features newly established *discipline discrimination* with the following categories obviously based on the ARWU listing with the categories Natural Sciences and Engineering, Mathematics and Computer Sciences, Life and Earth Sciences, Biomedicine and Health Sciences and Social Science and Humanities. This clearly proves the force of a trend towards performance and excellence measurement on a *disciplinary* basis.

3 Indicators and Data

The proposed institutional efficiency analysis is being conducted by utilizing a method from the DEA method family to calculate the relative efficiency of higher education institutions (Cooper et al. 2000; Madden et al. 1997; Sarrico and Dyson 2004; Taylor and Harris 2004). A major focus of DEA evaluations is the question of *viable objectives and performance data*, especially if universities from different countries are involved (Ramsden 1994; Sarrico et al. 2009; Stahl et al. 1998; Worthington and Lee 2008). The research question in general is usually hampered by the lack of quantitative comparative data on a disciplinary level. As in many cases the problem is constituted not by missing data but the *access* to *standardized* data comparable between different institutions of higher education (Kosmützky and Krücken 2014; Reale 2014). For the universities in Germany, several output and performance indicators can be discussed. Academic references point towards publications (e.g., Johnes and Johnes 1992) and third party funding (e.g., Jongbloed and Vossensteyn 2001) for the field of research and graduates (e.g., Brennan et al. 1994) for teaching. From these, the performance areas of third party funding 2009 (state competitive research funding) can be used due to a very detailed data supply from the German Deutsche Forschungsgemeinschaft (DFG) as main funding body in Germany (DFG, 2013). Graduate data 2012[1] from the federal data agency in Germany can be used for a performance comparison (Statistisches Bundesamt

[1] The time gap between 2009 and 2012 can be connected to the distinctively longer time lag of teaching as teaching processes take significantly longer to “produce” graduates, assumed between *two* (master) and *five* years (Ph.D.) – whereas third party funding usually is registered about up to *one year* after the proposal work input regarding the acquisition of research funding.

2014) as those datasets are highly standardized and comparable among different institutions. These output indicators can be combined with the number of professors as input indicator for a productivity measurement (DFG 2013—data 2009). As only the third party funding and professor numbers are subdivided by the four disciplines discriminated by the DFG (*humanities, life sciences, natural sciences, engineering*), the efficiency measurement with DEA will focus on *establishing the differences* of efficiency results for the listed German universities *with* and *without* these discipline categories.

The following indicators are finally used in the calculation as outlined in Sect. 4: Number of professors (FTE), Number of professors (FTE) in the four disciplines (*humanities, life sciences, natural sciences, engineering*), Number of graduates on PhD level, Share of female graduates on PhD level, Number of graduates on master level, Number of graduates on bachelor level, Third party funding in total, Third party funding in total in the four disciplines (*humanities, life sciences, natural sciences, engineering*).

Based on the dataset provided by DFG (third party funding, professor by discipline, data 2009), 88 universities and universities of applied sciences are listed due to the following data quality improvement steps: Addition of graduates numbers (total, Ph.D., MA, BA) from the German federal statistics offices; the Lausitz FH (UAS) was deleted due to the recent merger with TU Cottbus; for the other *universities of applied sciences* the professor FTE numbers were acquired[2] through internet and telephone sources from the universities themselves.

4 DEA Calculation

The description and calculation of the following data for German universities shall outline and further the discussion on the disciplinary performance measurement topic at the same time. In order to highlight the effect of disciplinary data details, the following cases are calculated in a DEA application (BANXIA Frontier Analyst, Version 4.0). In combination with the four model options in the DEA calculation (CCR with fixed returns to scale, BCC with variable returns to scale, input- and output-oriented models) a total of 16 DEA calculation runs is applied (A1-D4) (Table 1).

A first ***test run*** was performed with all 88 universities showing that there are two *theoretical production anomalies* (Köln DSHS and Hagen FernU): As it can be argued, that these two institutions, the “Deutsche Sporthochschule Köln” (Köln DSHS) as well as the “Fernuniversität Hagen” (Hagen FernU) as distance learning

[2] DFG does not provide professor numbers for the FH/UAS; though the later data from 2012/2013 represents a rupture in data standards (university professor staff with data from 2009, different data source), the special interest in the performance of FH/UAS compared to the universities foregoes the data quality problem in this calculation. Added data is marked with an asterix (*) in the data tables.

Table 1 DEA calculation cases

Case	Inputs	Outputs	Variations (1–4)
A (A1-A4)	Professors total (FTE)	TPF[a] total, graduates total	CCR input, CCR output, BCC input, BCC output
B (B1-B4)	Professors total (FTE)	TPF total, graduates discriminated by levels (Ph.D., Ph.D. share women, MA, BA)	CCR input, CCR output, BCC input, BCC output
C (C1-C4)	Professors total (FTE)	TPF discriminated by *discipline*, graduates total	CCR input, CCR output, BCC input, BCC output
D (D1-D4)	Professors total (FTE)	TPF discriminated by *discipline*, graduates discriminated by levels	CCR input, CCR output, BCC input, BCC output

[a]Third party funding

university, have *significantly different production environments* at least for the teaching area—shown in the distinctive outlying positions of these two institutions in the frontier graph—they are excluded from the further calculations (then with only 86 institutions).

After this correction the following frontier is obtained for the calculated DEA Case A1 with the input professors (total number) and the outputs third party funding (total in Mio. €) and graduates (total) for the CCR constant returns to scale and input minimization mode (Fig. 1); this distribution is more dispersed with even four institutions making up the *production frontier* (Hohenheim U, München TU, Aachen TH and Freiberg TU).

A further comparison of the four different sub-cases A1 to A4 is depicted in Table 2. Interestingly, comparing CCR and BCC cases the calculated efficiency score increase which proves the existence of variable returns to scale (VRS) and the BCC cases shall provide the more realistic efficiency values. Besides the first four there are three more efficient institutions (München LMU, Friedrichshafen ZU and Hannover MedH). Between input and output orientation there are only minor (BCC) or no differences (CCR).

Furthermore, the concept of *super-efficiency calculation* (i.e. Zhu 2001) is used in DEA in order to distinguish the institutions constituting the efficiency frontier (correctly announced in a first step with 100 % efficiency score, exemplified in the first column—A1**); this is done by calculating how much an efficient unit can lose out on output (or gain in input) and still be on the production frontier compared to the other efficient units and adding this to the already 100 % efficiency score. For example, *Hohenheim University* has the highest super-efficiency score as it can lose 10.3 % of all output and still would be 100 % efficient (110.30 %).

Finally, in the A4 case the combination of efficiency values and the number of professors (as an indicator for institutional size) are matched (correlation value $r = 0.11$). As shown in Fig. 2, no clear "economies of scale" picture is recognizable, meaning that high as well as low efficiency values are realized in small (low number of professors) as well as large (high number of professors) institutions; this is a result hypothesis as "*independency of efficiency and size*".

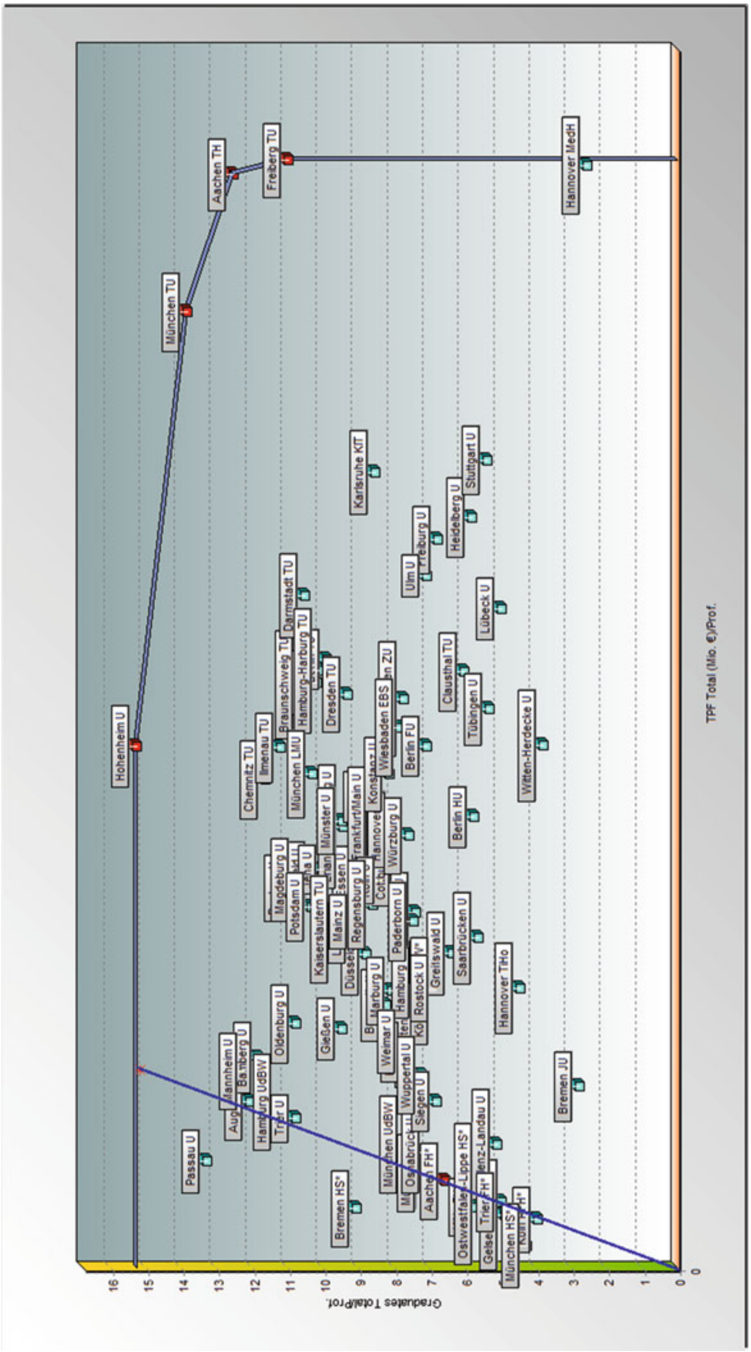


Fig. 1 DEA case A1 (86 institutions)—Frontier plot

Table 2 DEA cases A1 to A4 including super-efficiency values

	A1**	A1	A2	A3	A4
	Input Prof. Total; Output TPF Total, Graduates Total				
University	CCR input (%)	CCR input (%)	CCR output (%)	BCC input (%)	BCC output (%)
Aachen FH*	42.90		42.90	43.70	45.60
Aachen TH	**100.00**	**104.00**[1]	**104.00**	**1,000.00**	**114.10**
Augsburg U	79.20		79.20	80.90	82.50
Bamberg U	77.40		77.40	78.10	78.90
Bayreuth U	53.90		53.90	54.30	56.90
Berlin FU	52.50		52.50	52.80	70.40
Berlin HU	43.70		43.70	44.10	58.00
Berlin TU	70.00		70.00	70.10	71.10
Bielefeld U	68.20		68.20	70.60	72.40
Bochum U	71.80		71.80	78.10	79.00
Bonn U	60.40		60.40	61.90	67.60
Braunschweig TU	74.00		74.00	75.00	74.10
Bremen HS*	59.40		59.40	60.50	61.50
Bremen JU	19.70		19.70	28.50	21.20
Bremen U	71.20		71.20	71.80	71.90
Chemnitz TU	77.90		77.90	78.70	78.70
Clausthal TU	53.90		53.90	60.80	54.70
Cottbus TU	54.30		54.30	58.60	55.40
Darmstadt TU	74.90		74.90	76.00	75.20
Dortmund TU	50.30		50.30	50.70	52.30
Dresden TU	65.90		65.90	66.40	73.50
Duisburg-Essen U	60.30		60.30	63.50	64.90
Düsseldorf U	57.90		57.90	59.80	62.10
Erlangen-Nürnberg U	64.00		64.00	67.30	73.50
Frankfurt/Main U	59.40		59.40	60.80	62.20
Freiberg TU	**100.00**	**101.20**	**101.20**	**109.60**	**110.90**
Freiburg U	65.70		65.70	66.40	66.50
Friedrichshafen ZU	57.20		57.20	**180.00**	**1,000.00**
Gelsenkirchen FH*	32.00		32.00	35.00	33.50
Gießen U	62.00		62.00	66.20	67.70
Greifswald U	43.50		43.50	45.80	43.90
Göttingen U	56.00		56.00	56.90	60.00
Halle-Wittenberg U	47.60		47.60	49.30	51.80
Hamburg U	48.90		48.90	52.70	70.70
Hamburg UdBW	73.70		73.70	78.40	75.50

(continued)

Table 2 (continued)

University	A1**	A1	A2	A3	A4
	Input Prof. Total; Output TPF Total, Graduates Total				
	CCR input (%)	CCR input (%)	CCR output (%)	BCC input (%)	BCC output (%)
Hamburg-Harburg TU	71.10		71.10	77.70	75.50
Hannover MedH	99.40		99.40	**100.10**	**100.10**
Hannover TiHo	31.50		31.50	45.60	36.00
Hannover U	55.70		55.70	56.00	57.30
Heidelberg U	67.70		67.70	68.50	77.40
Hohenheim U	**100.00**	**110.30**	**110.30**	**114.10**	**115.30**
Ilmenau TU	76.10		76.10	80.80	78.90
Jena U	66.50		66.50	70.00	71.30
Kaiserslautern TU	63.50		63.50	63.90	66.00
Karlsruhe KIT	72.30		72.30	72.50	72.60
Kassel U	50.20		50.20	51.70	54.30
Kiel U	46.30		46.30	47.50	49.40
Koblenz-Landau U	33.30		33.30	38.00	34.00
Konstanz U	57.50		57.50	60.90	58.70
Köln FH*	37.50		37.50	38.50	41.10
Köln KatHO NRW*	45.60		45.60	51.50	46.70
Köln RFH*	25.80		25.80	28.80	27.30
Köln U	56.60		56.60	59.10	64.10
Leipzig U	60.40		60.40	64.80	66.10
Lübeck U	59.50		59.50	64.80	59.50
Lüneburg U	47.00		47.00	49.80	48.40
Magdeburg U	70.90		70.90	73.20	75.10
Mainz U	60.50		60.50	64.70	65.90
Mannheim U	80.10		80.10	80.50	81.80
Marburg U	53.00		53.00	55.70	57.70
München HS*	27.90		27.90	28.50	32.10
München LMU	70.30		70.30	**1,000.00**	**115.20**
München TU	**100.00**	**103.60**	**103.60**	**1,000.00**	**112.60**
München UdBW	50.50		50.50	52.30	52.50
Münster FH*	47.20		47.20	47.70	50.80
Münster U	64.40		64.40	67.60	72.70
Oldenburg U	70.50		70.50	70.50	73.20
Osnabrück U	46.30		46.30	47.40	48.90
Ostwestfalen-Lippe HS*	36.60		36.60	39.80	38.00
Paderborn U	50.60		50.60	52.20	51.70
Passau U	87.00		87.00	89.40	88.10

(continued)

Table 2 (continued)

University	A1**	A1	A2	A3	A4
	Input Prof. Total; Output TPF Total, Graduates Total				
	CCR input (%)	CCR input (%)	CCR output (%)	BCC input (%)	BCC output (%)
Potsdam U	67.80		67.80	69.90	72.00
Regensburg U	57.70		57.70	58.60	60.90
Rostock U	45.70		45.70	45.90	48.30
Saarbrücken U	39.50		39.50	41.40	39.70
Siegen U	44.60		44.60	45.00	47.70
Stuttgart U	72.90		72.90	73.60	73.70
Trier FH*	32.60		32.60	36.10	33.90
Trier U	70.60		70.60	70.70	73.20
Tübingen U	50.40		50.40	50.90	51.10
Ulm U	62.60		62.60	62.70	63.00
Weimar U	50.90		50.90	57.50	52.00
Wiesbaden EBS	56.50		56.50	80.00	73.50
Witten-Herdecke U	47.10		47.10	68.00	56.10
Wuppertal U	47.10		47.10	47.20	50.40
Würzburg U	52.90		52.90	53.30	54.20

[1]Super-efficiency values in bold.

