SSTC Report 3/2013

Performance Measurement and Quality Assurance in Higher Education and Research

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Schweizerischer Wissenschafts- und Innovationsrat
Conseil suisse de la science et de l'innovation
Consiglio svizzero della scienza e dell'innovazione
Swiss Science and Innovation Council

The Swiss Science and Innovation Council

The Swiss Science and Innovation Council SSIC is the advisory body to the Federal Council for issues related to science, higher education, research and innovation policy. The goal of the SSIC, in line with its role as an independent consultative body, is to promote a framework for the successful long term development of Swiss higher education, research and innovation policy.

Der Schweizerische Wissenschafts– und Innovationsrat

Der Schweizerische Wissenschafts– und Innovationsrat SWIR berät den Bund in allen Fragen der Wissenschafts–, Hochschul–, Forschungs– und Innovationspolitik. Ziel seiner Arbeit ist die kontinuierliche Optimierung der Rahmenbedingungen für die gedeihliche Entwicklung der Schweizer Bildungs–, Forschungs– und Innovationslandschaft. Als unabhängiges Beratungs–organ des Bundesrates nimmt der SWIR eine Langzeitperspektive auf das gesamte BFI–System ein.

Le Conseil suisse de la science et de l'innovation

Le Conseil suisse de la science et de l'innovation CSSI est l'organe consultatif du Conseil fédéral pour les questions relevant de la politique de la science, des hautes écoles, de la recherche et de l'innovation. Le but de son travail est l'amélioration constante des conditions–cadre de l'espace suisse de la formation, de la recherche et de l'innovation en vue de son développement optimal. En tant qu'organe consultatif indépendant, le CSSI prend position dans une perspective à long terme sur le système suisse de formation, de recherche et d'innovation.

Il Consiglio svizzero della scienza e dell'innovazione

Il Consiglio svizzero della scienza e dell'innovazione CSSI è l'organo consultivo del Consiglio federale per le questioni riguardanti la politica in materia di scienza, scuole universitarie, ricerca e innovazione. L'obiettivo del suo lavoro è migliorare le condizioni quadro per lo spazio svizzero della formazione, della ricerca e dell'innovazione affinché possa svilupparsi in modo armonioso. In qualità di organo consultivo indipendente del Consiglio federale il CSSI guarda al sistema svizzero della formazione, della ricerca e dell'innovazione in una prospettiva globale e a lungo termine.

SSTC Report 3/2013

Performance Measurement and Quality Assurance in Higher Education and Research

Swiss Science and Technology Council SSTC

A More Purposeful and Sensible Use of Performance Measurement and Evaluation in Science – Ten Theses

Adopted by the SSTC on the 25th of June 2013

Sybille Reichert

Beyond Performance Review – The Search for a New Approach to Quality Assurance at Universities

Contents

Summary
Zusammenfassung
Résumé

E

The present paper was written by Dr Sybille Reichert on behalf of the Swiss Science and Technology Council (SSTC), and is based on a review of the key literature. It was drafted between October 2012 and March 2013, and follows up on research Dr Reichert undertook in 2010 on behalf of the SSTC about the quantification of research assessment. In its further investigation of the limits of research assessment, the paper examines the qualitative methods which are used particularly in peer review processes. The analysis takes into account the context of the use of different assessment methods, their cognitive, psychological and sociological circumstances, and the benefits and effects they have not only on the science system but on scientists as well. The study is not based on the independent collection of empirical data but is rather a critical evaluation and synthesis of the relevant international literature. A variety of European science systems are taken into account, as research assessment methods are in widespread use internationally.
In addition to a critical analysis of the possibilities, limits, and effects of prevailing research assessment methods, this study also served the SSTC by giving it a basis for discussing alternate ways to improve quality, and to strengthen confidence in publicly-funded research and in the government that finances it. The result of the discussion led the SSTC to formulate the ten theses which precede the study by Dr Reichert. The SSTC would like to thank Prof. Dr Dr h.c. Margit Osterloh of the Warwick Business School and Prof. Dr Jacques Lanarès of the University of Lausanne for their valuable comments on the paper and for their contributions to the discussion during the SSTC's plenary meeting in June 2013.

D

Das folgende Diskussionspapier wurde im Auftrag des Schweizerischen Wissenschafts- und Technologierats (SWTR) auf der Basis einer eingehenden Sichtung einschlägiger Forschungsliteratur zwischen Oktober 2012 und März 2013 von Dr. Sybille Reichert verfasst. Es schliesst an die Fragestellungen einer früheren Untersuchung im Auftrag des SWTR zur Quantifizierung der Forschungsbewertung (2010) an und führt die Frage nach den Grenzen der Forschungsbewertung fort, indem auch qualitative Methoden der Forschungsevaluation, insbesondere im Rahmen von Peer Review-Verfahren, ins Visier genommen werden. Dabei werden Verwendungskontexte, kognitive, psychologische und soziologische Bedingungen, Nutzen und Auswirkungen verschiedener Forschungsevaluationsmethoden auf Wissenschaftssystem und Wissenschaftlerinnen und Wissenschaftler untersucht. Das Diskussionspapier stützt sich dabei nicht auf eine eigene empirische Datenerhebung, sondern setzt sich kritisch-synthetisierend mit der internationalen Forschungsliteratur zu dem Thema auseinander. Da die Forschungsevaluationsmethoden international verbreitet sind, werden Untersuchungen verschiedener europäischer Wissenschaftssysteme einbezogen.
Neben einer kritischen Auseinandersetzung mit den Möglichkeiten, Grenzen und Auswirkungen gängiger Forschungsbewertungsmethoden diente die vorliegende Untersuchung dem SWTR auch als Grundlage für die Diskussion alternativer Wege der Qualitätsentwicklung und der Vertrauensbildung für die öffentliche Forschung und den sie finanzierenden Staat. Die Diskussionsergebnisse fanden Eingang in die dem Diskussionspapier vorangestellten Thesen des SWTR. Der Dank des SWTR gilt Prof. Dr. Dr. h.c. Margit Osterloh, Warwick Business School, sowie Prof. Dr. Jacques Lanarès, Universität Lausanne, für ihre wertvollen Kommentare zum vorliegenden Diskussionspapier und ihre Beiträge zur Diskussion anlässlich der Plenarsitzung des SWTR im Juni 2013.

F

Le présent document de discussion, basé sur une revue systématique de la littérature secondaire du domaine conduite entre octobre 2012 et mars 2013, a été réalisé par la Dr Sybille Reichert sur mandat du Conseil suisse de la science et de la technologie (CSST). Il fait suite à la recherche effectuée pour le CSST en 2010 au sujet de la quantification de l'évaluation de la recherche scientifique. Le document reprend la question des limites de l'évaluation, en intégrant aussi les méthodes qualitatives, comme l'évaluation par les pairs (Peer Review). L'analyse porte ainsi sur les contextes d'utilisation de différentes méthodes d'évaluation de la recherche, sur leurs conditions cognitives, psychologiques et sociologiques, ainsi que sur leurs bénéfices et effets pour le système scientifique et pour les chercheurs. Le document ne se fonde pas sur des données statistiques compilées par l'auteur, mais expose une synthèse critique de la littérature internationale à ce sujet. Compte tenu du fait que les méthodes d'évaluation de la recherche sont largement répandues, le document prend en considération un éventail de différents systèmes scientifiques en Europe.
Outre le but de fournir une analyse critique des possibilités, des limites et des effets des méthodes actuelles d'évaluation de la recherche, le présent document a également servi de base de discussion aux membres du CSST pour explorer la question des méthodes alternatives visant à assurer la qualité et à instaurer la confiance dans la recherche publique et dans son financement par l'Etat. Les résultats de cette discussion se manifestent dans les thèses formulées par le CSST, qui se situent dans la première partie de la publication. Le CSST exprime sa gratitude à la Prof. Dr Dr h.c. Margit Osterloh, Warwick Business School, ainsi qu'au Prof. Dr Jacques Lanarès, Université de Lausanne, qui ont commenté le présent document et participé à sa discussion lors de la séance plénière du CSST de juin 2013.

Part One

Swiss Science and Technology Council SSTC

A More Purposeful and Sensible Use of Performance Measurement and Evaluation in Science – Ten Theses

In the following theses, the Swiss Science and Technology Council recommends the use of methodologically well-considered measures of performance, or evaluations of scholarship, that are also purposeful, subject-oriented, and guided by reason. Each thesis is backed by a brief description or justification. The theses are based on Dr Sybille Reichert's "Beyond Performance Review – The Search for a New Approach to Quality Assurance at Universities" paper of April 2013.

1 The Use of Time

The time researchers and teaching staff spend in Switzerland on evaluating scholarly accomplishments and writing reports should be empirically investigated in order to assess how proportionate this effort is, relative to the time actually spent on scholarly research and teaching.

Researchers and academics report that since the 1990s, assessments of scholarly activity have become increasingly frequent. Time that might have been devoted to teaching or research must now be spent on assessing or reviewing academic achievements. This is related to the fact that more and more resources are competitively awarded, both in Switzerland and abroad. Decisions about resource allocation, positions to be filled, and being able to publish, hence also about reputation in the discipline, rely as a rule on peer review. In addition to the increasing effort within the science system this represents, trustees, funders, politicians, and the public increasingly demand to know what scholars and scientists have accomplished.