As the research question addressed a possible change by adding a disciplinary differentiation, especially the *cases C2-C4* are reported in comparison with the A4 case and the B2-B4 cases: For the *B cases* with the differentiation in graduate levels (PhD, MA, BA) the further added indicator of *PhD by women* as *share* of the total PhD has been selected as this has shown the highest discriminatory value of the three options. This is consistent with the use of this special indicator also, for example, in indicator-based budgeting and funding schemes (e.g., MIWF NRW 2014). There are specific universities gaining in the efficiency calculation for the DEA *C cases*, namely Augsburg U, Bamberg U, Berlin FU, Bielefeld U, Bochum U, Dresden TU, Hamburg UdBW, Hannover TiHo, Heidelberg U, Kaiserslautern TU, Leipzig U, Mainz U, Passau U, Witten-Herdecke U (Table 3).

The following Fig. 3 is outlining the aforementioned connection of efficiency scores and institutional size represented by the total number of professors for the DEA cases B4 (same result for C4).

Again, it is obvious that there is only a very tiny connection ($r = 0{,}23$) identically for both cases—this highlights again that there are *no significant economies of scale* present in university production processes as smaller as well as larger institutions both come up with high as well as low efficiency values.

The further DEA cases D1-D4 (combining the output differentiation for TPF in different disciplines and for the graduates on different levels) have been calculated and can be seen in the data annex but do not provide further insights as the numbers

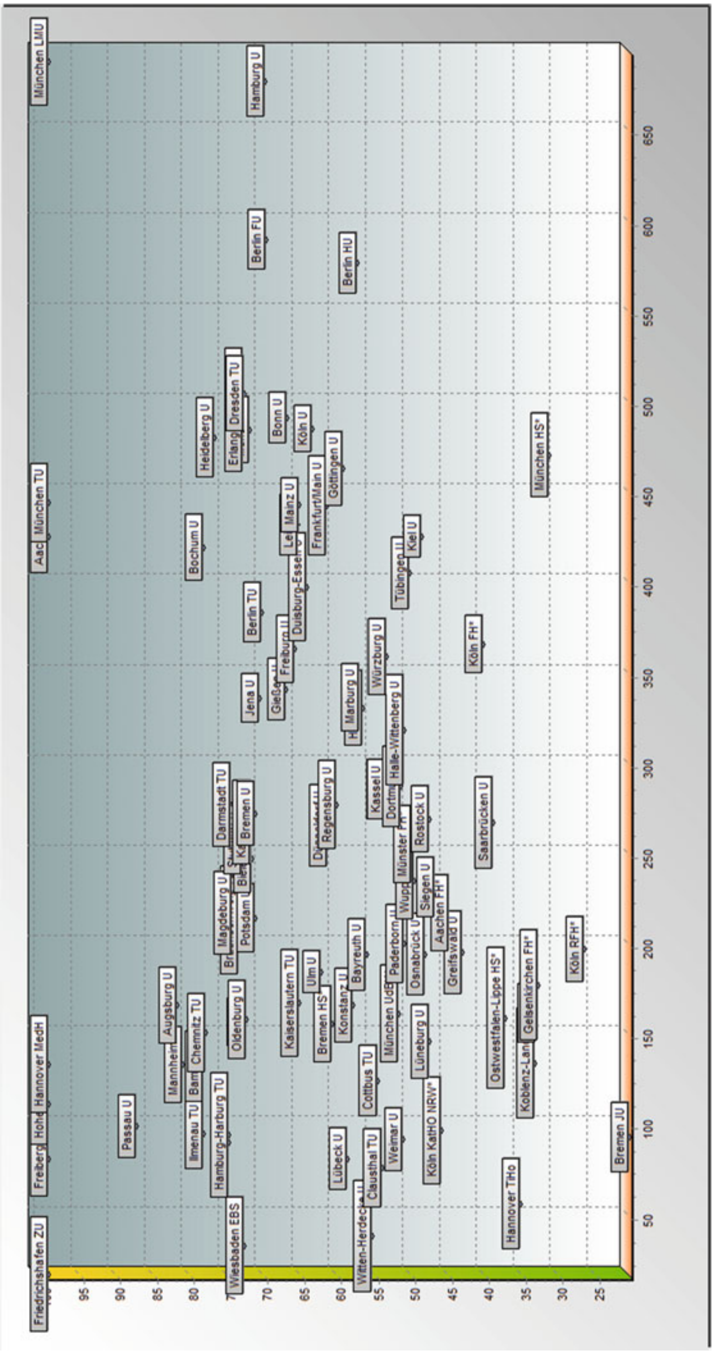


Fig. 2 Efficiency (y-axis) and number of professors (x-axis), A4

Table 3 DEA cases A4 and C1–C4 including super-efficiency values

University	A4 BCC output (%)	B2 CCR output (%)	B3 BCC input (%)	B4 BCC output (%)	C2 CCR output (%)	C3 BCC input (%)	C4 BCC output (%)
Aachen FH*	45.60	56.70	61.20	73.80	56.70	61.20	73.80
Aachen TH	**114.10**[1]	**108.20**	**1,000.00**	**114.10**	**108.20**	**1,000.00**	**114.10**
Augsburg U	82.50	**104.30**	**104.40**	**104.50**	**104.30**	**104.40**	**104.50**
Bamberg U	78.90	**103.30**	**105.30**	**104.60**	**103.30**	**105.30**	**104.60**
Bayreuth U	56.90	60.70	61.00	70.70	60.70	61.00	70.70
Berlin FU	70.40	58.80	**1,000.00**	**109.50**	58.80	**1,000.00**	**109.50**
Berlin HU	58.00	50.70	78.10	93.30	50.70	78.10	93.30
Berlin TU	71.10	78.70	78.70	80.50	78.70	78.70	80.50
Bielefeld U	72.40	75.70	**130.30**	**109.80**	75.70	**130.30**	**109.80**
Bochum U	79.00	76.00	**115.40**	**103.20**	76.00	**115.40**	**103.20**
Bonn U	67.60	84.90	89.40	92.30	84.90	89.40	92.30
Braunschweig TU	74.10	82.80	83.00	83.00	82.80	83.00	83.00
Bremen HS*	61.50	79.40	86.60	91.90	79.40	86.60	91.90
Bremen JU	21.20	24.50	31.50	25.50	24.50	31.50	25.50
Bremen U	71.90	77.60	79.60	87.60	77.60	79.60	87.60
Chemnitz TU	78.70	89.00	89.60	89.30	89.00	89.60	89.30
Clausthal TU	54.70	64.80	69.60	65.70	64.80	69.60	65.70
Cottbus TU	55.40	63.20	67.00	63.40	63.20	67.00	63.40
Darmstadt TU	75.20	79.10	79.10	83.70	79.10	79.10	83.70
Dortmund TU	52.30	58.70	59.00	63.80	58.70	59.00	63.80
Dresden TU	73.50	97.60	**104.10**	**103.30**	97.60	**104.10**	**103.30**
Duisburg-Essen U	64.90	66.60	69.60	84.30	66.60	69.60	84.30
Düsseldorf U	62.10	75.50	77.50	86.80	75.50	77.50	86.80
Erlangen-Nürnberg U	73.50	74.50	78.90	85.30	74.50	78.90	85.30
Frankfurt/Main U	62.20	79.70	86.10	92.50	79.70	86.10	92.50
Freiberg TU	**110.90**	**123.80**	**132.10**	**140.30**	**123.80**	**132.10**	**140.30**
Freiburg U	66.50	69.20	70.20	89.00	69.20	70.20	89.00
Friedrichshafen ZU	**1,000.00**	74.90	**180.00**	**1,000.00**	74.90	**180.00**	**1,000.00**
Gelsenkirchen FH*	33.50	44.70	46.10	54.30	44.70	46.10	54.30
Gießen U	67.70	75.50	91.90	97.80	75.50	91.90	97.80
Greifswald U	43.90	68.70	72.00	79.50	68.70	72.00	79.50
Göttingen U	60.00	58.50	73.70	93.30	58.50	73.70	93.30
Halle-Witten-berg U	51.80	61.10	62.60	83.10	61.10	62.60	83.10
Hamburg U	70.70	62.10	81.10	97.20	62.10	81.10	97.20
Hamburg UdBW	75.50	**144.10**	**179.00**	**148.30**	**144.10**	**179.00**	**148.30**

(continued)

Table 3 (continued)

University	A4 BCC output (%)	B2 CCR output (%)	B3 BCC input (%)	B4 BCC output (%)	C2 CCR output (%)	C3 BCC input (%)	C4 BCC output (%)
Hamburg-Harburg TU	75.50	82.50	89.40	87.40	82.50	89.40	87.40
Hannover MedH	**100.10**	**100.30**	**121.70**	**116.30**	**100.30**	**121.70**	**116.30**
Hannover TiHo	36.00	**131.20**	**1,000.00**	**144.20**	**131.20**	**1,000.00**	**144.20**
Hannover U	57.30	61.10	61.10	70.40	61.10	61.10	70.40
Heidelberg U	77.40	68.10	**105.90**	**101.60**	68.10	**105.90**	**101.60**
Hohenheim U	**115.30**	**132.10**	**136.40**	**133.20**	**132.10**	**136.40**	**133.20**
Ilmenau TU	78.90	78.80	85.10	82.80	78.80	85.10	82.80
Jena U	71.30	93.00	98.00	98.50	93.00	98.00	98.50
Kaiserslautern TU	66.00	**100.70**	**103.00**	**103.30**	**100.70**	**103.00**	**103.30**
Karlsruhe KIT	72.60	85.60	85.70	86.60	85.60	85.70	86.60
Kassel U	54.30	64.00	64.40	72.80	64.00	64.40	72.80
Kiel U	49.40	58.50	59.20	77.00	58.50	59.20	77.00
Koblenz-Landau U	34.00	68.50	69.20	75.20	68.50	69.20	75.20
Konstanz U	58.70	62.00	62.90	66.70	62.00	62.90	66.70
Köln FH*	41.10	47.60	62.90	71.20	47.60	62.90	71.20
Köln KatHO NRW*	46.70	60.30	64.30	61.00	60.30	64.30	61.00
Köln RFH*	27.30	29.20	31.30	36.90	29.20	31.30	36.90
Köln U	64.10	82.10	85.90	89.10	82.10	85.90	89.10
Leipzig U	66.10	77.60	**112.20**	**103.30**	77.60	**112.20**	**103.30**
Lübeck U	59.50	86.20	93.00	94.40	86.20	93.00	94.40
Lüneburg U	48.40	63.90	79.40	96.10	63.90	79.40	96.10
Magdeburg U	75.10	80.70	81.10	80.90	80.70	81.10	80.90
Mainz U	65.90	98.00	**109.90**	**107.80**	98.00	**109.90**	**107.80**
Mannheim U	81.80	90.70	97.00	98.30	90.70	97.00	98.30
Marburg U	57.70	68.10	69.10	88.00	68.10	69.10	88.00
München HS*	32.10	34.40	43.90	56.00	34.40	43.90	56.00
München LMU	**115.20**	89.50	**1,000.00**	**148.50**	89.50	**1,000.00**	**148.50**
München TU	**112.60**	**111.00**	**1,000.00**	**126.10**	**111.00**	**1,000.00**	**126.10**
München UdBW	52.50	65.10	66.20	77.00	65.10	66.20	77.00
Münster FH*	50.80	61.10	76.80	83.50	61.10	76.80	83.50
Münster U	72.70	73.40	98.60	99.60	73.40	98.60	99.60
Oldenburg U	73.20	79.80	91.50	96.50	79.80	91.50	96.50
Osnabrück U	48.90	55.70	56.70	82.80	55.70	56.70	82.80
Ostwestfalen-Lippe HS*	38.00	51.10	52.50	59.50	51.10	52.50	59.50
Paderborn U	51.70	54.80	55.10	63.30	54.80	55.10	63.30
Passau U	88.10	**106.80**	**110.80**	**106.80**	**106.80**	**110.80**	**106.80**
Potsdam U	72.00	83.10	85.60	92.50	83.10	85.60	92.50
Regensburg U	60.90	66.80	72.40	93.10	66.80	72.40	93.10
Rostock U	48.30	52.20	52.30	71.30	52.20	52.30	71.30

(continued)

Table 3 (continued)

University	A4 BCC output (%)	B2 CCR output (%)	B3 BCC input (%)	B4 BCC output (%)	C2 CCR output (%)	C3 BCC input (%)	C4 BCC output (%)
Saarbrücken U	39.70	51.10	51.20	82.80	51.10	51.20	82.80
Siegen U	47.70	50.10	50.40	63.00	50.10	50.40	63.00
Stuttgart U	73.70	74.60	75.00	75.50	74.60	75.00	75.50
Trier FH*	33.90	43.60	45.50	51.20	43.60	45.50	51.20
Trier U	73.20	**102.90**	**112.50**	**110.90**	**102.90**	**112.50**	**110.90**
Tübingen U	51.10	51.50	56.10	86.30	51.50	56.10	86.30
Ulm U	63.00	70.20	71.40	75.80	70.20	71.40	75.80
Weimar U	52.00	69.50	73.40	70.60	69.50	73.40	70.60
Wiesbaden EBS	73.50	69.80	88.50	85.30	69.80	88.50	85.30
Witten-Herdecke U	56.10	**115.60**	**122.50**	**130.40**	**115.60**	**122.50**	**130.40**
Wuppertal U	50.40	54.00	56.30	78.30	54.00	56.30	78.30
Würzburg U	54.20	80.40	85.50	88.40	80.40	85.50	88.40

[1]Super-efficiency values in bold.

are identical to the DEA Cases B and C (featuring the maximum values for each unit thereof); this is consistent with DEA methodology as combinatorial patterns usually do not produce additional outer production function boundaries.

5 Discussion

The aforementioned DEA results can be used for a discursive approach and the identification of further research fields and questions. The main question of representing a discipline differentiation in performance and efficiency measurement was shown to be valuable for third party funding indicators as many universities gained in their efficiency scores compared to a calculation case with only an accumulated total TPF indicator. In this context especially *mid-size universities* (e.g., Universities of Mainz, Leipzig, Passau) as well as smaller and *specialized universities* (e.g., Hannover TiHo) and the *universities of applied sciences* (e.g., Köln FH, Münster FH, Bremen HS) profited from these disciplinary details as especially universities of applied sciences have usually a focused disciplinary approach (e.g., none of them attains faculties in the field of life sciences including medicine). The same result can be shown for a differentiation of graduates by degree levels including the question of the share of PhD by women: With the *introduction of differentiation* in the output area mainly mid-size and smaller universities as well as the universities of applied sciences "gain ground" on the very large and also on the technical universities (as they are recognized to have advantages in acquiring TPF, for example, due to their engineering profile). This should be detailed in further research, combined with influencing factors like

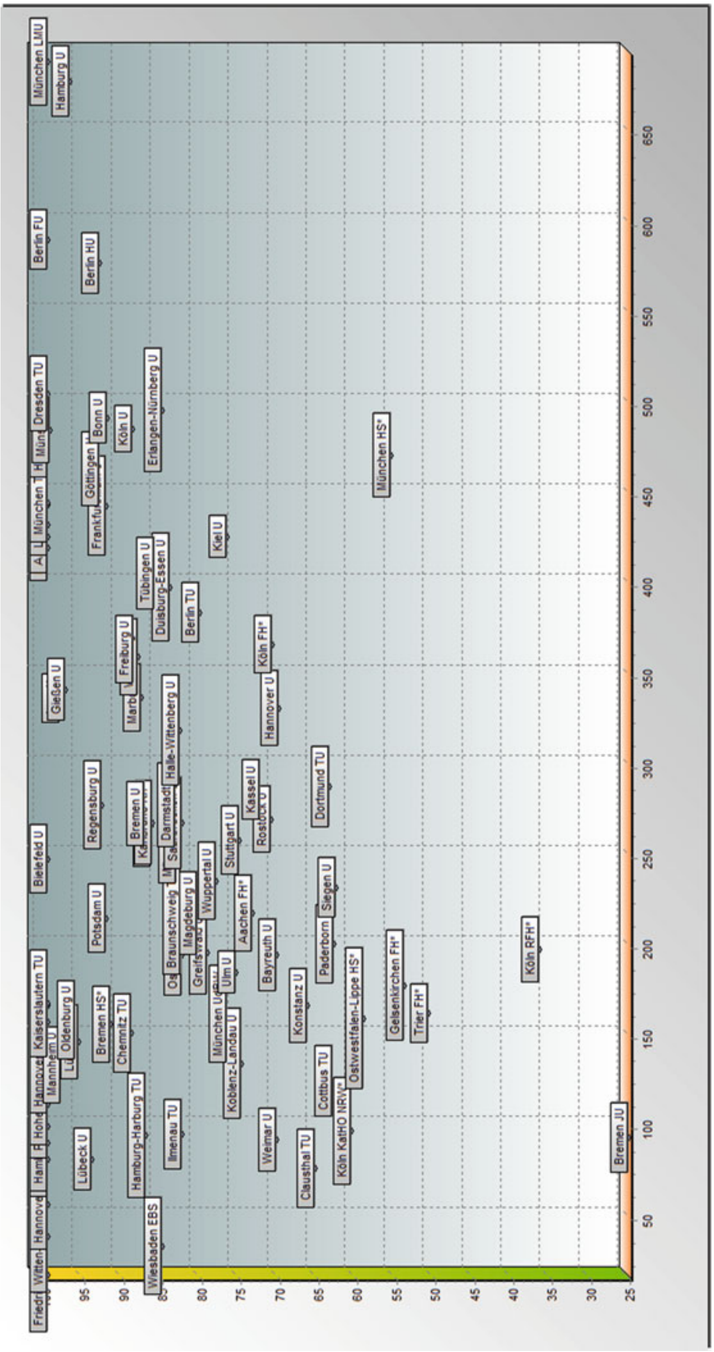


Fig. 3 Efficiency (y-axis) and number of professors (x-axis), B4

university type, disciplinary focus etc. as outlined above—though in this DEA calculation no specific correlations were discernible which is in line with several calculation results also in an international comparison: Usually no specific indicators defining efficiency can be recognized but it becomes obvious that efficiency is an independent perspective. Therefore especially university performance evaluations, e.g., within rankings as well as within indicator-based funding schemes are urged to apply multi-output measurement methods as, for example, DEA—as otherwise the higher education concept of "profiling" and specialization as well as smaller institutions are "punished" by lower financial budgets.

Altogether the research presented has shown that the detailed insight into disciplinary differences is *crucial* to a fair and productive performance measurement and management approach for universities. Otherwise especially specialized institutions without the full disciplinary range are at a discernible disadvantage. Therefore further research has to establish is this is true for all disciplines, if this result for Germany holds also true for other countries and if there are other patterns and indicators to explain the results (e.g., the influence of the 16 different higher education systems of the German states).

References

Abbott M, Doucouliagos C (2003) The efficiency of Australian universities: a data envelopment analysis. Econ Educ Rev 22(1):89–97

Agasisti T, Johnes G (2009) Beyond frontiers: comparing the efficiency of higher education decision-making units across more than one country. Educ Econ 17:59–79

Ahn T, Arnold V, Charnes A, Cooper WW (1989) DEA and ratio efficiency analyses for public institutions of higher learning in Texas. Res Govern Nonprofit Account 5:165–185

ARWU (2014) About the ARWU ranking (Shanghai). Download at http://www.shanghairanking.com/aboutarwu.html. Accessed 31 Mar 2014

Beasley JE (1995) Determining teaching and research efficiencies. J Oper Res Soc 46:441–452

Brennan J, Lyon ES, Schomburg H, Teichler U (1994) The experiences and views of graduates: messages from recent surveys. High Educ Manag 6(3):275–304

Butler TW, Li L (2005) The utility of returns to scale in DEA programming: an analysis of Michigan rural hospitals. Eur J Operati Res 161(2):469–477

Charnes A, Cooper W, Rhodes E (1978) Measuring the efficiency of decision making units. Eur J Oper Res 2:429–444

Charnes A, Cooper WW, Thrall RM (1991) A structure for classifying and characterizing efficiency and inefficiency in data envelopment analysis. J Prod Anal 2(3):197–237

Coelli T (1995) Estimators and hypothesis tests for s stochastic Frontier function: a Monte Carlo analysis. J Prod Anal 6:247–268

Cooper WW, Seiford LM, Tone K (2000) Data envelopment analysis: a comprehensive text with models, applications, references and DEA-solver software. Kluwer, Boston

DFG (2013) Funding Atlas 2012 – key indicators for publicly funded research in Germany. Wiley, Weinheim

Dundar H, Lewis DR (1995) Departmental productivity in American universities: economies of scale and scope. Econ Educ Rev 14:199–244

Glass JC, McKillop DG, O'Rourke G (1998) A cost indirect evaluation of productivity change in UK universities. J Prod Anal 10:153–175

Ministerium für Innovation, Wissenschaft und Forschung des Landes Nordrhein-Westfalen (MIWF NRW) (2014) Ziel- und Leistungsvereinbarung V (ZLV 2014–2015) zwischen der RWTH Aachen und dem Ministerium für Innovation, Wissenschaft und Forschung des Landes Nordrhein-Westfalen, Düsseldorf
Jacobs R (2001) Alternative methods to examine hospital efficiency: data envelopment analysis and stochastic Frontier analysis. Health Care Manag Sci 4(2):103–115
Johnes G, Johnes J (1992) Apples and oranges: the aggregation problem in publications analysis. Scientometrics 25(2):353–365
Jongbloed B, Vossensteyn H (2001) Keeping up performances: an international survey of performance-based funding in higher education. J High Edu Policy Manag 23(2):127–145
Kao C, Hung HT (2008) Efficiency analysis of university departments: an empirical study. Omega 36(4):653–664
Kocher GM, Luptacik M, Sutter M (2006) Measuring productivity of research in economics: a cross-country study using DEA. Socio Econ Plan Sci 40:314–332
Korhonen P, Tainio R, Wallenius J (2001) Value efficiency analysis of academic research. Eur J Oper Res 130:121–132
Kosmützky A, Krücken G (2014) Growth or steady state? A bibliometric focus on international comparative higher education research. High Educ 67:457–472
Kumbhakar SC, Lovell CAK (2000) Stochastic Frontier analysis. Cambridge University Press, Cambridge
Lewin AY, Morey RC (1981) Measuring the relative efficiency and output potential of public sec-tor organizations – an application of data envelopment analysis. Int J Policy Anal Inf Syst 5 (4):267–285
Lindsay AW (1982) Institutional performance in higher education: the efficiency dimension. Rev Educ Res 52(2):175–199
Liu NC, Cheng Y (2005) The academic ranking of world universities. High Educ Eur 30:127–136
Madden G, Savage S, Kemp S (1997) Measuring public sector efficiency: a study of economic departments at Australian universities. Educ Econ 5(2):153–168
McMillan ML, Chan WH (2006) University efficiency: a comparison and consolidation of results from stochastic and non-stochastic methods. Educ Econ 14(1):1–30
McMillan ML, Datta D (1998) The relative efficiencies of Canadian universities: a DEA perspective. Can Pub Policy 24(4):485–511
Ng YC, Li SK (2000) Measuring the research performance of Chinese higher education institutions: an application of data envelopment analysis. Educ Econ 8:139–156
Ramsden P (1994) Describing and explaining research productivity. High Educ 28(2):207–226
Reale E (2014) Challenges in higher education research: the use of quantitative tools in comparative analyses. High Educ 67:409–422
Sarrico CS (2010) On performance in higher education – towards performance government. Tert Educ Manag 16(2):145–158
Sarrico CS, Dyson RG (2004) Restricting virtual weights in data envelopment analysis. Eur J Oper Res 159(1):17–34
Sarrico CS, Hogan SM, Dyson RG, Athanassopoulos AD (1997) Data envelopment analysis and university selection. J Oper Res Soc 48(12):1163–1177
Sarrico CS, Teixeira P, Rosa MJ, Cardoso MF (2009) Subject mix and productivity in Portugese universities. Eur J Oper Res 197(2):287–295
Seiford LM (1996) Data envelopment analysis: the evolution of the state of the art (1978–1995). J Product Anal 7:99–137
Shin JC, Totkoushian RK (2012) The past, present, and future of university rankings. In: Shin JC, Totkoushian RK, Teichler U (eds) University rankings – theoretical basis, methodology and impact on global higher education. Springer, Dordrecht, pp 1–18
Stahl MJ, Leap SL, Wei ZZ (1998) Publication in leading management journals as a measure of institutional research productivity. Acad Manag J 31(3):707–720

Statistisches Bundesamt (2014) Statistisches Bundesamt, Fachserie 11, Reihe 4.2, Prüfungsjahr 2012, Wiesbaden

Stevens PA (2005) Stochastic Frontier analysis of English and Welsh Universities. Educ Econ 13 (4):355–374

Taylor B, Harris G (2004) Relative efficiency among South African universities: a data envelopment analysis. High Educ 47(1):73–89

Thannasoulis E, Portela MCS, Despic O (2008) Data envelopment analysis: the mathematical programming approach to efficiency analysis. In: Fried HO, Knox Lovell CA, Schmidt SS (eds) The measurement of productive efficiency and productivity growth. Oxford University Press, Oxford, pp 251–419

Van Vaught FA, Ziegele F (2012) Multidimensional rankings – the design and development of U-multirank. Springer, Dordrecht

Worthington AC, Higgs H (2011) Economies of scale and scope in Australian higher education. High Educ 61:387–414

Worthington AC, Lee BL (2008) Efficiency, technology and productivity change in Australian universities, 1998–2003. Econ Educ Rev 27:285–298

Zangoueinezhad A, Moshabaki A (2011) Measuring university performance using a knowledge-based balanced scorecard. J Product Perform Manag 60(8):824–843

Zhu J (2001) Super-efficiency and DEA sensitivity analysis. Eur J Oper Res 129(2):443–455

Zhu J, Cook WD (eds) (2007) Modeling data irregularities and structural complexities in data envelopment analysis – a problem-solving handbook. Springer, New York

Consideration of Knowledge and Technology Transfer Characteristics for Research Evaluation

Harald F.O. von Kortzfleisch, Matthias Bertram, Dorothée Zerwas, and Manfred Arndt

Abstract Knowledge and Technology Transfer (KTT) is currently becoming the third mission for the scientific community in addition to research and education. Therefore, there is a growing need to evaluate the impact of KTT, both directly and indirectly, on industry and society. However, despite the growing importance of KTT and the considerable amount of research that has already been conducted in this field, existing approaches to research evaluation primarily focus on quantitative determinants (e.g., number of publications, patents and licenses, number of collaboration projects with industry, or of companies founded) thereby neglecting transfer-oriented aspects of research evaluation. Therefore, in this article we investigate the characteristics of KTT, and to what extent they are taken into account by existing research evaluation approaches. Our results confirm that, up until now, KTT has been infrequently considered as an approach toward the evaluation of current research. Existing evaluation approaches focus on quantitative determinants, but to some extent they fail to realize that those determinants are not equally appropriate for evaluating KTT in different scientific disciplines or traditions. Based on our results, we call for more integrative and systematic research, building a foundation to meet the requirements of the growing importance of KTT in research evaluation.

1 Introduction

The importance of KTT for economic development cannot be overstated. The term KTT generally refers to the process of transforming intellectual property into products, and transferring research results from research organizations to industrial partners. "Both the acquisition of technology and its diffusion foster productivity growth" (Hoekman et al. 2005, p. 1587). For this reason, the European Union has

H.F.O. von Kortzfleisch (✉) • M. Bertram • D. Zerwas • M. Arndt
University of Koblenz-Landau, Koblenz, Germany
e-mail: harald.von.kortzfleisch@uni-koblenz.de

I.M. Welpe et al. (eds.), *Incentives and Performance*,
DOI 10.1007/978-3-319-09785-5_27

made KTT one of the central pillars of its new Horizon 2020 funding program (Horizon 2014). The aim of this program is not merely to promote excellent scientific results, but also to make an impact on industry and society. In fact, Horizon 2020 explicitly encourages the transfer of knowledge and technological advances along the innovation chain, from scientific excellence to industrial leadership and sociological challenges. However, as many examples show a priori KTT evaluation is difficult.

One example for economically successful technology transfer is the MPEG Audio Layer III encoding, more commonly known as MP3 compression format.[1] The MP3 format was designed in the early 1990s by the Movie Picture Experts Group (MPEG) including, the German Fraunhofer-Gesellschaft, as an integration of the MPEG-1 standard (and later extended in the MPEG-2 standard). However, in the first years after its development MP3 did not gain much attention. The need for personal and portable audio at that time was met by cassette players or CD players. It took several innovations in other areas to create a growing interest in the MP3 format, such as the increasing access to personal computers and an improvement in storage capacity as well as the growth of the internet and the upcoming of file-sharing software in the late 1990s. Today, MP3 is integrated in many music players, mobile phones or tablets and companies like Apple, Google, Amazon and many other successfully use the MP3 format to sell music or audio books. MP3 is a huge commercial success and generates licensing revenues that can be used for founding new research projects. The MP3 format was accepted by the market. This corresponds to an indirect positive evaluation of the MP3 format by the market.

An example for an unsuccessful technology transfer is the Transrapid project.[2] The Transrapid is a German high speed monorail train using magnetic levitation and a cruising speed of currently 500 km/h. Planning for the Transrapid started in the late 1960s and its technological readiness was in 1991. From a technological point of view the Transrapid uses a superior technology which allows higher speed and even lower energy consumption as classical railway solutions. With its maximum speed of 550 km/h it is placed between classical high speed trains and air traffic. Despite these advantages the Transrapid—at least so far—was not accepted by a broad market. The competition with high speed trains using the ‘traditional’ railway system as well as cheaper air traffic tickets are seen as reasons for the unsuccessful implementation of the Transrapid. In fact, the Transrapid technology has only been commercially implemented in China so far. In Germany, the initial test track in Emsland is the only remaining Transrapid track. Thus, although providing a better technological foundation with better properties in speed, costs and ecological impact the Transrapid can be regarded as an example for unsuccessful technology transfer. In contrast to the MP3 format, the Transrapid was not positively received and therefore indirectly evaluated negatively by the market.

[1] See http://www.iis.fraunhofer.de/de/bf/amm/produkte/audiocodec/audiocodecs/mp3.html

[2] See www.transrapid.de and http://www.hochgeschwindigkeitszuege.com/deutschland/transrapid.php

Nevertheless from the point of view of engineers it is considered as a great innovation. What can be learned from the examples is that an evaluation by the market is not always sufficient. Rather, an evaluation by peers appears to be conducive to assess complex technology innovations like the Transrapid. Especially with regard to complex technology innovations, a positive evaluation by peers cannot be put on an equal footing with market success. Although the Transrapid is a great technological innovation from a scientific perspective, the technology has not penetrated the market due to a lack of demand and due to alternative technologies.