2 Explanatory Power

The critique of methods used in, and the limited reliability of, performance assessments should be better communicated both within the research community and to the public. As a rule, every assessment of scholarly accomplishment should be accompanied by an explanation of the data it is based on so as to connect the assessment to an understanding both of its explanatory power and of its limitations.

Both performance measures using quantitative indicators and those involving qualitative assessments by expert evaluators (peer review) have methodological drawbacks. The reliability of statements about quality is limited, and both approaches lead to behavioural accommodations – not always in the desired direction – by those being assessed. In addition, peer review strains the capacity of science systems to their limits. Those who examine science as a practice, along with statisticians, thus increasingly criticize not just the process employed but also how the results generated by such assessments are used.

3 Quantitative Performance Indicators

Quantitative performance indicators should always be placed in context. If possible, in evaluations or reports, such indicators ought only to provide background information. Where quantitative assessments have little or no explanatory power, they should not be used at all.

The unthinking weight accorded to quantitative performance indicators or to positions in well-known rankings leads to a distorted picture of scholarly accomplishment. When such data has direct effects on resource allocation, undesirable effects may result, including leading researchers to avoid projects which have uncertain outcomes, scholars fragmenting results to increase the number of publications, or creating citation circles to increase impact.
The use of quantitative indicators only has the effect of increasing academic productivity when resource allocation strongly depends on such indicators. Effects of this kind are not the same as improving quality, however. Existing studies of the effects of such incentive systems show instead that they have a deleterious influence on quality.

4 Basic Financing and External Funding

The relationship between basic financing for research at universities and fixed-term external funding (project funding) should be so structured that a balance is created between the longer-term orientation of the research and the shorter, more dynamic development of projects. Success at obtaining external funding must not become the main hallmark of success, either of a person or of a research unit, in an assessment. Rather, it needs to be seen in conjunction with other success markers, institutional parameters, and against the backdrop of the particular discipline in question.

The capacity of a science system to sustain innovation depends on finding the right balance between basic funding that can be freely allotted and competitively-awarded external funding for fixed-term projects. The proportion of fixed-term external funding has increased dramatically in recent years, to the detriment of basic financing for research. This is troubling, because success at obtaining external funding is used to calculate the share research funds should have in future basic funding provided to universities, a "bonus" that reinforces the shift away from longer-term and freely allottable resources and towards short-term, project-bound means.

5 The Utility and Frequency of Evaluations

If evaluations with clearly formulated goals are used to identify and resolve particular problems, they can be useful, justifying the effort they take. But in their current forms, not all evaluations are necessary, and for that reason one also should avoid their routine use. Instead, one should consider each time whether existing discipline-specific self-evaluation practices or alternate reporting forms (such as particular types of annual reporting) could not replace evaluations. In any case, it is imperative that one reduce the frequency with which evaluations are carried out at Swiss research institutions.

Formative evaluation, among all the forms of performance assessment, is assessed the most positively not only in empirical studies but also by those who work in higher education. In it, peers engage in a dialogue about a research unit's potential for improvement and development; formative evaluation is recognized as an instrument for improving quality. Yet when such kinds of evaluations come from outside an institution, and are used in increasingly frequent and routine cycles, both their effectiveness and acceptance quickly dwindle, and their marginal utility declines. The effort of evaluation is not compensated for by the learning effects or an improved quality that it might generate. After two decades of developing and consolidating quality assurance more generally, and in evaluating disciplines more specifically, the utility of evaluations at many Swiss universities has become quite small.

6 The Use of Evaluations in Resource Allocation

Open dialogue is key if there is to be a genuine improvement in quality. Evaluations ought to make it possible to openly reflect on weaknesses as well as about the future. For that reason, they should not have *automatic* effects on the allocation of resources.

Peer review remains highly regarded, particularly when part of a formative evaluation in which self-evaluation and external assessment initiate a dialogue about a research unit's potential. But given its limited reliability and its potential for triggering unintended effects, peer review is insufficiently robust for making allocation decisions be *directly* dependent on it. The potential effect such review might have on resource redistribution severely curtails the willingness of those being evaluated to work together constructively, and it considerably limits institutional learning as well.

7 Intrinsic Motivations

Science systems are based on cultural practices that are rooted in the intrinsic motivations of those who work in them. Intervening only to control should therefore – and to the greatest extent possible – be avoided so as to not endanger this motivational basis. Instead, performance assessments should primarily be used to support the self-reflective processes of scholars as well as the inherent orientation they have to produce quality work.

The key characteristics of academic practice include personal initiative, creative freedom, and self-determination with respect to the content, goals, methods and processes involved in research and teaching. These, in turn, are based to a high degree on the intrinsic motivations of scholars. For many of them, freedom and decision-making autonomy justify the high personal costs and uncertainties associated with pursuing an academic or research career. Constant reporting requirements, along with routinized appraisal and evaluation processes, erode this intrinsic motivation and undermine the considerable willingness researchers and teachers have not just to make an effort but also to take risks. Forms of performance assessment, in accreditation or evaluation processes, whose primary purpose is to control, weaken this cultural basis – as do the actual conditions under which scholarly work is carried out, and with it, the attractiveness of pursuing an academic career.

8 *Ex ante* Selection and Monitoring

Universities and institutions which fund academic research should fundamentally rethink the use of *ex post* evaluation methods. They should also rethink the extent to which performance assessments are used as measures of quality assurance. Instead, they should rely more on effective methods of quality improvement, including better *ex ante* hiring processes. They also should support assessment processes which are transparent, use coherent criteria, and of high quality. Just as effective is long-term monitoring of teaching staff and researchers.

If the quality of academic research is guided by training and selecting scholars and researchers *ex ante*, in accordance with the goals of the hiring institution, then quality assurance is considerably more effective and the relationship between cost and benefit better than attempts to control quality *ex post*, or by trying to improve performance through incentive systems that fail to understand the role intrinsic motivations play. Longer-term, continuous monitoring of research accomplishments by a group of peers who engage in a critical dialogue with academic staff has proven an effective complement to *ex ante* evaluations.

9 Trust and Creative Freedom

Achievement should be promoted primarily by strengthening motivations that already exist. Trust should be offered in advance, which means creative freedom should be guaranteed and promising developments or initiatives supported.

An academic life geared to confirming and steady "production" runs the risk over the longer term of becoming sclerotic. The growing culture of mistrust and the increasingly closely-woven fabric of controlling mechanisms promote mediocrity rather than excellence.

10 A Culture of Empowerment

A culture of empowerment characterized by mutual trust and ongoing dialogue should replace the increasing tendency of state and society to subject academic research to close control.

Rather than granting them greater autonomy, governments and political actors often express an increased desire to control the universities. Teaching staff and researchers therefore need to reach out to the public more, to report on what they are doing, and need to initiate a dialogue to help promote the culture shift on the part of those who support and fund the universities. But if New Public Management is trying to increase efficiency in research by using performance incentives and controls, it misjudges the nature of academic work. Detours and unforeseen developments, failed experiments or the risk an idea might fail, the falsification of a hypothesis – all these are the very preconditions for progress, and crucial aspects in the long-term ability of academic researchers to innovate.

Part Two

Sybille Reichert

Beyond Performance Review – The Search for a New Approach to Quality Assurance at Universities

"Externally and internally driven research assessment has incontestably become part and parcel of the research university enterprise. However, the explosion of various types of research assessment for a variety of users and purposes in recent years risks to create – or, one may argue, has already started to create – an obsession with measurement and monitoring, which may result in a 'bean counting' culture detracting from the real quality of research and the boundless search for new knowledge. [...]

The demands of producing large quantities of data may create or extend a culture of regulations, instructions, lists of good practice etc., marked by an obsession with measurement and monitoring, all of which might well detract from creative freedom, flexibility and productivity."

(The League of European Research Universities, LERU 2012)

1 Introduction

The following paper was written on behalf of the Swiss Science and Technology Council (SSTC); it was drafted between October 2012 and March 2013 on the basis of a detailed survey of the relevant research literature. It follows up on research undertaken in 2010 for the SSTC about the quantification of research assessment. In further pursuing questions about the limits of research assessment, it examines qualitative methods used in research evaluation, particularly peer review methods. The investigation includes the contexts of use, the cognitive, psychological and sociological circumstances, and the benefits and effects different research assessment methods have on the science system and on researchers. The study was not based on an independent collection of empirical data but is instead a critical evaluation and synthesis of the relevant international research literature. A variety of European science systems are taken into account, as research assessment methods are widely and internationally used.

The goal of the investigation is to not only critically examine the limits, opportunities and effects of current research assessment methods. The goal is also to create a basis for a discussion of alternate ways to improve quality, and to strengthen confidence in publicly-supported research and in governments that fund research.