Despite the complexity of KTT, Horizon 2020 and many other research programs show that KTT is steadily becoming what is called the third mission for communities within scientific organizations. The importance of analyzing KTT in a research context cannot be overstated, too. Thus, with KTT becoming the third mission for communities within scientific organizations, there is a growing need not only to evaluate research results, but also to evaluate their direct and indirect impact on industry and society. According to Kuhlmann and Heinze (2003), internal and external research evaluation can be depicted in a three-layer model. The first layer, known as 'the core', describes the evaluation of individual research efforts; the second layer documents an evaluation of research and innovation policy programs; and the third layer includes an evaluation of the performance of research institutions. Within each of these layers, scholars and practitioners alike have developed a considerable number of taxonomies, frameworks and methods. For instance, research evaluation has been investigated within the context of responsive program evaluation (Stake 1983), effect-oriented evaluation in poverty and wealth reports (Beywl et al. 2004), and value, or purpose and method-oriented, research evaluation (Altschuld and Zheng 1995; Patton 2000; Scriven 2013). Despite the obvious importance of KTT, to the best of our knowledge, none of the identified approaches for research evaluation has put a specific, explicit spotlight on the evaluation of KTT, but rather existing approaches for research evaluation only focus on implicit aspects of KTT.

As emphasized above, KTT is an important aspect of modern research communities. A considerable amount of research has already been conducted in the field of research evaluation. However, little attention has been paid to the importance of KTT in the context of research evaluation. To fill this gap, this paper addresses the following research question: *To what extent do existing methods of research evaluation take into account specific KTT characteristics?* To answer this question, we conducted an extensive literature review in the interdisciplinary area of research evaluation.

The remainder of the paper is organized as follows: The second section will briefly introduce the topic of KTT. The third section describes our research approach, derives characteristics of KTT from the literature and provides an overview of existing taxonomies and concrete approaches of research evaluation. Section 4 discusses how the characteristics of KTT feature in existing evaluation methods, and to what extent transfer-oriented characteristics have been considered up until now. The paper closes with a conclusion and discusses the future practical and theoretical implications of KTT.

2 Knowledge and Technology Transfer

Broadly speaking, the term KTT means the dissemination of knowledge and technology between partners. More specifically, exchanging successful methods and research results is not a new phenomenon for research institutions, especially in industry. However, the goals of KTT have changed over time. Whereas in the past, KTT mostly concerned the transfer of knowledge and technologies from industries of developed countries to developing countries, today it refers to both intra-organizational and inter-organizational transfer in the sense of collaboration and cooperation in knowledge and technology-intensive fields (Meissner and Sultanian 2007). In this respect, more recently the term KTT has been associated with innovation.

In academic circles, defining the transfer of technology and knowledge has been, and still is, difficult. From a scholarly perspective KTT has been defined as “the movement of know-how, technical knowledge, or technology from one organizational setting to another” (Roessner 2000). However, conceptualizing KTT depends very much on the situational context in which it takes place. Up until now, scholars have conceptualized it in the following areas: industry-science-relationships (Polt et al. 2001a, b), technology transfer and public policy (Bozeman 2000), international comparison of transfer activities (Meissner and Sultanian 2007), transfer practices in universities (LeBris et al. 2010), and transfer strategies in research programs (Schmoch 2000). Ambiguous definitions complicate discussions regarding KTT, especially in the area of research evaluation. In this area exact definitions are crucial because they guide the evaluator’s perception.

The following definitions exemplarily show that technology transfer and knowledge transfer are often closely related. On the one hand, Meissner and Sultanian (2007) define technology transfer as the “targeted transfer of technological or technology-oriented know-how between transfer partners” (p. 21), such as individuals, organizations or firms. On the other hand, Argote et al. (2000) define knowledge transfer as “the process through which one unit (for example, group, department, or division) is affected by the experience of another” (p. 151). Transactive memory refers to the knowledge about ‘who knows what’ and thus to “the combination of individual minds and the communication among them” (Wegner et al. 1985), is very important for this effect. According to Bozeman (2000), separate definitions have led to confusion in scholarly discussions for a long time. Combining the two definitions, however, seems to solve this confusion: “focusing on the product is not sufficient to the study of transfer and diffusion of technology; it is not merely the product that is transferred but also knowledge of its use and application” (Bozeman 2000, p. 151). However, a more detailed consideration of KTT is needed in order to really understand what KTT is all about.

Practical and scholarly contributions have contributed to a considerable body of knowledge about KTT. Authors have identified varieties of motives, transfer channels, barriers, influencing factors, qualitative and quantitative dimensions and determinants, and characteristics of transfer activities. For instance, with respect to technology transfer in universities, Beise and Spielkamp (1996) have identified the following motives for transfer activities on the sides of the agent and the recipient: economic interests, qualification of human capital, acquisition of know-how, and scientific dialogue. LeBris et al. (2010) listed three distinct forms of transfer interaction, serving as channels for KTT of universities: information transfer, personal transfer, and goods (or product) transfer. In an additional study, Markowski et al. (2008) discovered that, whereas about 90 % of large enterprises undertake cooperative research and development activities, only about 36 % of small and medium enterprises (SME) have these activities. According to their research, the main factor hindering KTT is information asymmetry between the scientific and industrial sectors, closely followed by the scarcity of resources for implementing and maintaining transfer structures in universities. Additional barriers are: cultural differences between the perspectives of the scientific and industrial communities, and amount of administration required during transfer interactions, especially with respect to rigid organizational structures on the side of the universities. Analyzing the structures of transfer partners, Czarnitzki et al. (2000) identified three major organizational factors influencing KTT: nature of the research, whether it be basic research or applied research; degree of financial support for transfer activities; and qualifications of employees.

Regarding the effectiveness of KTT, Bozeman (2000, p. 155) has summarized dimensions and determinants for KTT, and integrated them into an effectiveness model of technology transfer (see Fig. 1). According to this model, the effectiveness of transfer activities depends on the characteristics of the transfer agent and the recipient, including research organizations and enterprises; the transfer object, for instance technology or knowledge; the transfer medium, for instance patents, licenses or publications; and the demand environment, such as supply and demand and the economical character of the transfer object.

Table 1 summarizes the dimensions of Bozeman (2000) Contingent Effectiveness Model.

Effectiveness embraces several different criteria, such as: (1) opportunity cost, which focus not only on alternative uses of resources but also possible impacts on other missions of the transfer agent; (2) scientific and technical human capital, which considers the impact of KTT on existing scientific and technological skills; (3) political reward, which refers to, e.g., increased funding from participation in KTT; (4) market impact and (5) economic development, which refer to the effect of KTT on single firms as well as regional or national markets; and finally, (6) 'out-the-door' transfer results, which refer to the effect of KTT provided by other organizations or firms (Bozeman 2000).

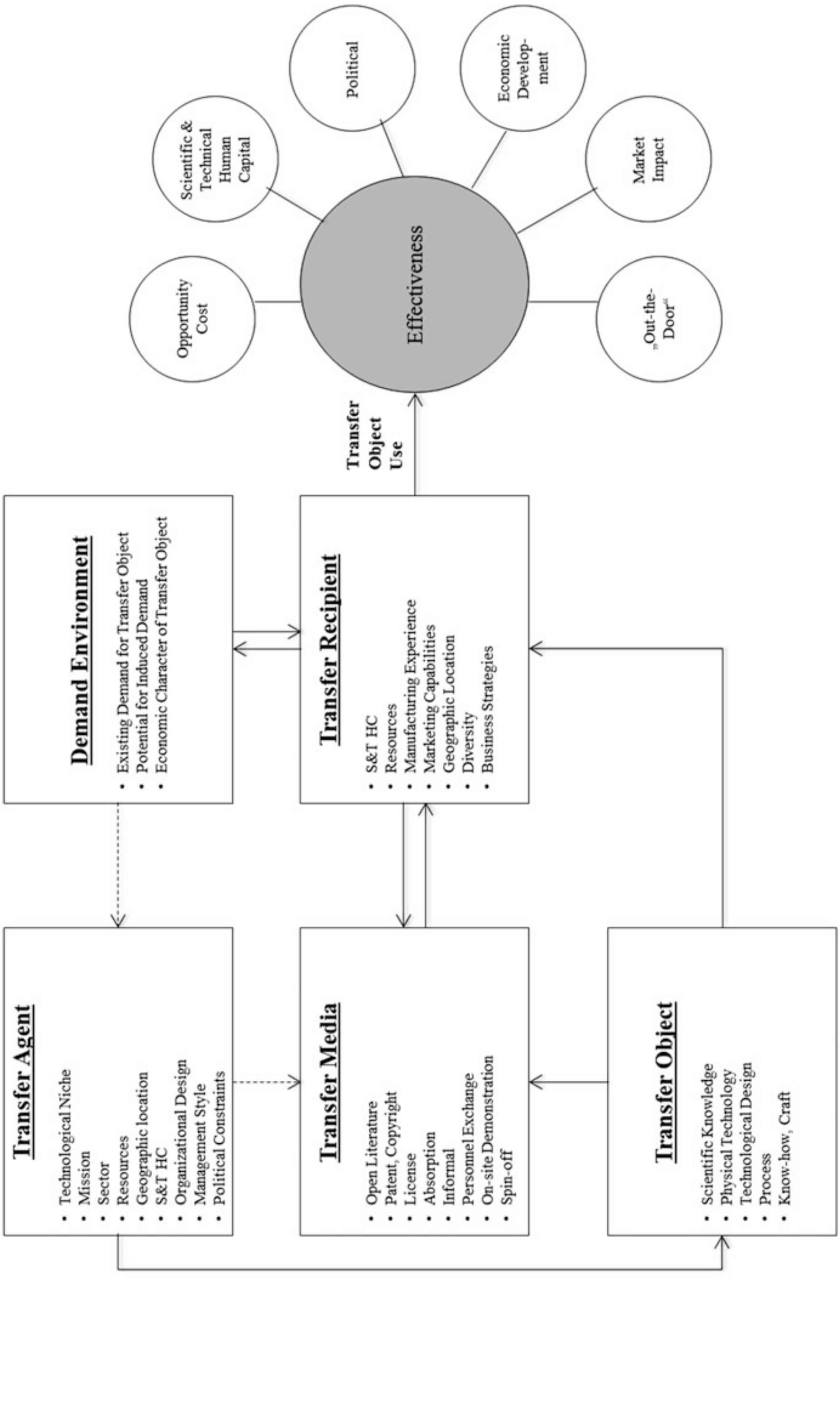


Fig. 1 Contingent effectiveness model of technology transfer according to Bozeman (2000)

Table 1 List of KTT characteristics according to Bozeman (2000)

Dimension	Focus	Example
Transfer agent	The institution or organization seeking to transfer the technology	Government agency, university, private firm, characteristics of the setting, its culture, organization, personnel
Transfer medium	The vehicle, formal or informal by which the technology is transferred	License, copyright, person-to-person, formal literature
Transfer object	The content and form of what is transferred, the transfer entity	Scientific knowledge, technological device, process, know-how, and specific characteristics of each
Transfer recipient	The organization or institution receiving the transfer object	Firm, agency, organization, consumer, informal group, institution and associated characteristics
Demand environment	Factors market and non-market pertaining to the need for the transferred object	Price for technology, substitutability, relation to technologies now used, subsidy, market shelters

3 Analysis of Knowledge and Technology Transfer Characteristics for Research Evaluation Approaches

To draw a coherent picture of the characteristics of KTT and their significance in existing approaches to research evaluation, our study followed a three-step research design: (1) identifying and summarizing the characteristics of KTT (Sect. 3.1); (2) selecting relevant evaluation approaches from scholarly and practical literature (Sect. 3.2); and (3) interpreting those evaluation approaches according to our set of characteristics (Sect. 4).

3.1 Characteristics of Knowledge and Technology Transfer

To identify the characteristics of KTT, we examined existing definitions, models, and frameworks in this field, and accomplished a list of five categories as a framework for our analysis. Our brief overview of KTT in Sect. 2 illustrates the broadness of this topic. Discussions among a group of authors, e.g., Bozeman (2000), Barjak (2011), Meissner and Sultanian (2007), and Polt et al. (2001a, b), led us to argue that a transfer-oriented research evaluation has to consider eight characteristics of KTT: (1) tradition and culture of a research discipline; (2) general understanding of research (for example, basic research vs. applied research); (3) degree and extent of transfer interaction; (4) existing transfer structures, processes, methods and instruments; (5) incentives and barriers for transfer activities; (6) (not surprisingly) existing financial support for transfer; (7) environmental conditions regarding research transfer; and (8) qualitative and quantitative performance indicators. Although we have discussed the KTT models and framework intensively, we do not claim that our list is complete. However, for the purposes of

Table 2 List of KTT characteristics

Characteristics	Description
Research tradition and culture	Scientific tradition and culture of the transfer agent. This includes several factors about the status of the research organization, such as natural or social sciences, basic or applied research, established or new disciplines
Transfer interaction	Characteristics of the interaction between the transfer agent and recipient. This includes, for instance, degree and frequency of transfer, transfer experience of each party, capacity of each of the transfer parties to desorb or absorb information
Internal conditions	Internal conditions that influence the transfer of knowledge and technology for both the transfer agent and the recipient, for example, financial support, transfer structures, processes, methods, instruments, and internal barriers and incentives
Environmental conditions	External incentives and barriers that influence the transfer of knowledge and technology, for instance market demands, regulations and promotion programs
KTT dimensions and determinants	Dimensions and determinants for measuring KTT performance and effectiveness

our analysis, we further summarized those eight characteristics to include a set of five distinct categories of characteristics: research tradition and culture (1 and 2), transfer interaction (3), internal conditions (4 and 5), environmental conditions (6 and 7), and KTT dimensions and determinants (8; see Table 2).

3.2 Approaches of Research Evaluation

Evaluation is a widely used word, and is used in different contexts with different meanings (e.g., Kromrey (2000); Stockmann et al. (2007)). Stockmann et al. (2007) highlights that evaluation in a scientific context is characterized by the following attributes: an evaluation has clearly defined objectives; it uses empirical and objective data collection methods for information generation; it uses precise, open and explicit evaluation criteria; it is based on systematic procedures applied by experienced evaluators; and it is the basis for the decision-making processes. Following the procedures of our research design, and considering this working definition, the next paragraphs present several evaluation taxonomies and concrete approaches extracted from academic literature. Table 3 summarizes the evaluated taxonomies of research evaluation.

In addition to general taxonomies of research evaluation, we also investigated concrete research evaluation approaches (see Table 4), namely Patton (2000) Utilization-Focused Evaluation, Stake (1983) Responsive Evaluation, Altschuld and Zheng (1995) Goal-Attainment Approach, and Beywl et al. (2004) Experimental Studies.

Table 3 Taxonomies of research evaluation

Taxonomy approach	Author	Description
Fourth generation evaluation	Guba and Lincoln (1989, 2001)	"In *Fourth Generation Evaluation* we described the historical evolution of evaluation practice: a first generation focused on measurement, a second generation focused on description, a third generation focused on judgment, and a fourth generation focused on negotiation (the hermeneutic/dialectic)." (Guba and Lincoln 2001, p. 8)
Program evaluation	Stufflebeam (2001)	Stufflebeam's taxonomy is one of the most foundational approaches for program evaluation adapted by various authors (cf. Beywl et al. 2004; Hansen 2005). Evaluation approaches are separated into: Pseudo-evaluation, Question/Methods-oriented Approaches, Improvement/Accountability Approaches, and Social Agenda/Advocacy Approaches
Program evaluation	Fitzpatrick et al. (2004)	Fitzpatrick et al. (2004) focus on the practical purpose of evaluation, and organize the approaches to program evaluation into six categories: objectives-oriented, management-oriented, consumer-oriented, expertise-oriented, adversary-oriented, and participant-oriented
Evaluation theory tree	Alkin (2004); Christie and Alkin (2008)	"This branch we have designated methods since in its purest form, it deals with obtaining generalizability, or knowledge construction, [. . .]. Another branch we call the valuing branch. [It] firmly establishes the vital role of the evaluator in valuing. The third major branch is use, which, [. . .] focused on an orientation toward evaluation and decision making." (Alkin 2004, p. 12)

In the next section the resulting relationships between the characteristics of KTT and the taxonomies and concrete approaches of research evaluation will be summarized.

4 Findings and Discussion

In our study, we investigated the characteristics of KTT, and to what extent they are taken into account by existing research evaluation approaches. The resulting relationships between the characteristics of KTT, namely research tradition and culture, transfer interaction, internal conditions, environmental conditions and KTT-dimensions and determinants, and the approaches by Guba and Lincoln (1989), Stufflebeam (2001), Fitzpatrick et al. (2004), Christie and Alkin (2008), Patton (2000), Stake (1983), Altschuld and Zheng (1995), Beywl et al. (2004) are summarized in Table 5. For our analysis, we used the broad categories of low (L),

Table 4 Concrete research evaluation approaches

Evaluation approach	Author	Description
Utilization-Focused Evaluation	Patton (2000)	Approach is built upon the evaluation standards of the Joint Committee of Standards; evaluation should be judged by the utility and actual use of evaluation results; ideal process for a Utilization-Focused Evaluation consists of five sequential phases
Responsive Evaluation	Stake (1983)	Approach aims at integrating the interests of all stakeholders of an evaluation project; evaluation should be a response to the interests of all stakeholders; evaluator's role is to promote communication between stakeholders, document the evaluation progress, and to make sure that all stakeholders are equally integrated
Goal-Attainment Evaluation	Altschuld and Zheng (1995)	Approach aims to identify if, and to what extent, concrete program goals have been fulfilled; evaluation based on pre-established research program purposes and goals; typically is implemented internally and addresses those responsible for research program or its sponsors
Experimental Studies	Beywl et al. (2004)	Approach aims to identify causal effects between a research program and its implications; effect-based evaluation of research programs; two random samples are taken: one that participates in the program, and one in a control group that does not; considered to be helpful in identifying small but crucial effects

medium (M), and high (H). Therefore, the entries in this table should be interpreted in relative rather than in absolute terms. The table is based on a literature-review, and thus does not reflect empirical relationships.

Our analysis revealed that most research evaluation taxonomies consider the importance of KTT to be low to medium. However, our KTT characteristics are given some consideration, particularly in Guba and Lincoln (1989) negotiation category of their generation model, Stufflebeam (2001) improvement/accountability and social agenda/advocacy approaches, Fitzpatrick et al. (2004) goal, management and consumer-oriented categories of purpose-oriented model, and Christie and Alkin (2008) 'use' and 'valuing' branches of the tree model. With respect to research tradition and culture, all of the presented models provide support for this category at least to a medium level. However, especially within Fitzpatrick et al. (2004) management-orientation category, we saw a slightly higher level of consideration of research tradition and culture. In a very similar vein, all taxonomies do consider transfer interactions, but only at a low level. With the exception of Stufflebeam (2001) social agenda/advocacy approach, which concentrates on societal change, all taxonomies describe mechanisms that regard internal conditions at a medium level, for instance, by respecting organizational goals or stakeholder perspectives. Finally, with respect to external conditions, we argue that Fitzpatrick et al. (2004) consumer-oriented category, in particular, points towards a high level of consideration of external environments. In summary, although we identified

Table 5 KTT in research evaluation

Approach	Research tradition and culture	Transfer interaction	Internal conditions	Environmental conditions	KTT-dimensions and determinants
Generation model (Guba and Lincoln 1989)	M	L	M	M	–
Program evaluation taxonomy (Stufflebeam 2001)	M	L–M	M	M	–
Purpose-oriented program evaluation (Fitzpatrick et al. 2004)	M–H	L	M	H	–
Evaluation tree (Christie and Alkin 2008)	M	L	M	L	–
Utilization-Focused Evaluation (Patton 2000)	L	L	L	L	L
Responsive Evaluation (Stake 1983)	M	M–H	M	M	L
Goal-Attainment Approach (Altschuld and Zheng 1995)	L	M	M	L	L
Experimental Studies (Beywl et al. 2004)	L	L	M	M	L

L not included or little included, *M* partly included, *H* highly included

some consideration of transfer activities in the investigated taxonomies, we state that the taxonomies do not take KTT into account in a way that merits its growing importance.

It is apparent from Table 5 that the concrete evaluation approaches consider knowledge and technology similarly at a low to medium level. This is especially true of Patton (2000) utilization-focused approach, which does not predetermine any content of an evaluation, and only provides low support for our categories. For the other evaluation approaches, we argue that research tradition and culture are only supported to a low level, with the exception of Stake (1983) approach, for which the holistic perspective on an evaluation objective would properly include these categories. Similarly, transfer interaction is regarded at a low level, again with the exception of Stake (1983) approach, which aims to consider the interests of various stakeholders and to establish a dialogue during the evaluation process. Thereby, we argue, this approach promotes dialogue on KTT between the transfer participants. With respect to internal and external research evaluation, we see no significant difference between Responsive Evaluation, Goal-Attainment Evaluation and Experimental Studies. However, due to its focus on the evaluation goals, we see a lower effect of external conditions in the case of Goal-Attainment Evaluation. Finally, all of the evaluation approaches analyzed in this paper only provide a low level of support when it comes to concrete dimensions or determinants for measuring KTT.

Summing up, we conclude that the approaches discussed above have so far shown little support for KTT. With respect to the evaluation taxonomies, we see most support for it in Stufflebeam (2001) systematic categorization of evaluation approaches. Regarding concrete evaluation approaches, we argue that Stake (1983) early approach of evaluation provides most support for our analysis criteria. However, as we will further conclude in the next section, we assert that a lot of development and adaption will be needed to prepare the approaches described above for the growing importance of KTT.

Conclusion

In this paper we have investigated the characteristics of KTT, and to what extent they are taken into account by existing research evaluation approaches. Despite the growing importance of KTT as the third mission in the scientific community, evaluation practice still overlooks transfer activities, or mainly focuses on quantitative determinants. However, at least to some extent, it ignores the fact that those determinants are not equally appropriate for the evaluation of KTT, especially in different scientific disciplines or traditions. Our findings, therefore, contribute in two main areas towards a better understanding of how research evaluation can consider KTT.

Our first contribution was the identification of the characteristics of KTT that should be considered in a transfer-oriented research evaluation and develop a set of distinctive categories for our analysis, consisting of: (1) research tradition and culture; (2) transfer interaction; (3) internal conditions; (4) environmental conditions; and (5) KTT dimensions and determinants. These characteristics have several implications for our analysis and interpretation: First, KTT is strongly affected by the transfer partners underlying research tradition and culture. For instance, transfer activities might differ in both goals and importance in different scientific disciplines. Second, the nature of the transfer interaction is an important characteristic of KTT. This implies that KTT is not only dependent on the transfer agent, but also its interplay with the transfer recipient as well as potential intermediates. Third, KTT depends on internal and external conditions that might result in transfer drivers or barriers. Internal drivers and barriers, for instance, are the transfer agent's financial support for transfer activities, its procedural, methodological or instrumental maturity. External drivers and barriers, on the other hand, are the existing and potential market demands, legal regulations and external promotion programs. Finally, KTT is defined through many different

(continued)

dimensions and determinates. Existing literature suggests that quantitative indicators alone are not sufficient for evaluating KTT.

Second, our findings contribute to the existing body of knowledge in the academic field of research evaluation by analyzing the state-of-the-art fields of research evaluation, and by identifying the potential for improvement with respect to KTT evaluation. With respect to taxonomies of research evaluation, such as the Guba and Lincoln (1989) fourth generation evaluation, Stuffelbeam (2001) and Fitzpatrick et al. (2004) program evaluation approaches and Alkin (2004) evaluation theory tree provide different categories for ordering evaluation approaches. However, our results provide evidence for the argument that, so far, KTT is either neglected in research evaluation or only considered implicitly. In a very similar lane, the results from the concrete research evaluation approaches indicate that KTT has not yet found its way into concrete research evaluation. None of the analyzed approaches provided explicit consideration of KTT but integrate it implicitly in terms of research aims or goals. However, in those cases the existing approaches mostly refer to quantitative determinants, such as number of publications or patents.

The evidence from our study suggests several implications for academic inquiry and evaluation practice in the future. Our study offers new implications for academic inquiry in the domain of research evaluation. From our perspective, it is necessary to further develop and adapt the existing approaches towards a transfer-oriented research evaluation. In this respect, we see Stufflebeam (2001) taxonomy and Stake's Responsive Evaluation (1983) as the most promising candidates for such enhancements. For instance, based on the KTT characteristics presented in Sect. 3.1 Stufflebeam (2001) systematic taxonomy including pseudo-evaluation, question/method-oriented approaches, improvement/accountability approaches, and social agenda/advocacy approaches could be extended by an additional, fifth category regarding transfer-oriented evaluation approaches. With respect to the concrete research approaches we regard Stake (1983) dialog-based, responsive evaluation approach as a good starting point for a transfer-oriented research evaluation. This relates also to the fact that in research—especially basic research—process and input control are more important than output control (Ouchi 1979; Schaarschmidt 2012). Regarding, for example, the KTT effectiveness of an individual research or research program through scientific publications, we argue that a dialog-based evaluation of the quality of publications is more accurate and expressive than a mere quantitative number of articles or journal rankings.

Finally, our review of literature provides valuable insights for evaluation practice. Regarding the goal-orientation of most of the existing approaches, evaluators today have to predefine explicit transfer goals to be able to

(continued)

evaluate them. However, this is not always possible, especially in complex research programs. In such cases, we suggest using Stake (1983) responsive approach and its participative mechanisms to help stakeholders to continuously identify and evaluate ongoing transfer activities. These implications are both new and important to research evaluation literature, and complement the existing body of knowledge in this field.

However, as is the case for any research project, our study choices create some limitations, which offer fruitful avenues for research. Depending on our research approach, our study lacks empirical evidence in favor of our investigations and conclusions. Much more empirical research is needed to further develop and test our findings in qualitative and quantitative research designs from different perspectives (for example, transfer agent or transfer recipient), or on different levels (for example, program, institution, group, or individual). Although those limitations must be kept in mind, we hope our findings provide new insights for academics and practitioners alike.

References

Alkin MC (2004) Evaluation roots. Sage, Thousand Oaks

Altschuld JW, Zheng HY (1995) Assessing the effectiveness of research organizations: an examination of multiple approaches. Eval Rev 19(2):197–216

Argote L, Ingram P, Levine JM, Moreland RL (2000) Knowledge transfer in organizations: learning from the experience of others. Organ Behav Hum Decis 82(1):1–8

Barjak F (2011) Wissens- und Technologietransfer als Interaktion: Theoretische Überlegungen und Fallbeispiele aus der Schweiz. Peter Lang, Bern

Beise M, Spielkamp A (1996) Technologietransfer von Hochschulen: Ein Insider-Outsider-Effekt. ZEW discussion papers

Beywl W, Speer S, Kehr J (2004) Wirkungsorientierte Evaluation im Rahmen der Armuts-und Reichtumsberichterstattung. Perspektivstudie. Bundesministerium für Gesundheit und Soziale Sicherung (BMGS), Bonn

Bozeman B (2000) Technology transfer and public policy: a review of research and theory. Res Policy 29(4):627–655

Christie CA, Alkin MC (2008) Evaluation theory tree re-examined. Stud Educ Eval 34(3):131–135

Czarnitzki D, Rammer C, Spielkamp A (2000) Interaktion zwischen Wissenschaft und Wirtschaft in Deutschland. Ergebnisse einer Umfrage bei Hochschulen und öffentlichen Forschungseinrichtungen. ftp://ftp.zew.de/pub/zew-docs/docus/dokumentation0014.pdf. Accessed 24 Mar 2014

Fitzpatrick J, Sanders J, Worthen B (2004) Program evaluation: alternative approaches and practical guidelines. Pearson, Boston

Guba EG, Lincoln YS (1989) Fourth generation evaluation. Sage, Newbury Park

Guba EG, Lincoln YS (2001) Guidelines and checklist for constructivist (aka fourth generation) evaluation

Hansen HF (2005) Choosing evaluation models: a discussion on evaluation design. Evaluation 11 (4):447–462

Hoekman BM, Maskus KE, Saggi K (2005) Transfer of technology to developing countries: unilateral and multilateral policy options. World Dev 33(10):1587–1602

Horizon (2014) Horizon 2020. The EU framework programme for research and innovation. http://ec.europa.eu/programmes/horizon2020/. Accessed 24 Mar 2014

House ER (1978) Assumptions underlying evaluation models. Educ Res:4–12

Kromrey H (2000) Qualität und Evaluation im System Hochschule. In: Stockmann R (ed) Evaluationsforschung. VS Verlag für Sozialwissenschaften, Wiesbaden, pp 233–258

Kuhlmann S, Heinze T (2003) Informationen zur Forschungsevaluation in Deutschland – Erzeuger und Bedarf. Gutachten für die Geschäftsstelle der Deutschen Forschungsgemeinschaft. Frauenhofer ISI Discussion Papers Innovation

LeBris J, Pousttchi N, Suszka M (2010) Good Practices im Wissens- und Technologietransfer von Hochschulen. Handlungsempfehlungen, Rechercheergebnisse und Literatursammlung aus der TEA-Grundlagenstudie, Augsburg

Markowski N, Grosser K, Kuhl R Analyse von Barrieren und Hemmnissen beim Wissenstransfer zwischen Hochschulen und KMU. Düsseldorf Working Papers in Applied Management and Economics, vol 5. University Library, Düsseldorf

Meissner D, Sultanian E (2007) Wissens-und Technologietransfer: Grundlagen und Diskussion von Studien und Beispielen. CEST, Zentrum für Wissenschafts-und Technologiestudien

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

Patton MQ (2000) Utilization-focused evaluation. Springer, The Netherlands

Polt W, Gassler H, Schibany A, Rammer C, Schartinger D (2001a) Benchmarking industry—science relations: the role of framework conditions. Sci Public Pol 28(4):247–258

Polt W, Rammer C, Schartinger D, Gassler H, Schibany A (2001b) Benchmarking industry-science relations in Europe—the role of framework conditions. Sci Public Pol 28(4):247–258

Roessner JD (2000) Technology transfer. In: Hill C (ed) Science and technology policy in the US. A time of change. Longman, London

Schaarschmidt M (2012) Firms in open source software development: managing innovation beyond firm boundaries. Springer, Wiesbaden

Schmoch U (2000) Wissens- und Technologietransfer in Deutschland. Fraunhofer-IRB, Stuttgart

Scriven M (2013) Key evaluation checklist (KEC). http://www.michaelscriven.info/images/KEC_3.22.2013.pdf. Accessed 12 Mar 2014

Stake R (1983) Program evaluation, particularly responsive evaluation. In: Evaluation models, vol 6. Springer, The Netherlands, pp 287–310

Stockmann R, Caspari A, Meyer W (2007) Evaluation. Eine Einführung. UTB L (Large-Format), vol 8337. Budrich, Leverkusen

Stufflebeam D (2001) Evaluation models. New Dir Eval 2001(89):7–98

Wegner DM, Giuliano T, Hertel PT (1985) Cognitive interdependence in close relationships. In: Ickes W (ed) Compatible and incompatible relationships. Springer, New York, pp 253–276

Performance Assessment and Professional Development in University Teaching

Tina Seidel

Abstract The quality of instruction at university level is typically measured by means of student evaluation. Student evaluation is based on scientific knowledge about effective instruction and includes aspects such as instructional clarity, teacher support, and a positive learning climate. Student evaluation has become a widely and commonly accepted approach of measuring instructional quality. However, with regard to professional development, this approach has a number of limitations, mainly due to the fact that the reflection about results and possible improvement is reserved only for the individual instructor being assessed. Therefore, new approaches have been developed that allow for collaborative reflection and systematic training. One example is the 1-year training program "Learning to Teach" in which novice university instructors are trained and systematically evaluated. Implications for the governance of knowledge-intensive organizations, such as universities, are drawn with regard to the importance of systematic training of novices.

1 Introduction

Universities are among the most important sources for young adolescents to acquire scientific knowledge and to prepare for future careers in knowledge-intensive occupations. They are accountable for providing this knowledge in a way that supports students in their development of general cross-curricular competencies, such as critical thinking, analytic reasoning, problem solving, and communicating knowledge in a scientific way (Shavelson 2012). In this way, knowledge is not just acquired in a factual way, but in a way that represents a deep understanding of subject matter that can be used flexibly and adopted for various applications.