2 Defining the Problem

The accountability of public universities to their funders has increased in the last decade as universities have gained more autonomy. Political decision-makers and the public expect that universities, now acting with greater freedom, both justify their decisions and demonstrate results. Throughout Europe, this increased university autonomy has led to developing new steering instruments and ever more differentiated reporting systems. Their purpose is to demonstrate to parliaments and public authorities that the result of relaxing state tutelage is responsible action, both individual and institutional, in and by universities, of a kind that simultaneously allows the realization of public goals. Hence, in the last decades strategy development, decision-making, and quality assurance processes have emerged, designed to make goals and their implementation as farsighted, sustainable, comprehensible, and verifiable as possible. The increase in institutional autonomy has therefore perforce been linked to new forms of steering and reporting obligations following the precept of "accountability", though it is an accountability which often restricts the individual autonomy of researchers and instructors. Increasingly, the question is whether or to what extent this limiting of research helps more than it hinders – or *vice versa*.

Even if the principle of "accountability" is largely beyond dispute today, opinions differ particularly with respect to steering instruments (their form, scope and implementation) and modes of verification about how effective they are. The hope that one can find a reliable way to depict and judge research achievements in order to fairly allocate resources, if not improve quality, appears impossible to fully realize. The path taken also seems to engender worrisome side effects in the form of escalating demands for expert opinions and still more reports. For that reason, in various countries, there have been increasingly critical calls for a review of performance assessments in order to estimate the added value the new steering instruments bring. Concerns have been voiced about both material and immaterial costs, and the effects evaluation has on research and on researchers. Over the last two decades, research assessment has continued to proliferate in Great Britain and evaluation itself has become a core controversy in science policy. Some countries, notably Norway, Germany and France, are only now introducing nationally comparative evaluations of disciplines and institutionalizing research assessment for the purpose of strategy formation. Forerunners such as Great Britain and the Netherlands now ask how they can adapt their methods to mitigate the side effects associated with performance assessment schemes. As described in Chapter 5, it is not just the quantitative, indicator-based performance assessments that have come under fire, but the qualitative, peer review-based evaluation methods as well, though for different reasons.

Beyond pragmatic questions about limiting the costs and effort of assessment, selecting the proper indicators, choosing forms of evaluation, or the cycles in which it is to occur, one can also discuss research assessment at a more basic level. It is precisely in the more mature systems, meaning those countries that have engaged in performance assessment and quality improvement efforts for decades, that questions are being raised about whether the highly developed and ever-growing set of tools for assessment and quality assurance has created a culture of mistrust which undermines, over the longer term, the foundations of an innovative, creative science sector. Are there alternatives to a system of accountability that rests on mistrust? Could not a higher education system like Switzerland's, one characterized by internationally recognized research results, afford a culture of trust? This would save time and support achievements and creativity by using different forms for reporting and more sensible means of quality improvement that both appear, and act, in a manner which controls less and develops achievement and creativity more.

The following discussion addresses both of these levels more thoroughly. After a brief historical introduction, a discussion of the central aspects, experiences and empirical studies of the various forms of research assessment follows. From them, core theses are derived that are meant to serve as a basis for further discussions in the SSTC. The focus here is not so much on the possible adoption of existing research assessment practices elsewhere. Instead, the emphasis is on finding new approaches that can strengthen the trust of political bodies which finance research, as well as that

of the public in general, both in the utility of scholarship and academic research practices and in its ongoing ability to produce results. Behind this stands a vision of a Swiss science sector characterized by a lively, content-filled dialogue between researchers and the social and economic environment they are in. It is also a vision in which the need for time-consuming, bureaucratic, invasive performance controls can be minimized, leaving scholars more time to engage in teaching, research, and communicating with those outside academe who are interested in the results of this research. Our society has never been as driven by, and as dependent on, research knowledge as it is today, yet the relationship of society and political bodies to research institutions is increasingly dominated by meaningless controls rather than meaningful dialogue. Such a dialogue should take account not just of innovations but also of the conditions under which they are generated, and this could help to mutually determine the terms and conditions of scholarship.

3 A Brief History of Research Assessment in the Context of Quality Assurance

Since the 1980s, research, first in English-speaking countries and the Netherlands and later throughout continental Europe, has been increasingly subjected to comprehensive review with the aim of increasing efficiency and enhancing achievement. This was seen as necessary for a variety of reasons:

1. Expanded reporting obligations and performance-based steering instruments were introduced to the science sector in conjunction with increased university autonomy and the introduction of New Public Management. This took place first in the English-speaking world, then in the Netherlands, and, since the 1990s, in Switzerland as well. The assumption was that one could record, verify, and reward research accomplishments and thereby supplant governmental micromanagement. Michael Power's characterization of Great Britain as an "audit society" (Power 1997) described how all realms of society were being overrun with rituals of verification and evaluation. Resources were increasingly being allocated based on performance to ensure "value for money", "accountability" and dynamic performance. This model was gradually adopted in the rest of Europe. The withdrawal of government from micromanaging the daily life of institutions was compensated for by expanding more remote state control using new performance review, assessment and steering instruments. Governments shifted "from interventionist to 'evaluative' governance" (Ferlie 2012; see also Moscati 2012; Paradeise 2012; Paradeise et al. 2009). The increase in institutional autonomy went hand-in-hand with greater limitation of the individual autonomy of researchers and scholars inasmuch as their actions were subject to a growing number of external steering efforts (Moscati 2012; Pechar 2012).
2. In the wake of globalization, research became increasingly important in high-wage countries as the motor of economic development. Their competitive disadvantage brought about by high wage costs could only be compensated for through increased productivity, and hence through having an

edge in research and technology. The increasing reliance on the value or worth of scientific research, however, went hand-in-hand with decreasing confidence in the ability of science to govern itself: "Indeed, it is precisely because it [i.e., science] is such a significant policy object that its fortunes can no longer be entrusted to scientists alone. Macroeconomic strategy in many countries demands that science be made governable" (Power 2008).

3. The strong weight given to research and the emphasis placed on an achievement orientation is also reflected in the growing emphasis on competition. This can be seen particularly in the relative increase in competitively-acquired external funding as compared to research funded through the basic financing provided to the universities. In some countries, the share of external funding has doubled in only a decade, leading to a corresponding increase in the number of experts involved in providing evaluations.
4. With the rapid but under-financed expansion of the university sector and of mass university education, combined with an increased permeability of the educated elite, calls for greater internal differentiation among the universities became louder, first from users and the public, then from the bodies providing support or funding. Users, funders and other organizations aiding universities became increasingly interested in identifying and promoting particularly high-achieving individuals (both students and researchers). The situation in the universities called for reforms. Ubiquitous deficiencies led to performance reviews and to a more differentiated allocation of resources based on achievement. The altered parameters suggested a greater monitoring of successful performance was needed (Weingart 1995). Germany and Austria were far from systematically introducing evaluation methods in 1995, and limited the information they provided to the public and to science policy authorities much as they limited themselves to input indicators when resources were being allocated. Only a few years later, they had introduced, and promoted, the comparison of institutions based on output or performance indicators, and they were using systematic quality assurance instruments (Müller-Böling 1995). As a result, in one federal state after the next in Germany evaluation and structure commissions were introduced, with the aim to assess the achievements of the universities with the help of data and experts' surveys. From this information, decisions about restructuring and financing were to be derived. Within the universities, initial attempts were made to allocate resources based on performance and achievement.
5. In Switzerland, quality assurance and the corresponding evaluation instruments were introduced earlier than in other German-speaking countries. Thus, systematic teaching evaluations by students were introduced in most universities already by the early 1990s. Systematic bibliometric investigations into research quality were carried out by the Swiss Science Council since the 1980s. Evaluations of disciplines in institutions were tried first at the Federal Institutes of Technology (ETH), first in the Department of Mathematics in 1989 at the ETH Zurich, at the behest of its Board of Trustees, and then, beginning in the early 1990s, in all departments at the ETH Zurich and ETH Lausanne, by order of the university administration (*Handbuch zum Qualitätsmanagement*, ETH Zurich, 2003). Other quality assurance instruments were also introduced. The University of Zurich followed this model, creating its own quality assurance bureau in 2001 so as to carry out peer review evaluations and monitoring at all levels. Institutionalized quality assurance through evaluation was enshrined into law by 1991 at the ETH (by the Federal Act of 4 October 1991 on the ETH), and by 1998 at the University of Zurich (by the University of Zurich Law of 15 March 1998). Other universities and trustees followed the model of systematic institutional evaluation and also established coordination offices at their universities. Various boards of trustees began using output or performance indicators for allocating resources, along with global budgets coupled to agreed-upon goals, by the second half of the 1990s. At the national level, these have been used since 1999 based on the Federal Act on University Funding (UFG 1999, Art. 15). Since 2006, national guidelines, so to speak as a retroactive imprimatur, require all universities to conduct periodic internal evaluations (SUK 2006).
6. In the thicket of a university landscape oriented to mass education, prospective students and other in-

terested parties needed some orientation or help in making decisions. This is particularly relevant in identifying islands of high achievement where advantageous study conditions might be found. Thus, at least in Germany, the comparison of universities and their achievements became popular, particularly in the mass media, spearheaded by *Spiegel* newsmagazine surveys (1993) of how attractive various courses of study were. This was soon followed by comparative investigations by other magazines (*Focus, Bild der Wissenschaft*) of the research achievements of entire institutions (1993). Later, the Centre for Higher Education (CHE) introduced a method of ranking in order to make the choice of where to study easier, subsequently marketed in cooperation with the respected weekly newspaper *Die Zeit*. Globalization and increased mobility among undergraduates, graduates, and researchers supported the interest in international rankings, the most prominent of which began being published in 2004 in the *Times Higher Education Supplement* Ranking and the ARWU Ranking produced in Shanghai's Jiao Tong University. Today, we can look back on two decades of repeated national and international comparisons of institutions. These comparisons and rankings claim to help enlighten readers, but their methodological shortcomings mean they can only do so to a limited extent (Rauhvargers 2011).