Given the many changes and reforms in university education worldwide, let alone the European Union's Bologna Reform, universities experience much

T. Seidel (✉)
TUM School of Education, Technische Universität München, Friedl Schöller-Endowed Chair for Teaching and Learning Research, Munich, Germany
e-mail: tina.seidel@tum.de

I.M. Welpe et al. (eds.), *Incentives and Performance*,
DOI 10.1007/978-3-319-09785-5_28

criticism. At the heart of this criticism is the discussion that university instruction is still oriented towards traditional teaching methods (Wissenschaftsrat 2008), methods that are strongly linked to approaches of direct instruction that offer limited opportunities for students to think critically, to develop reasoning skills, to learn to communicate knowledge in a scientific way, and to apply knowledge to new problems. Therefore, a call has been issued for instructional improvement within universities (Braun and Hannover 2008).

It is interesting to note that the discussion and criticism about the quality of university instruction is lacking empirical evidence. In educational research to date, the limited studies focus mainly on describing typical instructional approaches and rating the quality of instruction by means of student questionnaires. These subjective student ratings of instructional quality are then linked to students' test performances. The tests are typically developed and applied by the instructors or university departments themselves, and therefore often lack standardization and comparability that would allow for systematic comparisons across courses and universities (Pellegrino et al. 2001). However, a number of intervention studies have also been initiated. These studies investigate changes in instructional practices and their effects on student learning more systematically (cf. Stes et al. 2010).

Student evaluation, in this context, has evolved as a commonly implemented approach to assess instructional quality at universities, either with the aim to monitor the quality of instruction and to detect outliers, or with the aim to use these findings as a basis for professional development (Spiel 2011). In the first case, student evaluation is used as a performance assessment tool with different high- or low-stake consequences for instructors. In some countries, university teachers might face serious consequences regarding their careers if their student evaluation results are not satisfactory (or even outstanding). In other countries, obtaining positive evaluation results is "nice-to-have", but it is not connected to high-stake decisions regarding future university-based careers. Especially in the context of high-stakes consequences connected to student evaluation, there should be solid empirical evidence about the objectivity, validity, and reliability of the applied instructional assessment instruments.

In this chapter, my contribution to the question of incentives and performances in knowledge-intensive organizations will come from an educational science perspective: first, the chapter will outline the conceptual models about teaching and learning that are used as a basis for assessing instructional quality, since those models are decisive for questions of objectivity, validity, and reliability. Thereby, a focus will be given to both the student's perspective and to the instructor's perspective, in comparison to the perspectives of students and external experts. Research findings are presented that show the importance of using different perspectives in judging multiple facets of instructional quality. Furthermore, the professional development of university instructors will be addressed. Specifically, research questions regarding the pre-requisites, practices and learning developments of novice university teachers are targeted since these persons have a strong influence on effective reform in higher education. In concluding, it will be argued that university teachers should be trained more systematically rather than by using the currently applied "learning-by-doing" approach.

2 Assessing Quality in University Instruction

2.1 Underlying Models of Teaching and Learning

In assessing the quality of instruction, most studies worldwide are oriented towards a basic process-product model (Dunkin and Biddle 1974) that differentiates between instructional *processes*, learning *products*, *context* such as student characteristics, and *prerequisites,* such as teacher qualification (Fig. 1).

The idea behind the model is that good quality in instructional processes will lead to positive learning outcomes on the side of students. The effect of instructional processes is mediated by student characteristics such as pre-knowledge, interest in a topic, and self-concept of ability, but also by the socio-economic background or the influence of peers and media. The quality of instructional processes depends on the quality of the prerequisites of the instructors, meaning that the more professionally qualified instructors are, the better the quality of the instruction provided by them.

Based on this fundamental model, instructional effectiveness research has evolved this model into more and more differentiated components. The so-called "supply-use-models" are the basis of modern instructional effectiveness research (Brühwiler and Blatchford 2011; Fend 1998; Helmke et al. 2008; Pauli and Reusser 2003). An important development in this context is the fact that all the research has shown that one cannot assume a direct effect of instruction on learning (Winne 1987). Instead, multiple factors on the part of students and teachers, as well as within the institutional context, have to be considered in order to model a teaching and learning process appropriately. An example of a supply-use-model is given in Fig. 2.

A very important aspect of the supply-use-model is the fact that the students' perceptions of instruction are indeed a very important process factor (den Brok et al. 2010; Seidel 2006). If students perceive a learning environment as supportive, it is more likely that they will actively process the learning content and will be intrinsically motivated to learn. Deep-learning processing in connection with

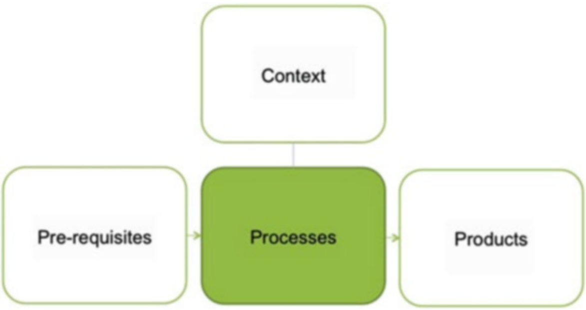


Fig. 1 Classic process-product model of teaching and learning

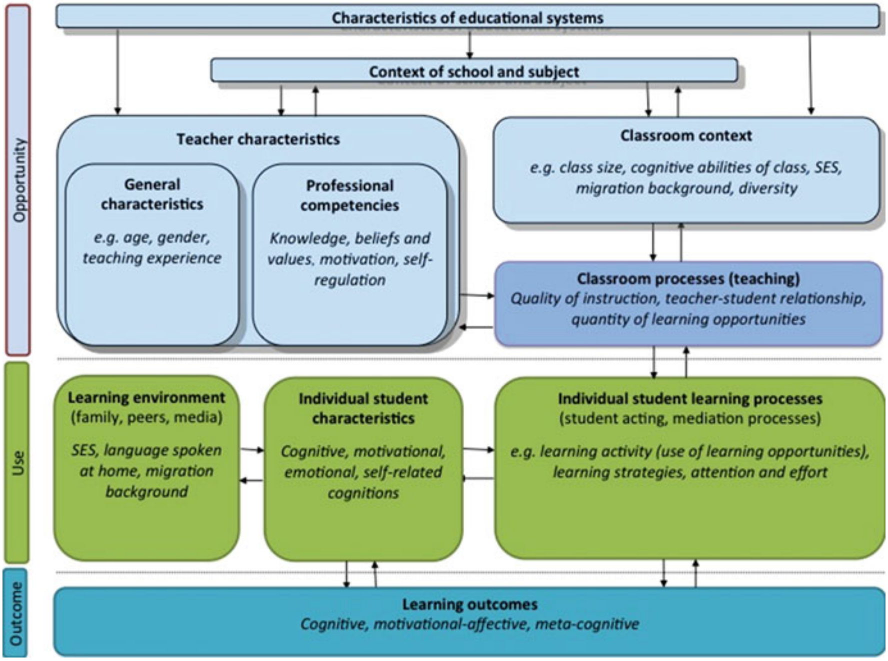


Fig. 2 Supply-use-model of teaching and learning [adapted from Brühwiler and Blatchford (2011)]

intrinsic learning motivation, in turn, is strongly connected to positive learning outcomes, such as a deep understanding of content, interest in a learning domain, and the use of learning strategies as a skill for life-long learning. Given these findings, it makes sense to ask students to evaluate the instructional quality of university teaching. The question is, however, "What are they asked to evaluate?" Research has shown quite some important factors to consider when answering this question.

2.2 *Instructional Quality from Student Perspective*

In assessing instructional quality from the student perspective, three important factors have to be considered. First, students can evaluate only those instructional aspects that they are able to perceive and to which their attention is drawn. This aspect is not trivial, since one's perception is strongly influenced by one's knowledge. Think about the situation in which an accident occurs, and different persons pass by. If one person has a professional background in medicine, the person will apply this specialized knowledge immediately to the situation; for example, the person may check whether someone is injured and, if so, to what degree. Another person with a background in law will perceive different aspects; for example, he or she might try to determine who may have caused the accident or which persons might have witnessed it. This process, called *professional vision*, is regarded as a key component of professional expertise (Goodwin 1994). When applying this concept to instruction, it is very important that persons who judge the quality of the instruction actually notice those aspects that are relevant for student learning and, in addition, use their professional knowledge in order to reason adequately about the observed instruction. To reason means to adequately describe what happens during instruction, to explain the situation with regard to what constitutes professional instruction, and to provide predictions regarding possible consequences of the instruction on student learning (Seidel and Stürmer 2014; Sherin et al. 2011). The fact that knowledge drives our vision and that the knowledge should be professionally related has important consequences for using student evaluation for assessing instructional quality. It means that student evaluation must be based on elements that students can actually notice, and that the students have enough knowledge to be able to reason about the observed instruction. The kind of elements that have been shown to be "noticeable" for students are more general, didactical aspects, such as goal clarity, classroom management, teacher support, and learning climate (Clausen 2001; Klieme and Tippelt 2008). However, when it comes to aspects that require deep understanding of the content taught, as well as the specific pedagogy related to the content (pedagogical-content knowledge; Shulman 1987), students are not a valid source of information. It has been shown that students rarely meet an adequate judgment of these aspects since they are simply not able to notice these aspects and to reason adequately about them (Clausen 2001).

The second factor to be considered when using student evaluation is the timeframe used for students to judge instructional quality. In many cases, students are generally asked whether the teaching of an instructor throughout a course has been clear, whether there was a positive learning atmosphere, and similar questions. However, instructional research has shown that student judgments of instruction are more valid if the ratings are rather very specific and situative. This has to do with the fact that instruction can vary considerably during the course of a semester and that students are asked rather generally about the instruction, and therefore find it difficult to provide an appropriate aggregation of their perceptions. Thus, the validity of student ratings can be systematically improved if student ratings are either focused on specific instructional events or if multiple measurements are aggregated to an overall value for instructional quality (Cohen et al. 2003).

The third factor to be considered refers to the question, "What should students evaluate regarding the importance of instruction on their learning outcomes?" It has been shown that the evaluation of the above-mentioned general-didactical aspects, such as goal clarity, classroom management, teacher support, and learning climate are systematically related to learning outcomes (Hattie 2008; Seidel and Shavelson 2007). In this sense, these are indeed relevant aspects for student evaluation of instructional quality at universities. However, judging the quality of these instructional approaches does not explain the mediating processes involved in student learning. In this context, psychological theories are used in educational research. One prominent theory is the "self-determination theory" (SDT; Deci and Ryan 2004), which allows one to model under which instructional conditions students are intrinsically motivated to learn and, in turn, to use active learning strategies for a deep understanding of learning content (Krapp and Prenzel 2011). SDT assumes three basic psychological needs that have to be fulfilled for a learner in the process of instruction: experience of competence, autonomy, and relatedness to others. If these three needs are addressed in the process of instruction, students are highly likely to experience self-determined motivation and to acquire a deep understanding of the content to be learned. In this vein, instructional effectiveness research, in the school context, is increasingly focusing student evaluation of instruction on these aspects since they provide a better basis to understand the processes involved between instruction and learning outcomes. In the context of student evaluation of university instruction, these elements are not yet being implemented in a systematic way.

2.3 Instructional Quality from the Instructor Perspective

Instructional quality can be also assessed from the perspective of instructors, either by using self-reports or by judgments of peer-instructors or external professional experts. In the case of instructors' self-reports, research to date provides evidence in the direction that instructors tend to overestimate the quality of their instruction (Könings et al. 2013). In a study by Johannes et al. (2013), for example, university teachers ($N = 20$) and their students ($N = 325$) rated the instructional quality by means of a questionnaire targeting their teaching approaches (teacher-focused or student-focused), the use of various didactical methods, and goal clarity. The instruction of one course session was videotaped, and the students were asked to rate the instructional quality directly after videotaping by means of a questionnaire. In addition, external experts rated the videotaped instruction. Mean judgments of instructors, students, and external experts were calculated, and means of the three perspectives were compared (Table 1). In addition, correlations between self-report and student judgments respectively expert judgments were calculated.

Table 1 Perspective differences between instructor self-reports, student judgments, and expert judgments (Johannes et al. 2013)

	Means			Correlations	
Instructional quality	Self-report	Students	Experts	Students	Experts
Teacher focus (TF)	3.23 (0.91)	4.02 (0.48)**	3.63 (0.84)	0.46*	0.29
Student focus (SF)	4.45 (0.72)	4.34 (0.55)	3.27 (0.71)*	0.50 *	0.63**
Didactical methods (DM)	4.16 (1.08)	4.18 (1.06)	3.92 (1.06)+	0.45*	0.44 *
Goal clarity (GC)	4.70 (0.64)	4.89 (0.45)	4.25 (0.97)*	0.42+	0.47 *

+$p < 0.10$; *$p < 0.05$; **$p < 0.01$; Range: 1 = "does not apply" to 5 "fully applies"

Comparisons between perspectives show that, from the perspectives of the students, the instruction was more teacher-focused (i.e., focus was on transmitting knowledge by means of direct instruction by the teacher) compared to the instructors' self-reports. Student-focused instruction referred to elements that allowed for student cooperation, exchange of ideas, and more constructivist orientations in teaching. Regarding these elements, self-reports and student judgments did not differ; however, experts tended to be more critical regarding the implementation of these aspects. This might have to do with the fact that external experts know about a wider variety of possibilities to implement student-focused teaching and therefore are more critical with regard to the opportunities that could be used. The same pattern in perspective differences is also found regarding the implementation of a variety of didactical methods in instruction, as well as in goal clarity. If instructors themselves, as well as their students, lack professional knowledge about opportunities that could be used in the area of student focus, didactical methods, and goal clarity, their judgments are likely to be biased, since their frames of reference are more limited compared to those of external experts.

In order to investigate whether rankings of instructors are comparable across perspectives, correlations between instructors' self-reports and student judgments as well as expert judgments, were calculated (see Table 1, right columns). A significant, coherent pattern emerged from these correlations. Instructor self-reports were systematically intercorrelated with the corresponding student, as well as with the expert judgments, and all correlations were comparable in size. Taken together with the comparison of mean differences, this indicated a more homogeneous perception of instructional quality, with a tendency towards a more critical expert perception.

A comparison of perspectives is also of interest when it comes to relationships between instructional quality and learning outcomes. As pointed out above, intrinsic learning motivation, as well as the deep-level processing of learning content, are two aspects in the learning process that are decisive for a deep understanding of subject matter. In the study of Johannes et al. (2013) judgments of the three perspectives of instructors' self-reports, student ratings, and expert ratings were correlated with students' reports about intrinsic motivation and deep-level processing directly after the assessed instruction. The correlational pattern between assessment of instructional quality from different perspectives and the quality of student learning processes is quite systematic for the perspectives of students and external experts. From these two perspectives, students are more likely to report about intrinsic learning motivation and a deep processing of learning content if their instructors' teaching was rated as student-focused, using a variety of teaching methods, and providing goal clarity. Teacher-focused approaches were not related to these two learning outcomes from any of the three perspectives. In comparison to these two perspectives, instructors' self-reports were not systematically related to the learning outcomes, with the exception of goal clarity and deep level processing.

The fact that it is challenging for university teachers to judge their own instruction is reasonable, since the process of teaching is highly dynamic and complex. In such a situation, reflection is difficult and strongly influenced by outstanding events in the complex flow of activities (Berliner 2001). For this reason, instructional effectiveness research typically focuses on the perspectives of students or external experts (e.g., by means of video-analysis; Seidel and Shavelson 2007). However, whether instructors actually meet the perspective of their students (in the sense of congruity) is also regarded as an important indicator for the professionalism of an instructor (Könings et al. 2013; Vermunt and Verloop 1999). Frictions respectively the degree of congruity between perspectives is therefore often used as a basis for teacher professional development.