7. The increase in evaluation activity in higher education institutions is also due to homogenizing quality assurance standards in the context of the Bologna reforms, to establishing common procedural standards for quality assurance through the European Standards and Guidelines for Quality Assurance, and to registering evaluation and accreditation agencies that meet certain standards in the European Quality Assurance Register for Higher Education (www.eqar.eu). Through the strengthened intra-European dialogue, in the context of the Bologna reforms, between actors involved in higher education policy, these evaluation practices – regarded as progressive – have spread that much more rapidly throughout Europe. The "Trends" studies examining the implementation of Bologna reforms in universities reflect the importance of quality assurance reforms in Europe: "Institutions find quality assurance reforms to have been among the most important developments that have shaped their strategy in recent years, since 63 % of the respondents to Trends mentioned it among the top three developments" (Loukkola & Zhang 2010, p. 12; EUA 2010). Most universities report they have established internal quality assurance systems in the last decade in response to demands from national quality or accreditation agencies (Loukkola & Zhang 2010, p. 23). Establishing external and internal quality assurance systems are, together with the new curricular structures, among the most important innovations the Bologna reforms have introduced. Both those with expertise in a discipline (in evaluations of that discipline) and management and quality control experts (in audits and accreditations) are called upon to serve as external evaluators.

The evaluation and performance review movement which emerged in the 1990s thanks to the aforementioned higher education policy contexts and the justificatory discourses associated with them, expanded not only throughout Europe but also into increasing spheres of activity. Fed "by a situation in which resources were stagnating, the reporting on the use of public monies for scientific research needed to be defended politically, and a concern about the dwindling ability to compete and reform conveyed by the mass media" (Hornbostel 2008, p. 74), evaluations and performance reviews expanded to encompass not just teaching and research but also transfer payments, internationalization, university cooperation with extramural institutions and enterprises, gender questions, marketing and visibility, and access to universities. The circle of those concerned with collecting and dissemination of information, additionally, also widened.

4 Research Evaluation Forms, Utilization and Stakeholders

The aforementioned rationales have led to a range of performance reviews and expert assessments that have considerably increased the overall cost and effort involved in reporting and assessing.

Among them, one may include performance assessments and appraisals inherent to the science system, including:

1. proposal and reporting requirements for, as well as appraisals in, the external funding of research projects (individual and joint funding); in recent years, these have also increased within universities in granting support for research;
2. appraising manuscripts for publication in scientific journals;
3. in appointment or hiring decisions at universities, both external and internal appraisals;
4. answering surveys which form part of studies of universities.

Performance reviews may also be demanded by actors external to the science system, including politicians, funders, or other stakeholders, coming in forms such as:

1. reports, including annual reporting for a university's administrators, for trustees or external funders, as part of performance controls linked to agreements on goals;
2. the evaluation of research units (departments or faculties) and research groups as part of prescribed institution-internal or nationwide assessments of disciplines;
3. collecting data on research output to be used, either internally by the university or externally by funders, as indicators to determine the allocation of resources;
4. collecting data on research output or on steering processes to be used by external accrediting or quality assurance agencies;
5. answering surveys for institutions which rank;
6. evaluating research programmes and their effectiveness;
7. evaluating research policy measures.

Of course data are also collected and evaluations carried out about teaching, sometimes also about knowledge transfer, so researchers and instructors are involved in collecting, reporting, and appraising in this area as well.

Today, evaluations take place at all levels. Entire institutions are subjected to audits and accreditation procedures; faculties and departments are evaluated so that resource allocations and strategic development plans can be justified; programmes of study and teaching modules are evaluated to improve teaching and ensure "studyability"; and individual researchers are evaluated so as to make decisions about project resources or hiring. Scholars are involved in all these evaluations, some of which eat up a considerable amount of time (Jacob & Teichler 2011).

A comprehensive overview of the various goals and users of research assessment, as well as their data needs, was provided by AUBR (Assessment of University-Based Research, 2010; see the table in the appendix), a group of experts appointed by the EU. It showed that there are data needs at all levels, and owing to the diversification of funding sources universities can tap (EUA 2008), these needs have been strongly increasing, particularly in and for funding research. There is an ever-increasing share of external funds in the financing of research, and a proliferation of sources from which funding comes, as well as the forms it takes. Acquiring research funds means an increasing set of obligations in terms of proposal writing, reporting, and assessment, to which one may add the evaluation or surveys of the achievement of research units themselves. The globalization of peer review, as an element of the evaluation of disciplines or of the assessments called for by organizations which fund research, has led in many European countries to a significant increase in the assessment obligations of international peers. They are increasingly called upon to take part in institutional or national evaluations and assessments, as recommended in the European Standards and Guidelines for Quality Assurance and by the European Association of Research Funding Organizations (EUROHORCs 2006).

The AUBR report indicates that the need for data varies not only by the goals and needs of stakeholders, but may strongly overlap between the user groups. The AUBR report and the most recent policy paper of the

League of European Research Universities (LERU 2012) hence emphasize that increasing diversity, in the context of expanded sources for financing universities, and the lack of coordination between needs for data and evaluation that are expressed, has meant that the reporting and assessment obligations among researchers and scholars have increased exponentially.

The increase in these expert assessments and evaluative activities in and around academic research that all can observe is meanwhile regarded with considerable concern, as they contribute to a considerable loss in time for other pursuits. In the view of many who work in academic settings, it has led more to a reduction than to an increase in efficiency. A recent, comparative study of the working conditions in academe in 20 OECD nations (*The Changing Academic Profession*) has underscored the disproportionate claims these evaluative activities make on researchers' time. Academics report spending between 23% and 45% of their time on self-governance, both of the university and of their discipline, which includes the time spent writing proposals, providing expert assessments, and reporting on research achievements (Jacob & Teichler 2011, p. 27). Of particular concern is that the most successful researchers are also those most called upon to serve as expert evaluators (as editors, proposal reviewers, the evaluators of research units), and are thus the most burdened. While nearly all scholars have reporting responsibilities of one kind or another, it is the most successful and renowned who are called upon to provide external evaluations of journal articles, professorial appointments, project proposals, and evaluations of the discipline; ironically, it is precisely the best researchers who have the least time left to conduct research – unless, through other means, they can be released from other obligations, including teaching.

5 The Problems with Research Assessment I: Pros and Cons of Various Research Assessment Methods

The assessment practices noted above should be critically evaluated, and in two directions:

— They can be critically examined in terms of making specific, pragmatic changes to current assessment practices, meaning in their selection, their frequency, or the context in which they are employed, as well as in their consequences and in how they interact.
— The underlying approach also can be more thoroughly scrutinized, particularly with respect to the culture of mistrust it engenders and its effects on the innovative drive in research endeavours.

In this section, we turn first to the pragmatic level before, in Chapter 6, examining assessment as a cultural practice, and its premises.

In pragmatic terms, discussions of research assessment and how appropriate it is range largely between how the two categories of methods are viewed. On the one hand, one has peer review, generally fairly well-established in science evaluation, and on the other, one finds the increasingly widespread use of quantitative performance indicators. Peer review is used primarily in qualitative, *ex ante* assessments of research project proposals, for articles submitted for publication in journals, and in the hiring and appointment of professors. Peer review is also used in the formative evaluation of researchers and their research units, as a means to provide advice or a monitoring of how they develop. In such evaluations, *ex post* appraisals merge into *ex ante* assessments. Quantitative indicators, by contrast, are always *ex post* assessments of previous achievements, and are used primarily to compare institutions or researcher groups. The pros and cons of both types of research assessment are discussed, in brief, below.

5.1 Peer Review

Evaluations carried out by colleagues in the discipline (peer review) have increased in importance, not just in response to demands from external stakeholders but also within the science system itself, as part of how it regulates itself. This increased significance is partly due to the growing share of external funding, competitively assessed through peer review. The demand for peer review also has increased due to the global competition for well-qualified scientists, for visibility and for resources; having peer-reviewed publications in international scientific journals, which continue to increase in importance, is a significant marker in this competition. Publications and external funding, both of which depend on peer review, are becoming more and more important as indicators for allocating resources, and this also drives the increase. Finally, peer review is a fixed aspect of evaluating research units and entire research centres, and at most institutions such evaluations are now carried out in periodic cycles.
The advantages of peer review have long been known. The method has been used since the 17th century, originally by the Royal Society in an effort, at a time when modern natural sciences were expanding, to stem the flood of observations, experiments, and theories that were of unknown validity (Hornbostel 2008). Among the advantages of peer review is that it strengthens the self-regulating capacity of a science system by having researchers evaluate one another as equals. Decisive here is an assumption that the intellectual level is high, but above all that the assessing peers have disciplinary expertise. A consensus also exists that peer review is the best possible form of external assessment of one's own research achievements. As long as the scope and effort involved in peer review does not get out of hand and the selection of expert evaluators, or the focus of the questions and topics of the evaluation, are felt to be appropriate, peer review is a widely accepted method of research assessment.
But despite its recognized role as a self-regulating mechanism, many critical objections have been raised against peer review. They rest on observations of those affected (both evaluators and the evaluated) by the expansion in its use, and are based in part on empirical studies (thus, see for example Peters & Ceci 1982; Moxham & Anderson 1992; Daniel 1993; Röbbecke & Simon 2001; Wenneras & Wold 1999; Lamont 2009). The reliability of peer review is criticized in several respects:

— The various judgements of peers often contradict one another. The most sensational proof of a lack of intersubjective reliability was provided by a field experiment conducted by Peters & Ceci. They submitted articles by well-known researchers – now under different names and addresses – to the same journals that had published these researchers' work 18 to 32 months previously. Less than 20 % of the articles were accepted again for publication (Peters & Ceci 1982). Given this evidence, one can hardly speak of the reproducibility of peer review judgements.
— Also problematic is the peer evaluation of emerging areas of study. Given the considerable increase in recent years of such new areas of study, as well as a growing interdisciplinarity, it is not easy for peers to come to valid judgements about new and interdisciplinary research fields (van Raan 1995, p. 88). This is not least because expert evaluators are often selected based on their accomplishments in a particular discipline, and then selected for discipline-based panels, thus reinforcing what is mainstream in the discipline (Lamont 2009; Martin & Whitley 2009). This is underscored by the many peer misjudgements of path-breaking research by various Nobel Prize winners or other prominent scientists, with examples including misjudgements about Enrico Fermi, Sir Hans Krebs, Rosalyn Yalow, Gerd Binnig, Hans Rohrer, Alfred Wegener, Alan Turing, Konrad Zuse, Hermann Oberth, Francis Peyton Rous, Mitchell Feigenbaum, Frank Rosenblatt, Stanley Prusiner, Günter Blobel, Andrei Linde, Noam Chomsky or Karl Popper (list assembled by Fischer 2003, p. 29).
— Conflicts of interest among expert evaluators can also lead to unfair or even false assessments.
— It has also been observed that peers more readily overlook the achievements of younger scientists and often have a negative bias against newcomers in the discipline (Fischer 2003).
— Some empirical studies on the effects peer review has on proposal submission and publication behaviour have observed mainstreaming effects and

a tendency to avoid riskier research or research that challenges paradigms. As a result, since there is increasing emphasis on being published in peer reviewed scientific journals, it is more and more the case that less risky research is being undertaken (Henkel 1999; Horrobin 1990; Lee 2007; Overbeke & Wager 2003).

Peer review, as a basis for evaluating individual scientists, is considerably more reliable as a way to assess weakness than strength in performance or achievement. In addition, the risk of misjudgement is that much greater the larger the unit to be assessed is, and correspondingly, expert evaluators lack sufficient unit-appropriate knowledge. The smaller the focus of an evaluation, however, the greater the risk of disciplinary tunnel vision and hence of misjudging newly emerging fields.

A suggested corrective for subjectivity problems is to embed the judgements of individual evaluators within an assessment by a larger group of evaluators. However, empirical studies indicate that the group barely acts in a corrective manner. Both Langfeldt (2004) and Olbrecht & Klein (2011) have observed that decision-making in a group of evaluators is marked by a clear division of tasks. There is little interaction between experts unless areas of expertise overlapped. Langfeldt hence demanded that: "Two experts assessing each object under review and some time for discussing the results would be the minimum needed if expert panel evaluations are to have some function exceeding individual review reports, when it comes to assessing the quality of research" (Langfeldt 2004, p. 60).

Investigations of cognitive and epistemological limits to the reliability of peer review assessments have focused on the individual and psychological circumstances of the individuals doing the assessing. To this, in recent years, have come sociological analyses of peer review as intersubjective practice. The work of Michèle Lamont, in particular, has drawn attention to the fact that peer review may also occur in teams or in the context of scientific communities and can thus be seen as communal practice. That makes it subject to common norms, behavioural rules, and intersubjective conventions that go beyond individual stances or prejudices and that legitimate the group's conclusions. She investigates "how panelists justify their judgments, but also how the processes and rules of practice set the stage for a sense of legitimacy" (Lamont 2009, p. 246). The way peer review functions as a core evaluative practice is precisely as a tension between the positions taken by individuals and the consensus of the peer panel. Evaluation is also always culturally embedded. Individual and intersubjective prejudices are a basic part of qualitative evaluation, so one should not expect universal rationality in evaluation results or expect a corresponding form of meritocratic fairness.

"What is presented as expertise may sometimes be merely preference ('taste'), described in depersonalized language. The reciprocal recognition of authority is central to the process, but it may lead to explicit horse-trading, which produces suboptimal results. Despite these potential hazards, however, panelists think the process works, in part because they adopt a pragmatic conception of 'truth' (or at least of what constitutes a 'fair evaluation') as something inevitably provisional and defined by the best standards of the community at the time" (Lamont 2009, p. 240).

In Lamont's view, the positive function and legitimacy of peer review in the science system is not thereby questioned, but instead the exaggerated expectations of what it can bear (Lamont 2009, p. 241).

Empirical investigations of the consequences of evaluation processes have only been carried out in a few European countries, and when they are, then primarily on evaluation processes related to teaching. Some of these are of longer standing, as, for example, in Belgium (Hulpiau & Waeytens 2003), Denmark (Bjørnkilde & Bason 2000; Thune & Kristoffersen 1999; The Danish Evaluation Institute 2004), Great Britain (Brennan, Frederiks & Shah 1997; Brennan, Shah & Williams 1996) and in the Netherlands (Frederiks, Westerheijden & Weusthof 1994; Jeliazkova 2002). Studies of the acceptance of research evaluation among those evaluated and those evaluating have only been carried out by the Scientific Commission of Lower Saxony (WKN), to a limited extent by the German Science and Humanities Council in evaluating its pilot project on "rating" research, and by various actors in Great Britain as part of the discussions about the Research Assessment Exercise (RAE) and the Research Excellence Framework (REF) superseding it. Catherine Paradeise, in a talk given in Lausanne, has

also presented some preliminary and unpublished results of her work on the acceptance of evaluation results (Paradeise 2011).

The British RAE is the most comprehensive, systematic peer review-based evaluation of research in Europe, and certainly also the most far-reaching in terms of its fiscal and system consequences. Its effects on the British research system have been thoroughly investigated, though largely in terms of changed governance structures, power relations, institutional differentiation, and relative costs. The most recent evaluation of the RAE (PA Consulting Group 2009) did not ask about the value it has added to the quality of research. Instead, it only inquired about the value added in terms of institutional and research group visibility, as well as their ability to acquire more resources. In this evaluation of the RAE, no information was provided about its effect on research topics chosen, their orientation, or the quality of publications. More recent studies, such as by Martin & Whitley (2010), have investigated the effects the RAE has had on authority relations, and argue it has strengthened governmental power *vis-à-vis* the universities via the funding councils and the allocation of resources. It has also strengthened scientific elites and traditional disciplinary boundaries, as they are reinforced by how the evaluating panels are organized (Martin & Whitley 2010, p. 65). Thus, in selecting expert evaluators, those who have been successful in their discipline are favoured, which means that "the RAE has reinforced the emphasis on conventional mainstream research, discouraging new developments and interdisciplinary research" (Martin & Whitley 2010, p. 65). By emphasizing peer review, these gatekeepers become the "arbiters of 'excellence'" and influence not only scientific production but over the middle term, even the selection of scientists themselves, as Lee & Harley showed (1998). The effects of evaluation on the quality of research outputs is otherwise not the subject of empirical investigation.

The Scientific Commission of Lower Saxony evaluated research undertaken in all the universities in this German state, and in one of the few empirical investigations of its kind, subsequently asked the evaluated researchers and the expert evaluators for an assessment of the evaluation itself. The judgement of the evaluation process and the utility of the evaluation tended to be critical among those who were being evaluated. Less than half (44 %) found the conversations conducted to have been effective, or that all the relevant strengths and weaknesses of the discipline had been addressed. Only 28 % judged the conversations to have been interesting and stimulating from an academic point of view, even though 72 % regarded the expert evaluators as respected representatives of their discipline. Only half regarded the experts' recommendations as sufficiently precisely formulated or as providing guidelines for action (in contrast to 90 % of the expert evaluators; see WKN 2006, pp. 19, 23). Only 43 % held the view that the recommendations would be accepted in the discipline (in contrast to 55 % of the expert evaluators). Even though 56 % of those evaluated felt it was important to implement the recommendations, a still larger percentage (62 %) felt there was insufficient financial or institutional support for actually doing so. 55 % also complained that it had not, or only barely, been possible to discuss the experts' report with the experts' commission itself. 48 % felt the costs of the evaluation did not correspond to its benefits. Attitudes about the optimal utilization of evaluations were correspondingly cautious among the evaluated. A little more than half (56 %) of those in Lower Saxony who had been evaluated using informed peer review regarded research evaluation as a suitable instrument for quality assurance. Nearly two-thirds (64 %) said that the results of an evaluation should be advisory only and not be used for allocating resources.