3 Professional Development in University Teaching

Being a university teacher means that one has to deal with multifaceted, instructional situations. Next to academic teaching goals, emotional, motivational, attitudinal, cultural, social, and other factors greatly affect teaching-learning interactions during instruction (Floden 2001; Shuell 1996). Hence, for university instructors to develop professionally, they must acquire professional knowledge of, and the ability to deal with, relevant teaching and learning factors. Nevertheless, university instructors in many countries are typically trained in their discipline and have no systematic training in teaching at the university level. Therefore, novice instructors often experience the classes they teach as "learning-by-doing." Exclusive use of a learning-by-doing approach leads to a situation in which novice instructors typically adopt traditional existing teaching approaches that they know from their own previous experiences as students (Seidel and Hoppert 2011). Furthermore, implicit and subjective theories of novice university instructors exert a major influence on the way instruction is provided. These subjective theories typically do not necessarily correspond to a research-based professional knowledge of teaching and learning (Berliner 2001; Hammerness et al. 2002). Therefore, it is argued that universities have to invest in the professional development of their instructors and that systematic training would be the best way to provide a sound scholarly basis for professional university instruction (Stes et al. 2010). As one example, the project "Learning to Teach" (LehreLernen) provided and tested a structured training program for novice university teachers (Johannes et al. 2012). Novice university teachers were targeted since they are more sensitive towards the new experience of teaching and motivated to learn about teaching. Furthermore, in comparison to more advanced researchers (post-docs, junior faculty professors) they do not yet have to face dealing with the multitude of tasks in research and teaching which often keeps more advanced instructors away from participating in 'extra' activities of teacher professional development. The "Learning to Teach" program included basic didactical training, as well as video-based workshops and feedback (see Fig. 3).

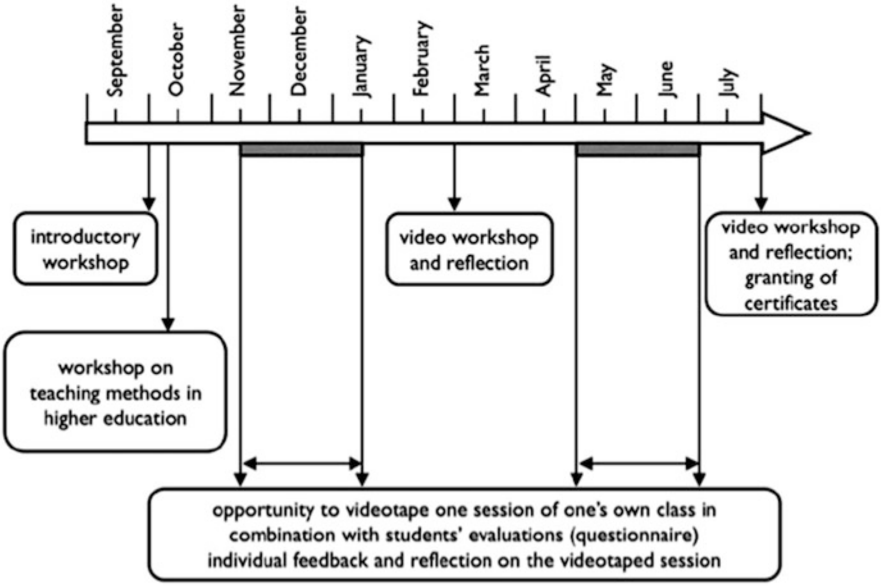


Fig. 3 Structure of 1-year training for novice university instructors in the program, "Learning to Teach" (Johannes et al. 2012)

In the training program, groups of approximately 15 novice university instructors were trained per course. Initially, an introductory workshop acquainted trainees with ways to discuss video cases constructively as a learner community. Basic professional knowledge about didactical teaching methods was provided in a second workshop. For individual and group reflection about the quality of instruction, trainees were videotaped while conducting one course session. They received individual feedback from the training team with regard to their instructional quality. In addition, short video clips were selected as a basis for collaborative discussions among trainees during two reflection workshops. The training program was conceptualized in a way that incorporates components of teacher professional development that have been shown to be effective. These include aspects such as reasonable duration, content focus, coherence, active learning, and collective participation (Desimone 2009). In addition, the incorporation of video-based learning tools in a learner community provides situative and concrete learning experiences that trainees can more easily put into action in their courses (Korthagen 2010). Overall, the evaluation of the training program has shown positive effects on the learning developments of the trainees. At the beginning of the training, the novice university instructors had general and unspecific goals and theories about instructional quality. A typical answer was "to teach well" or "to teach better than what I had experienced as a student" (Johannes et al. 2012). The general goals were unrelated to strategies they could use in order to achieve their goals. During the course of the training, they systematically specified their goals because they acquired professional knowledge about what constitutes instructional quality. Furthermore, they were also better able to notice these aspects when they observed their own teaching and they were better able to use their knowledge in order to reason adequately about their teaching. Despite the fact that their teaching still tended to be rather teacher-focused (Seidel and Hoppert 2011), they gained more confidence in teaching and, with this confidence, then started to try out new teaching approaches and methods that would allow for more student focus and collaboration (Johannes and Seidel 2012).

The "Learning to Teach" program is one example how teacher professional development components that have been shown as effective in research can be implemented. However, a variety of other approaches could be used as well: This would include for example, peer instructors or more advanced instructors observing and evaluating instruction and providing feedback regarding their perspectives. Peer-observation is particularly important regarding feedback about the quality of content and the specific content-pedagogy since student evaluations do not provide this kind of information. A second variation can be that instructors form a "learning club" in which instructors plan courses together, implement them and reflect upon their experiences collaboratively. In the context of school instruction, these forms are being already implemented, for example, as video clubs or reflection groups (Sherin et al. 2011).

Conclusions and Future Directions

This chapter contributed to the field of assessment of instructional quality from the perspective of educational science by demonstrating that underlying models of teaching and learning drive what is regarded as a "good quality" of instruction. Quality of instruction can be viewed from the viewpoint of its relevance for student learning, meaning that those aspects should be assessed that are systematically related to positive outcomes in student learning. In order to judge instructional quality, professional knowledge about these aspects of instructional quality is required. When students evaluate instructional quality, they possess knowledge about a number of these aspects, typically with regard to more general didactical approaches, as well as their subjective experiences of being supported in their learning. Research shows that, given these aspects of instructional quality, student ratings in the context of student evaluations provide a quite valid source of information. However, with regard to instructional aspects that require content knowledge as well as content-related pedagogy, students lack this kind of professional knowledge. In this respect, external experts or peer-instructors are better sources for judging these instructional aspects. General didactical approaches are systematically related to important indicators for student learning. However, in the school context, content knowledge, as well as content-related pedagogy, has been shown to be important aspects of instructional quality (Baumert

(continued)

et al. 2010; Shulman 1987). In university instruction, these elements are hardly ever addressed, and a focus is given to student evaluations of the above outlined general didactical aspects. For the future development of the field, however, content and content pedagogy should be addressed more intensively.

Furthermore, it must be noted that complex teaching and learning models are intensively applied in the context of teaching and learning in schools. Thereby, student as well as teacher prerequisites (knowledge, interest, home environment, etc.) are considered when investigating the influence of instruction on student learning. In the field of university instruction, research based on elaborated teaching and learning models is widely lacking. Most studies refer to a very basic process-product model by linking instructional variables to general student test scores as learning outcomes. As research within the school context has long demonstrated, these approaches are very limited, since they do not provide a sound basis for describing, explaining, and predicting teaching and learning processes (Winne 1987). Thus, future research in the context of university instruction should focus more intensively on applying complex models of teaching and learning in order to assess instructional quality. This would mean that more rigorous research programs have to be incorporated that include systematic testing of university students' prerequisites and developments during courses and study programs (e.g., pre-knowledge, motivation, self-regulation, problem-solving). Furthermore, instructor competencies (e.g., knowledge of content, pedagogy) would have to be assessed and instructional quality has to be assessed by multiple perspectives and multiple times. These assessments then have to be linked systematically to changes in student development (so-called "value-added"). These kinds of approaches could be started with selected university departments as pioneer and then systematically expanded throughout university.

Finally, when it comes to learning to teach in university contexts, most universities still rely on an approach in which novice instructors start teaching by applying a learning-by-doing approach, with the result that traditional teaching methods are transferred from one generation to the next (Wissenschaftsrat 2008). In this context, the implementation of new instructional approaches, as it is called for in the context of university reform, will hardly find entrance to university instruction. Therefore, the need to provide novice university instructors with more structured and systematic training is indicated.

References

Baumert J, Kunter M et al (2010) Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. Am Educ Res J 47(1):133–180

Berliner DC (2001) Learning about and learning from expert teachers. Int J Educ Res 35:463–482

Braun E, Hannover B (2008) Zum Zusammenhang zwischen Lehr-Orientierung und Lehrgestaltung von Hochschuldozierenden und subjektivem Kompetenzzuwachs bei Studierenden. Zeitschrift für Erziehungswissenschaft Sonderheft 9:277–291

Brühwiler C, Blatchford P (2011) Effects of class size and adaptive teaching competency on classroom processes and academic outcome. Learn Instr 21(1):95–108. doi:10.1016/j.learninstruc.2009.11.004

Clausen M (2001) Unterrichtsqualität: Eine Frage der Perspektive? Empirische Analysen zur Übereinstimmung, Konstrukt- und Kriteriumsvalidität. Waxmann, Münster

Cohen DK, Raudenbush SW, Loewenberg Ball D (2003) Resources, instruction, and research. Educ Eval Pol Anal 25(2):119–142

Deci E, Ryan RM (eds) (2004) Handbook of self-determination research. University of Rochester Press, Rochester

den Brok PJ, van Tartwijk J, Wubbels T, Veldman I (2010) The differential effect of the teacher-student interpersonal relationship on student outcomes for students with different ethnic backgrounds. Br J Educ Psychol 80:199–221. doi:10.1348/000709909X465632

Desimone LM (2009) Improving impact studies of teachers' professional development: toward better conceptualizations and measures. Educ Res 38(3):181–199. doi:10.3102/0013189x08331140

Dunkin M, Biddle B (1974) The study of teaching. Holt, Rhinehart & Winston, New York

Fend H (1998) Qualität im Bildungswesen. Schulforschung zu Systembedingungen, Schulprofilen und Lehrerleistung. Juventa, Weinheim

Floden RE (2001) Research on effects of teaching: a continuing model for research on teaching. In: Richardson V (ed) Handbook of research on teaching. American Educational Research Association, Washington, DC, pp 3–16

Goodwin C (1994) Professional vision. Am Anthropol 96(3):606–633. doi:10.1525/aa.1994.96.3.02a00100

Hammerness K, Darling-Hammond L, Shulman LS (2002) Toward expert thinking: how curriculum case writing prompts the development of theory-based professional knowledge in student teachers. Teach Educ 13(2):219–243

Hattie J (2008) Visible learning. A synthesis of over 800 meta-analyses relating to achievement. Routledge, New York

Helmke A, Rindermann H, Schrader F-W (2008) Wirkfaktoren akademischer Leistungen in Schule und Hochschule. In: Schneider W, Hasselhorn M (eds) Handbuch der Pädagogischen Psychologie. Hogrefe, Göttingen, pp 145–155

Johannes C, Seidel T (2012) Professionalization of university teachers—teaching approaches, professional vision and teacher identity in video-based training. Z Erzieh 15(2):233–251. doi:10.1007/s11618-012-0273-0

Johannes C, Fendler J, Seidel T (2012) Teachers' perceptions of the learning environment and their knowledge base in a training program for novice university teachers. Int J Acad Dev. doi:10.1080/1360144X.2012.681785

Johannes C, Soellner R, Seidel T (2013) Teaching and learning in university courses from teacher, student and video-based expert perspectives (Manuscript submitted for publication)

Klieme E, Tippelt R (2008) Qualitätssicherung im Bildungswesen: Eine aktuelle Zwischenbilanz. Zeitschrift für Pädagogik 53(Beiheft):7–13

Könings K, Seidel T, Brand-Gruwel S, van Merrienboer JJG (2013) Differences between students' and teachers' perceptions of education: profiles to describe congruence and friction. Instr Sci. doi:10.1007/s11251-013-9294-1

Korthagen FAJ (2010) Situated learning theory and the pedagogy of teacher education: towards an integrative view of teacher behavior and teacher learning. Teach Teach Educ 26(1):98–106. doi:10.1016/j.tate.2009.05.001

Krapp A, Prenzel M (2011) Research on interest in science: theories, methods, and findings. Int J Sci Educ 33(1):27–50. doi:10.1080/09500693.2011.518645

Pauli C, Reusser K (2003) Unterrichtsskripts im schweizerischen und im deutschen Mathematikunterricht. Unterrichtswissenschaft 31(3):238–272

Pellegrino JW, Baxter GP, Glaser R (eds) (2001) Knowing what students know. The science and design of educational assessment. National Academy Press, Washington, DC

Seidel T (2006) The role of student characteristics in studying micro teaching-learning environments. Learn Environ Res 9(3):253–271

Seidel T, Hoppert A (2011) Merkmale von Lehre an der Hochschule. Ergebnisse zur Gestaltung von Hochschulseminaren mittels Videoanalyse. Unterrichtswissenschaft 39(2):154–172

Seidel T, Shavelson RJ (2007) Teaching effectiveness research in the past decade: the role of theory and research design in disentangling meta-analysis results. Rev Educ Res 77 (4):454–499. doi:10.3102/0034654307310317

Seidel T, Stürmer K (2014) Modeling and measuring the structure of professional vision in pre-service teachers. Am Educ Res J. doi:10.3102/0002831214531321

Shavelson RJ (2012) Assessing business-planning competence using the collegiate learning assessment as a prototype. Empir Res Vocat Educ Train 4(1):77–90

Sherin MG, Jacobs VR, Philipp RA (eds) (2011) Mathematics teacher noticing: seeing through teachers' eyes. Routledge, New York

Shuell TJ (1996) Teaching and learning in a classroom context. In: Berliner DC, Calfee RC (eds) Handbook of educational psychology. Macmillan, New York, pp 726–764

Shulman LS (1987) Knowledge and teaching: foundations of the new reform. Harv Educ Rev 57 (1):1–22

Spiel C (ed) (2011) Evaluation universitärer Lehre – zwischen Qualitätsmanagement und Selbstzweck. Waxmann, Münster

Stes A, Min-Leliveld M, Gijbels D, Van Petegem P (2010) The impact of instructional development in higher education: the state-of-the-art of the research. Educ Res Rev 5(1):25–49. doi:10.1016/j.edurev.2009.07.00

Vermunt JD, Verloop N (1999) Congruence and friction between learning and teaching. Learn Instr 9:257–280

Winne PH (1987) Why process-product research cannot explain process-product findings and a proposed remedy: the cognitive mediational paradigm. Teach Teach Educ 3(4):333–356. doi:10.1016/0742-051x(87)90025-4

Wissenschaftsrat (2008) Empfehlungen zur Qualitätsverbesserung von Lehre und Studium. Wissenschaftsrat, Berlin

Part VIII
Incentives and Performance: What We Can Learn from a Fairy Tale

Cinderella Between Rigor and Relevance

Rolf Wunderer

Abstract Using a well-known fairy tale from the Brothers Grimm's collection as an example, I outline the patterns of behavior and consequences of a one-dimensional orientation towards quantitative measuring. This task occurs in connection with the boycott of an individualized ranking by economic researchers, which occurred in the context of their disciplines' history of dogmata.

1 Cinderella, Rigor and Relevance

The tales in the collection of the Brothers Grimm (2007/1997; Brüder Grimm 1999/1837), are called "world document heritage" and perceived as the most published work in the German language; they have been translated into 170 languages (Schede 2004, p. 60). The Brothers were also effective role models in higher education politics. Following their "protestation letter" written on 17 November 1837 to King Ernst August, their sovereign and employer, with five colleagues from the University of Goettingen ("Die Göttinger Sieben"), they were discharged without notice and expelled from the country. They accused the King of breaching the constitution (Martus 2010, p. 383ff.). They took this risk consciously, a decision that today's lecturers would scarcely make.