Noteworthy here is how much this appraisal diverges from that of the expert evaluators themselves, who were considerably less critical of the process and benefits of evaluation. In general, they regarded their own activities considerably more positively and ascribed considerably higher significance to their own recommendations. But they, too, critically noted that the diagnosis of the causes for variance in research performance was too weak, as that variance resulted from different framework conditions.

A similar discrepancy between the acceptance of evaluation by expert evaluators and scepticism on the part of those being assessed is also evident in the assessments of the process for evaluating teaching and courses of study by the Central Evaluation and Accreditation Agency Hannover (Germany) and the Associa-

tion of North German Universities. In a written survey, questions were asked about whether the evaluation process had proved itself, about whether the goal of quality assurance and quality improvement had been reached, about the general satisfaction with how the evaluation had been conducted, whether the relationship between the effort expended in the evaluation process and the results of the evaluation was appropriate, and the personal benefit obtained through participating in the evaluation. The overall appraisal of those evaluated was generally from 13 % to 32 % lower than the assessment by the expert evaluators. However, the results were generally more positive than in the abovementioned multilevel research evaluation in Lower Saxony which was conducted at the same time (Mittag, Bornemann & Daniel 2006).

According to Catherine Paradeise (2011), variability in the acceptance of evaluations depends not only on the use to which they are put but also on the nature of a researcher's engagement in the university and disciplinary community, and what was expected of these institutions. Based on a survey of evaluated researchers, Paradeise developed a typology of evaluation reception: it is decisive how weak or strong the identification of an academic is with the institution or the disciplinary community. The expectations placed on the institutions resulting from this identification determine how research evaluation is dealt with or accepted.

Given its limited reliability and its potential to produce unintended effects, the context of its use leads to quite different judgements about peer review. While it is sometimes criticized when used as performance assessment, it is generally judged more positively when it is part of a formative evaluation in which the potential of a research unit to develop is gauged.

Irrespective of the context it is employed in, and in light of the growing number, variety, and international diffusion of peer review, there are complaints about the overuse of a growing number of expert evaluators. Selecting them has become increasingly difficult. There is also critique of the falling marginal utility of evaluations the more routinely they are used (Hornbostel 2008). Wilhelm Krull (2011) warns of the dangers in escalating the frequency of evaluations: then even evaluations designed to be formative lose their character and degenerate into "a ritual drained of meaning".

What suggests itself in this context is to align evaluation practices in terms of scope, focus, frequency and intensity with previous evaluation experiences and in light of the track record of a research unit.

5.2 Research Assessment Using Quantitative Data or Indicators

Because of the considerable effort and time qualitative peer review assessment calls for, both from those evaluating and those being evaluated, quantitative indicators seem an attractive alternative to fall back on; it seems an easier way to review performance. In contrast to the qualitative data generated by peer review, quantitative data on research – understood at least in part as indicators of research quality – reputedly require much less effort. It was not for nothing that heated discussions surrounding reforming the RAE were triggered by the Chancellor of the Exchequer's concern about its costs. Quantitative methods of assessing research also have the advantage, at least to steering organizations outside of academe, of permitting an externalization of the assessment processes. Assessments of content are excluded. Using data produced by the science system itself, some the result of internal peer review, autonomous self-determination processes are encroached upon from without through decisions to allocate resources that are based on quantitative measures of success.

Thanks to their supposed objectivity, the hope is that quantitative indicators will allow for comparisons to be made between different research contexts (van Raan 1995, 2005). As a result, such indicators are used particularly when a large number of institutions or research groups are being compared to one another, such as when resources are being allocated, in indicator-based performance assessments, or in the comparative ranking of institutions. They are also used when data that is as objective as possible serves as background information to augment other intersubjective appraisals (as by expert evaluators), or when research sites are to be compared (as in the RAE or REF in Great Britain). The present norm in most European countries is to use indicator-based methods for resource allocation both in and for universities, though the selection and weighting of the indicators employed provoke fierce controversies.

The critique of the use of quantitative indicators in performance assessment, and especially in rankings or resource allocation, is directed in part against the significant distortion they create in the assessment of accomplishments, and in part at the undesirable behavioural accommodations they provoke among researchers.

Distorted appraisals are generated by reducing the multi-dimensionality of university activities to measurable research accomplishments. That adequate data on research is not available or comparable enough remains one of the biggest challenges in research assessment (AUBR 2010; LERU 2012). Data is frequently either completely unavailable or, as it is generated by surveys or analyses within universities, available only at considerable effort or in forms that are particularly context-dependent. This makes cross-university or cross-national comparisons impossible. Or it is possible, but only with an immense effort at data collection on the part of the universities that stands in no relation to the utility of the comparison, as in the EU-sponsored Multirank Project. The added effort to collect the data needs to be offset by the added value the information brings. Yet, if one confines oneself just to available and comparable datasets, one receives a distorted picture, as important research outputs are not recorded or included. So while bibliometric data about peer-reviewed journal articles are available, including commercially, other research results or publications (such as conference proceedings or monographs) which may be more important in a given discipline, are not compiled internationally and cannot be identified using database search functions.

The prioritizing of publication-intensive disciplines reliant on external funding is also criticized, as their accomplishments are more likely to be reflected in internationally recognized journals and in competitively-awarded external funds. Even specialists in citation studies like van Raan concede that bibliometric indicators may not have the same explanatory power in all disciplines: "Bibliometric indicators are very well applicable in the natural and life sciences. However, in the applied and engineering sciences as well as in the social and behavioural sciences (and even more in the humanities) international journals are often not the primary communication channel" (van Raan 2005, p. 9). Bibliometry has refined its methods for surveying publications in these other disciplines, but still cannot claim to have developed a satisfactory method for fully capturing research output. If one relied only, or too much, on quantitative indicators, the result in the

middle to long term would be the downsizing, if not dismantling, of the humanities and social sciences. This tendency has begun to make itself noticeable in a number of European countries, though this has also been for other reasons (Henkel 1999; LERU 2012).

A third significant objection to the excessive weight accorded quantitative performance indicators is related to their focus on research accomplishments. This leads to a creeping, long-term devaluation of the quality of teaching, which is less measurable. Despite the OECD's AHELO project, no nationally or internationally comparative data exists on the competencies of graduates or on their professional success.

The strong emphasis placed on performance indicators in appointments and in the internal allocation of resources, particularly regarding external funding and publications, has been shown to lead to undesirable behavioural accommodations. This can be seen, for example, in the selection and planning of academic research projects, but it can also be seen in the avoidance of projects whose results are open or unclear and hence of uncertain publishability. It is reputed to engender changes in publication practices, such as fragmenting research results into the largest possible number of individual (and "smallest publishable") units (Henkel 1999; Weingart 2005a). On behalf of the SSTC (*"Zur Quantifizierung wissenschaftlicher Leistungen in Forschungsevaluation and Hochschulrankings"*, August 2010, SSTC; the document is internal and unpublished), the present author has already completed an overview of studies about the effects of quantified research evaluation.

The use of specific indicators is also criticized. Thus, in gathering and using bibliometric performance indicators that measure productivity and research impact using citations, popular due to a belief in their explanatory power, certain conditions must be met so that how the data are presented does not lead to incorrect analyses and distorted perceptions:

— The level of aggregation cannot be too low.
— Research areas cannot be compared based on absolute frequency of citation; understanding the norms specific to one research area is precondition for comparing it to other areas.
— All authors need to be taken into account and their contribution accordingly weighted, and names and addresses need to be standardized.

One also needs to bear in mind that citations, which are regarded as the major indicators of research impact and thereby are also indirectly seen as indicators of quality, do not actually reflect the quality of research. An article which reviews an area of research may be particularly frequently cited even though it makes no claims to originality or to quality. Conference proceedings are another example.

Objections have been raised as well to using the external funds a research unit or individual acquires as an indicator of research intensity or competitive success. In answer, using external funding as an indicator is justified by the argument that such funding makes research projects possible in the first place: more external funding reflects more research projects. Peer review is also cited in this context, since at least for external funding from public sources, what is decisive is being successful in the competition between various applicants. External funding thereby becomes an indirect indicator of the recognition or acclaim of peers.

Also a cause for critique has been the observation that it is not equally meaningful in all disciplinary contexts to use the acquisition of external funding as a measure of research intensity. In the humanities, where individual achievement remains the standard, acquiring external funding is not necessarily beneficial for a researcher's own work. It is only in exceptional cases here that the support of doctoral students by acquiring external funding will actually contribute to a researcher's *own* work. Still, networking practices and joint initiatives have been increasing in the humanities. In the experimentally-oriented natural sciences, by contrast, where research work depends on the availability of (externally) financed young researchers as well as on the availability of appropriate scientific equipment, most research projects would not be possible without external funding. It is also misleading that some research institutes have generous internal budgets for research, making the acquisition of external funds not only superfluous but also making external funding not a genuine measure of research intensity.

What is noteworthy, however, is not just the limited explanatory power but the growing significance of external funding as a supposed indicator of academic success. In Germany, both its Research Foundation (DFG) and its Science and Humanities Council (Wissenschaftsrat) have watched with some concern how external funding

has grown. Between 1998 and 2010, basic financing increased by only 23% (from 12.6 to 15.5 billion Euro), while competitively-awarded external funding to universities increased by more than 100% in the same time period (from 2.5 to over 5.3 billion Euro). Thus, the "external funding share", meaning the share external funding contributes to the overall financing of the universities and research, has increased from 16% to 26% in only a decade. For many years, the "full costing" discussion in England has dominated science policy debates, as the increased share of external funding has largely eroded basic financing. There, it has become easier to apply for an expensive piece of new scientific equipment than repair a somewhat older one at a fraction of the cost.

The increasing share of external funding is worrisome primarily for two reasons:

— For one, this increasing share means demands for reviews rise, and still more time and effort on the part of expert reviewers. As these demands associated with externally funded projects and instruments increase, it becomes more and more difficult to find good (or willing) experts to carry these reviews out. Researchers nowadays say the time they can devote to any individual evaluation continues to decline, which in turn reduces the soundness of their assessment.

— For another, it is worrisome if the increase in external funding comes at the cost of a decrease in institutional support, because it means a creeping erosion of ongoing appropriations for infrastructure, personnel, equipment maintenance, repair and overhaul. This trend is due not only to the dependence on external resources but also because institutions increasingly need to contribute their own funding to externally supported projects. In this context, a recent iFQ study of the self-perception of professors is revealing. They said they competed for external funding because they otherwise could not address research questions or hire research staff, and because success at acquiring it affected the allocation of basic financing, at least in the context of performance-oriented fund allocation that is now practised at many German universities. When asked which measures would genuinely help in promoting research excellence, they said: secure career prospects and a basic financing independent of research performance (Böhmer et al. 2011).

5.3 Is There Consensus that "Informed Peer Review" Is a Solution?

To prevent potential distortion when using quantitative indicators, and to adequately take account of contexts, quantitative indicators of research accomplishments "need to be embedded in a systematic reflection about the nature of research […]", Peter Weingart noted already in 1995, adding that the right direction seemed to lie in "developing and using a combination of external indicators and internal evaluating boards" (Weingart 1995). At the time, the bibliometry specialist van Raan was also arguing that the more telling bibliometric indicators should be used primarily as an aid to peer review processes; despite its drawbacks, peer review was still the best method for evaluating research (van Raan 1995). While Weingart was then referring to models used by the American National Science Board and the Swiss Science Council, one can point today to the research evaluation practices of the British RAE and the REF, its successor, to Dutch reports about research fields, or to Australian and French comparisons of disciplines. After heated debates, particularly in England and Australia, for or against indicators as opposed to peer review, "informed peer review" seems to have established itself. It has become the compromise, fifteen years later, between the two major research assessment methods, and the basis for forming a consensus about the standard way research evaluations should be conducted (AUBR 2010). While quantitative indicators provide an initial basis for asking questions, and a possible corrective to peer bias, peer review itself provides an assessment of research potential by those who are experts in the discipline.

However, while informed peer review or quantitatively-based peer review may be seen as attempts to mitigate potential distortions in assessing research achievements, they are not solutions to all the aforementioned problems in assessment methods. The core problems of peer review remain, and have been exacerbated by new assessment contexts and goals:

For one, due to an increasing number and widened scope of evaluations, which ever more frequently apply not just to individual research proposals but to entire research units, peers increasingly rely on time-

saving overviews of output data just to be able to cope with the sheer quantity of individual pieces of information (Torka 2011; Kieser 2010). The distorting effects associated with indicators thereby increase.
For another, peers increasingly are asked to address objects of evaluation lying outside their particular areas of disciplinary expertise, either because what is to be assessed is organized in an interdisciplinary or multidisciplinary manner, or because organizational parameters are increasingly part of the evaluation. This latter aspect is precisely what German science policy increasingly wants, as the latest developments in the excellence initiative or the "rating" of research efforts launched by the German Science and Humanities Council indicates. From the improvement of the conditions up-and-coming researchers face, through the support provided for innovation and technology transfer, and to the strategic steering of departments: peers are increasingly asked to address issues that lie outside their own disciplinary briefs. Even when they bring the knowledge they themselves have gained over time, their authority with respect to such matters is potentially more disputable than their research authority in their area of expertise. Inasmuch as the aim is a genuine assessment rather than a dialogue about perspectives, evaluation would actually need to be conducted only by a small group of experts (as is usual in the classic form of experts evaluating individual project proposals) in order that the process really be a review by "peers". A genuine expertise, in which experts are able to judge other experts because they survey the same research landscape, would have to limit itself to that comparatively narrow field or at least to closely related specializations. This, however, is only true for assessments of research proposals and manuscripts submitted for journal publication. Even then, peer assessment can in some cases reinforce the aforementioned mainstreaming effects and thereby underestimate the potential of newly proposed research avenues.

5.4 Implications

The solution to this dilemma – of not being able to find a genuinely satisfying, sustainable way to lay out accomplishments or performance even when using informed peer review – therefore lies in modifying or adapting the frequency, form, and objectives of the evaluation process. Even if the assessment of achievement is subject to a high margin of error, evaluations can nevertheless contribute to quality improvement. However, this is only possible under particular communicative conditions in which open dialogue, and the disclosure of weaknesses forms part of an effort to rectify these weaknesses and play to one's strengths, is successful. Less routinized, meaning less frequent, evaluations emphasizing questions a unit actually asks itself hold out the promise of a more stimulating evaluation process that is genuinely oriented to improving quality, particularly for oft-scrutinized research units. Evaluations "at the level of entire disciplines or institutions should be used sparingly" and be conceived in a manner "related to the grounds for the evaluation and the institutional structure in which the evaluation is carried out" (Krull 2011).
It is especially important that a genuine dialogue can take place between the evaluators and the evaluated, and one in which the analysis of strengths and weaknesses allows for, or serves as, an appraisal of future potential. To permit an evaluation to generate "productive insecurity" or an "impulse to correct a path taken" (Matthies & Simon 2008, p. 340; Wiesenthal 2008, p. 324), evaluation procedures need to be decoupled from "gratification and sanctioning measures" (Matthies & Simon, loc. cit.).

6 The Problems with Research Assessment II: Between Achievement Orientation and a Culture of Mistrust

Fundamentally, the multiplication of assessment practices in research contexts and within the system of higher education, along with their premises, can also be examined as cultural and normative practices. The first question is what assumptions current performance assessment practices make about how research functions, what drives it, as well as what is implicit in performance incentives that are crafted and to what extent these hit the mark. The new manner in which governments and universities interact reflects changed assumptions, which should be critically examined, about the meaning and purpose of research as a human enterprise, as well as the role it plays in society and its utility.

6.1 Performance Incentives in Academe

What is new about the practices noted above, through which academic research is to prove, or at least be able to assess, its ability to adapt or perform? Peer review has been with us as long as established scientific research practices have existed. The assessment of research projects has long been part of everyday academic practice as well. What is new is not just the increase in what needs to be evaluated (due to changes in the nature of project funding) and the larger size of the units to be evaluated (entire research units and institutions), but the sheer ubiquity of performance reviews. They encompass all aspects of everyday research practice and academic life, and with cyclical regularity, every unit of a science system is scrutinized to establish whether its performance is adequate. This is often as part of an artificially generated competition: "There was an expectation that efficiency would be increased through the mobilization of 'entrepreneurial' energies and activities once political steering and controlling claims were reduced" (Hornbostel & Schelling 2011, p. 59). Yet, this stimulation has now degenerated into a culture of verification in which accomplishments only carry weight if, or when, they have been evaluated and measured. In the end, it is not the achievement itself that matters but rather the ability of a unit to verify its own achievements: "Management attention shifts from individual scientists as experts to the capacity of scientific organizations to self-regulate" (Power 2008). Michael Power's characterization of British society as an "Audit Society" that without scruples blankets everything with ritualized verifications and evaluations (Power 1997), is becoming more and more true of continental Europe. Even if performance can be increased this way, one must ask whether genuine stimulation has come about – or whether the aforementioned misdirections and unintended accommodation effects have not instead been generated (Henkel 1999; Frey 2006; Kieser 2010). Productivity may have gone up, but at the cost of discouraging innovative research.

One should also question whether the constant assessment of performance is an appropriate way to promote achievement: is it really suited as an incentive? A core justification for systematically evaluating

researchers and research units is to reward good performance. To what extent do such incentives, whether by allocating more resources or by other types of reward, in fact influence researcher performance?

In this context, one should take note of empirical studies carried out over the last decades in organizational and social psychology that have examined the effects incentives have on creativity.