With the fairy tale "Cinderella", the Brothers Grimm provide an impressive allegory on the formation of un-reflected and one-sided rigorous criteria with the prince's choice of his bride. If the reader lacks knowledge concerning fairy tales, here is a refreshment:

Cinderella is extremely bullied by her stepmother, her two daughters and even her father following her mother's death. She copes with strong resilience. The king of the realm plans to wed his son and invites Cinderella's family. She, the bullied person, would like to go with them. Her stepmother states that she is willing to let her go but only after successfully sorting good and bad lentils. It is only thanks to

R. Wunderer (✉)
Institut für Führung und Personalmanagement, Universität St. Gallen, c/o Hardungstr. 22, St. Gallen, Switzerland
e-mail: rolf.wunderer@unisg.ch

I.M. Welpe et al. (eds.), *Incentives and Performance*,
DOI 10.1007/978-3-319-09785-5_29

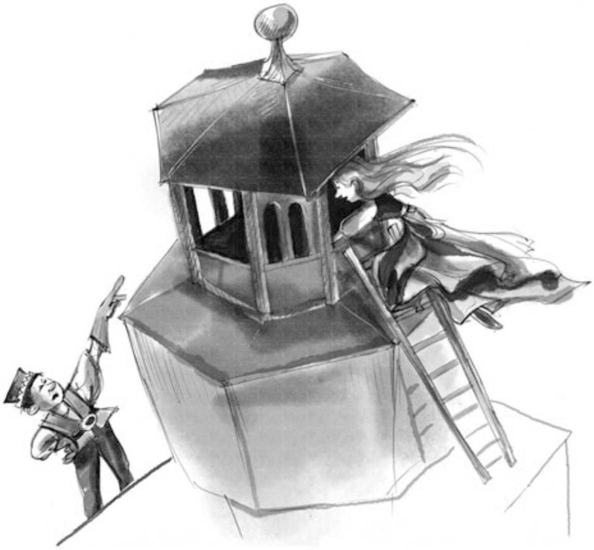


Fig. 1 Cinderella flees to the dovecote (Bromundt 2014). All right for this illustration belongs to R. Wunderer

her heavenly helpers, the doves sent by her deceased mother, that she passes the tests. Still, she is denied participation in the event three times in breach of contract. However, Cinderella decides with some more heavenly help to attend the event secretly in a beautiful evening gown.

The moment the prince sees the unknown beauty, he is enamored and makes only her "his dancer" out of a great number of attractive candidates. The following two events also occur. The desired Cinderella consistently flees the festivities right before the end. The prince follows her obsessively each time, stalking her. He makes her father destroy her possible hideouts: the dovecote and a tree in her family garden (Fig. 1).

After the third night, the suitor cunningly gets a hold of one of the shoes of the fairy tale heroine. Although he had close visual and communicational contact with her for three nights, he announces the examination of all aspiring females with only one measurement criterion: who fits in the tiny and delicate shoe? In this process, he first measures Cinderella's mean sisters, who, following their mother's explicit advice, shortened their toes and heels with a knife given by the stepmother. One after another is measured, with the one concept of quantitative measurement remaining the only valid method of evaluation (Fig. 2).

When Cinderella's mother's heavenly helpers (two doves), with their "*rucke di guck, rucke di guck, blood is in the shoe, the shoe is too small, the right bride is still at home*," show the "rigor-fetishist" his false decisions, he sends back the stepsisters with no reflection, exchanging them without remorse. Finally, Cinderella is brought from her stove and is "measured" successfully. The prince looks at the much more relevant face of this object of evaluation and promptly leads her to the altar, the only way for the bride to escape her extremely destructive family culture. The fairy tale does not tell whether the marriage was happy. There should be some doubts also in

Fig. 2 Consequences of a rigorous orientation for measuring (Bromundt 2014). All right for this illustration belongs to R. Wunderer

the precondition of choice of a partner on the one-dimensional criterion of shoe size.

Thus, how can the king and queen manage their succession? By means of the "Prince Charles solution", i.e., waiting for grandchildren? Even this test demonstrates that he was totally overwhelmed both mentally and socio-emotionally. Just by looking at this evaluation and cooperation strategy, it can be concluded that the prince would be totally overstrained in the social "governance" of the country (Fig. 3).

Conclusion: One-dimensional thinking and acting is not only relevant for research. Many mergers have failed by prioritizing the analysis of the usual balance sheets and finances, which are easier to measure than different leadership and organizational cultures (Wunderer 2010).

2 "Handelsblatt Ranking" and the Shoe-Size

Extremely constricted criteria such as "shoe size" are no less common in our science than in Grimm's fairy tale Cinderella. However, in contrast to this fairy tale, persons concerned with our science have resisted such criteria. This paper

Fig. 3 The prince at the third shoe assessment (Bromundt 2014). All right for this illustration belongs to R. Wunderer

focuses on the "Handelsblatt Ranking"—a person-oriented and individual ranking by a German business newspaper. There was an unprecedented collective opposition by professors. The ranking appeared with a high degree of pretension and the headline, "Germany in search for the super professor". The researchers had individual, external, public and very selective judgments of performance, which were criticized in 2010 by *Kieser* and *Osterloh* in their farewell lectures (Gysin 2010 regarding Osterloh; Kieser 2011). There were no consequences by "Handelsblatt" in the second edition of the rankings in 2012. Shortly before, Kieser and Osterloh (2012a) released an appeal with the maxim, "End this nonsense" (Kieser and Osterloh 2012b). 339 researchers followed this appeal for boycott within a few days.

Never had I experienced comparable reactions of concerned people in enterprises, public businesses or universities when it came to the evaluation or co-development of staff evaluation—not even in a two-year long study for the Federal Ministry for Research and Technology. The researchers interviewed in 1976 rated the assessment of publications for performance measurement as generally "minor" and the number of citations as "mediocre" and asked for a "quantitative-qualitative combination of criteria" (Wunderer et al. 1979, p. 126ff.).

In 1980, *Neuberger* even characterized employee appraisals as "ritual of self-deception". His reason is the behavior of participants in this concept of evaluation: "They accept the procedure and blame the short-comings to themselves and not to the system. Furthermore, the system boosts rule knowledge, gives and stabilizes power and status, can be taken as anticipatory socialization and for ex post legitimation" (Neuberger 1980, p. 42; Wunderer 2012).

Although the critiques of a ranking of researchers have reached high levels, concrete suggestions for improvement are still hard to come by. In addition, the general rejection of researcher rankings with measurable and comparable criteria or substitutes with a still-limited realization of feasible alternatives seems to be premature to us. Extensions via monographs, more evaluators and objects of rating, central demands for function and roles for researching lecturers are being tested or already used for evaluation purposes by universities.

3 Another Controversy of Methods in Business Economy's History of Dogmata?

This spontaneous boycott could enter into the business' history of dogmata as the newest "dispute of methods". Business studies seldom address the history of dogmata and ideas anymore. Therefore, one cannot find a single keyword about it in the encyclopedia of business (Wittmann et al. 1993), which encompasses 5,070 columns. At the end of the nineteenth century, the national economists in universities founded a closed shop. In 1898, the excluded business economists founded—following the role model of technical universities—their own business schools with lasting success (e.g., St. Gallen). Afterwards, economic "Humboldtians" tried to "integrate" the young business science by "take over", but only a few succeeded. The second struggle of methods began in the thirties inside the just-establishing discipline between two fractions. The "private business studies" (Rieger 1928) put return on investment and shareholders as entrepreneurs in their focus. Furthermore, they criticized the lack of scientific character and value orientation of certain colleagues. These focus on "Gemeinwirtschaft" (common values) and the system of forming inclusion for society and employees, partly as "plant communities" with an early stakeholder-like approach in their business economics and practical regard for applied science (Nicklisch 1922/1932; Schmalenbach 1920—see also Kruk et al. 1984; Schneider 1981; Weibler and Wald 2004; Wunderer 1967). These academic devaluations led, in my view, to a "stable" inferiority complex in front of the elder stepsister and her scientific self-understanding. This most likely encouraged the orientation of economic researchers in the mainstream of the "believing communities of economists" (Binswanger 1998) as well as, perhaps, a psychological "identification with the devaluating researchers".

4 Regarding the Weighting of Rigor and Relevance

A ranking of researchers by publications and citations in very few journals with a high impact seems to be as one dimensional as the classification of brides by shoe size. It does not take into account books or monographs, even if these books are dissertations or habilitations. It hides central requirements on function and the role of professors, like providing expert advice, teaching and further training, and orientation toward students. In addition, it discriminates against some behavioral scientific research.

The A-B/C-journals legitimize themselves more and more with the “rigor approach”, using it as science in front of “relevance” for their own discipline or the practice of business (Kieser 2011, 2013), which is no more useful than shortening the feet of stepsisters with a knife to fit the criteria of shoe size. Beyond the fairy tale, there is a decisive side-effect: it leads to an early imprinting of young lecturers and, with it, of students. Instead of risky innovation, statistical rigor is encouraged without equivalent weighting of the relevance of the themes dealt with at least in academic practice. Academic practitioners answered in a survey that even relevant results of scientific research “do not find attention” and stated that “the scientific system with its A-B-C- journals, its duct of methods, and its systems of gratification does not reward necessarily a transfer-oriented, comprehensibly edited research” (Armutat 2014, p. 55).

In a recent study, young scholars in academia from the fields of business and economics advocate including qualitative criteria in research assessments. Specifically, the findings of the study suggest that young scholars in academia from the fields of business and economics call for considering more strongly the extent to which knowledge is gained or to which a project is innovative and creative. But the young scholars also advocate taking the benefits for practice and society into account (Wollersheim et al. 2014); not only for emeriti, has this given reason to hope.

5 Contra Shoe Sizes as Measuring Criteria for Professors

Which conclusions can be drawn to avoid one-sided criteria for judging complex activities such as those of professors?

- Understand business administration as applied science, thereby orientating the relevance of theoretical and empirical research to this subject.
- Install qualitative research and its judgment adequately into evaluation concepts, and do not simply strengthen quantitative research “rigorously”.
- Lower the extreme weighting of publication in high-impact journals only for researcher rankings and the selection of young scientists and professors.
- Weight the relevance of research projects higher in advance and with multiple evaluations, but differentiating in a way appropriate for the situation.

- Fill the positions of heads of institutes and projects complementary-wise, not by "assumed similarity". Seldom can university lecturers individually fulfill all important requirements. Some are gifted critics, and others are curious and creative.
- Build rough clusters instead of seemingly exact range of rankings and differentiate them with subjective judgments (also see Dagett 2012).
- Include central stakeholders in the evaluation and thus—via multiple evaluations—increase the "subjective objectivity". Make potential differences in evaluation transparent.
- Include book publications; especially multiply ranked dissertations and habilitations.
- Do not align the intelligibility of scientific publications based solely on insiders.
- Why not include academic practitioners of specialized fields in the editor's committee to judge the relevance of publications additionally?
- Do not reduce cooperation with academic practitioners to sponsoring, congresses, interviews, counseling or leading exercises.

Conclusion: Excellent researchers and "rigor standards" are necessary but are not enough for determining the excellence of universities in the way that shoe size is not sufficient for correctly and relevantly assessing the choice of a partner.

References

Armutat S (2014) Zwei Seiten einer Medaille. Personalwirtschaft Extra 1:52–56

Binswanger H-Chr (1998) Die Glaubensgemeinschaft der Ökonomen. Murmann, München

Bromundt C (2014) Three original illustrations (Figs. 1, 2, 3). All rights for the illustrations belong to R. Wunderer

Brüder Grimm (eds) (1999) Kinder und Hausmärchen (KHM), 19th edn. Artemis und Winkler, Düsseldorf, Zürich

Dagett W (2012) Achieving excellence through rigor and relevance. In: Bambrick-Santoyo P (ed) Leverage leadership: a practical guide to building exceptional schools. Wiley, San Francisco

Gysin R (2010) Man redet nicht mehr, man zählt. http://www.uzh.ch/news/articles/2010/man-redet-nicht-mehr-man-zaehlt.html. Accessed 16 Sep 2012

Handelsblatt (2012) Betriebswirte bekämpfen Forschungsranking. http://www.handelsblatt.com. Accessed 10/11 Sep 2012

Kieser A (2010) Akademische Rankings: Die Tonnenideologie der Forschung. Leicht gekürzte Fassung seiner Ende Mai gehaltenen Abschiedsvorlesung. In: Frankfurter Allgemeine Zeitung, 11 Jun 2010

Kieser A (2011) Between rigour and relevance: co-existing institutional logics in the field of management science. Soc Econ 33:237–247

Kieser A (2013) JOURQUAL – der Gebrauch, nicht der Missbrauch, ist das Problem. Die Betriebswirtschaft 72(1):93–210

Kieser A, Osterloh M (2012a) BWL- Forscher rufen zum Boykott des Handelsblatt-Rankings auf – mit Update. http://www.handelsblatt.com

Kieser A, Osterloh M (2012b) Den Unfug schnell beenden – Professoren boykottieren das BWL-Ranking des Handelsblatts. Forschung Lehre 19(10):820–821

Kruk M, Potthoff E, Sieben G (eds) (1984) Eugen Schmalenbach: der Mann, sein Werk, die Wirkung. Schäffer, Stuttgart

Martus S (2010) Die Brüder Grimm – eine Biographie, 2nd edn. Rowohlt, Berlin

Neuberger O (1980) Rituelle (Selbst-)Täuschung – Kritik der irrationalen Praxis der Personalbeurteilung. Die Betriebswirtschaft 40(1):27–43

Nicklisch H (1932/1922) Die Betriebswirtschaft, Stuttgart

Rieger W (1928/1959) Einführung in die Privatwirtschaftslehre, 2nd edn. Palm & Enke, Erlangen

Schede HG (2004) Die Brüder Grimm. Deutscher Taschenbuch Verlag, München

Schmalenbach E (1920) Grundlagen dynamischer Bilanzlehre, 2nd edn. Gloeckner, Leipzig

Schneider D (1981) Geschichte betriebswirtschaftlicher Theorie. Oldenbourg, München/Wien

The Brothers Grimm (2007/1997) The complete fairy tales, with illustrations by A. Rackham. Wordsworth Edition, Hertfordshire, pp 121–128

Weibler J, Wald A (2004) 10 Jahre personalwirtschaftliche Forschung – Ökonomische Hegemonie und die Krise einer Disziplin. Die Betriebswirtschaft 64(3):259–272

Wittmann W et al (1993) Handwörterbuch der Betriebswirtschaft, vol 5, 3rd edn. Poeschel, Stuttgart

Wollersheim J, Welpe IM, Ringelhan S (2014) Sein und Sollen: Bewertung von Forschungsleistungen in den Wirtschaftswissenschaften. Forschung Lehre 4(2014):276–278

Wunderer R (1967) Systembildende Betrachtungsweisen der Allgemeinen Betriebswirtschaftslehre und ihr Einfluss auf die Darstellung des Unternehmers. Duncker und Humblot, Berlin

Wunderer R (2010) Führung in Management und Märchen. Luchterhand, Köln

Wunderer R (2012) Rituelle (Selbst)täuschungen? Zur Beurteilung von Wissenschaftlern durch externes Ranking mit kollektiver Enttäuschung. In: Stein V, Müller S (eds) Aufbruch des strategischen Personalmanagements in die Dynamisierung. Nomos, Baden-Baden, pp p165–p171

Wunderer R, Boerger M, Löffler H (1979) Zur Beurteilung wissenschaftlich-technischer Leistungen – eine empirische Studie zur Personalbeurteilung in Forschungsorganisationen des Bundes. Nomos, Baden-Baden