In her monograph, based on empirical investigations, Teresa Amabile examined the social psychology of creativity (Amabile 1996). She found that evaluation (even including the expectation of an evaluation), along with other "constraints" that are perceived as controlling, can have negative effects on the creativity of adults (and children). In later studies, she differentiated this finding, noting that it could depend on the characteristics of the person or the form of the evaluation. Thus, the "prospects of threatening critical evaluation" were correlated with verifiably lower creativity, while "informative, constructive feed-back and evaluation that conveys positive recognition of creative work often co-occur with high levels of creativity". Thus, not every form of performance assessment inhibits creativity, just those that are perceived as controlling, scrutinizing, and restrictive:

"[…] in real world settings, creativity may be undermined by evaluation that conveys incompetence or threatens self-determination, but creativity may be supported by evaluation that is work-focused and constructive (even when negative), that provides information about competence improvement, or that conveys positive recognition of competence and valued work. These positive effects may be due to the motivational synergy mechanisms (extrinsics in service of intrinsics)" (Amabile 1996, p. 152).

Bruno Frey, in a similarly differentiated manner, has examined the effect of evaluations, and distinguishes between a reinforcement and a displacement (or crowding-out) effect on the intrinsic motivations of researchers (Frey 1997b; Frey 2006). The displacement effect, also empirically investigated by psychologists, argues that intrinsic motivation is displaced by extrinsic motivation, such as that provided by external monetary incentives (Deci, Koestner & Ryan 1999). This always occurs when external intervention is seen as controlling. Limiting the freedom to act and limiting autonomy through the non-acknowledgment of intrinsic motivations can even lead to a reduction in self-esteem (Frey 1997a, pp. 23–25). Empirical studies suggest that making rewards for work depend on performance may even have a negative effect on motivation and performance itself (Fehr & Falk 1999). Limits on the freedom to act have particularly deleterious effects on the creativity of intrinsically motivated actors: "The characteristic most likely to kill creativity is not inadequate pay or tight deadlines, but a lack of freedom in deciding what to do or how to accomplish a task, lack of sense of control over one's own work and ideas" (Kohn 1993). Incentives that are aimed at increasing creative freedom in future projects, by contrast, act to strengthen motivation: "The creativity of work groups within organizations correlates positively with the degree to which members of those work groups report feeling that they have freedom in their work, a sense of autonomy and control over their own work and their own ideas" (Amabile 1996, p. 177). The exact interplay between extrinsic and intrinsic motivation, and the conditions which mutually suppress or enhance them, have only recently become objects of empirical study, so specific suggestions for institutions can be expected in the future.

What is clear is that market-based notions of increased efficiency brought about through performance-based incentives, sanctions, and regular performance assessments that serve as a basis for allocating resources, fail to address the socio-psychological realities and conditions of academic life. That life is characterized by a large number of intrinsically motivated actors and creative activities which do not follow strict specifications. Precisely these activities, however, are ill-suited to external and controlling steering mechanisms. Only evaluations that address the self-selected interests and issues of those being evaluated, and evaluations that are not tied to consequences regarded as controlling, have a chance of having a performance-enhancing effect.

What does this mean for steering science systems?

First, it means achievement incentives need to be conceptualized in a manner that they strengthen creative freedom and individual responsibility. Otherwise external incentives will have hardly any positive influence on creative research achievements even when they result in increased productivity, at least in the sense of accommodating to external steering incentives. In

fact, this corresponds to current arguments against the RAE. It has brought unquestioned increases in productivity, yet it is criticized for encouraging an increasingly short-term perspective in research, a diminished willingness on the part of researchers to take risks, and for leading to decreased innovative research (Evidence Ltd 2005; Henkel 1999; May 2004).

A second conclusion is that steering should not focus on trying to influence established actors already in the system but should instead focus on *ex ante* decisions, meaning on the selection and education of academic researchers (Frey & Osterloh 2006; Krull 2011). Such appraisals, long the basis for elite institutions such as the American Ivy League universities, the Swiss Federal Institutions of Technology, or the German Max Planck Institutes, seem to slowly be gaining in popularity, as witnessed by Germany's Excellence Initiative. No evaluation can replace such *ex ante* efforts to ensure quality. Requiring a qualitatively demanding education and rigorously selecting academic researchers would render *ex post* controls, in the sense of examining performance adequacy, largely superfluous. Maximal autonomy and creative freedom are already granted in these research contexts, though the trust of those who disburse public monies in the track records of their researchers is not large enough for them to feel they can do without evaluations.

One should note the objection that replacing *ex post* control by *ex ante* selection only makes sense in the case of top researchers and excellence-oriented institutions. Yet, this is not necessarily the case, since it is precisely *ex ante* selection that can be tailored to the differing goals an institution has, and thereby be oriented to different criteria. Other goals can be just as decisive for the success of a national science system as asserting oneself in the international competition for the best researchers, including generating innovations in economy and society, and contributing to the education and training both of future citizens and of highly qualified employees. A science system is noteworthy, after all, not just for the achievements of its top performers but also for its breadth and diversity. Researchers and scientists are needed not just for top-ranked, internationally competitive research but also supply numerous other important markets with training, research, and services, and these may well call for other selection criteria.

Beyond the necessary diversity in goals set and approaches taken, it is also unavoidable that there is a certain variation in the quality of research. Bruno Frey, who favours the model of *ex ante* selection, concedes that misjudgements in selection are unavoidable. Although unsatisfactory and irritating, misjudgements are nevertheless easier to bear than the transaction costs and cultural effects of *ex post* controls. It is not only a question of individual "errors", but also unavoidable variability in the distribution of achievements in any system. To this one should count the broad mass of good achievements which nevertheless cannot be counted as a part of top-ranked international performance: "A broad fundament of qualitatively high-grade research work is essential and the basis for peak performance in the science system" (Wissenschaftsrat 2011). The so-called 80/20 Pareto principle by which 80% of the achievements are provided by only 20% of the actors may be an exaggeration, but a degree of unequal distribution, both quantitative and qualitative, is part of every performance system regardless of how high-achieving it is overall. For that reason, evidence of less high-achieving individuals or units should not lead to a general mistrust of the ability to perform, or to raise doubts about the condition of the system as a whole.

6.2 Between Benefits and Freedom: Do We Need a New Social Contract for Academic Research?

If trust in the intrinsic motivation of researchers and creative freedom are important conditions for their creative energy, under which conditions are government funders in a position to win such a trust – and thereby abstain from elaborate reporting and from using controlling instruments that harm the culture of research? Under which conditions, and with which expectations, can funders justify abstaining from certain forms of control? What kind of communication between researchers and the public needs to exist to allow such trust?

Those who study research at universities see the rise of new performance assessment practices as part of the creation of new models of governance that call for a completely altered understanding of academic research (and of its institutions in their social contexts) but also of the role of government *vis-à-vis* the research which it helps promote (Power 2008; Neave 2012). Four profound changes are of decisive importance here:

1. *Academic research increasingly justifies itself by its economic utility.* By emphasizing how important its contributions are to the "knowledge economy", academic research makes its accomplishments, and the institutions in which it occurs, more central to politics and to the economy. At the same time, those accomplishments are defined more and more by their utility for the economy. In so doing, contributions private institutions make and the interplay between academic research and the private sector is both promoted and desired. But the rhetoric of utility obscures the other contributions academic research makes, including to the drive to investigate the human condition and the world without constraints, to ask about the reason for life along with its physical and historical conditionality, and to examine social practices with an eye to imagining other forms of coexistence. The new discourse about "value added" tends to be one focused on its material aspects. The adaptive capacity of academic research instead should be measured by its ability to respond to the needs and demands of an economy oriented to "adding value". The distance of the university from economy and society used to be what increased the university's value. Today, it is instead its close cooperation.
2. *Academic research institutions are losing their special status.* The new proximity of research to external partners dissolves the boundaries between academic institutions devoted solely to research and research-based commercial enterprises. Purely research-oriented institutions like universities thereby increasingly lose their special status ("Mode 2 Science", Nowotny, Scott & Gibbons 2001) and become increasingly subject to performance criteria and steering processes of other institutions.
3. *Academic research, and its associated institutions, no longer purely serve the public interest.* The dovetailing of academic research interests with the needs of external and often private interests or partners leads to creating numerous and new or mixed forms lying somewhere between public and private concerns. Public-private partnerships are called for, intellectual property rights are secured, and strategies are developed to create "research clusters" where universities, working together with private actors, pursue common goals and create infrastructures. Endowed professorships, funding from private and external sources, and services that promote innovation and fundraising have become symbols of progressiveness at universities. With this, universities and their researchers no longer act purely in the public interest. The mistrust of "distortions generated by competition" and of private benefits reaped under the aegis of public institutions grows as university and research funding comes increasingly from private sources.
4. *Academic research institutions gain greater autonomy.* As institutional autonomy increases, governments lose some of their control over, as well as access to, universities as institutions and along with it, the daily identification with its concerns. At the same time, the now more autonomous universities search for new methods of control and organization, and initially look to models from other institutional contexts. Often, these are also demanded by public authorities. The development of self-governance methods that are compatible with academic research and internal leadership is far from com-

plete, as older models of consensus-oriented collegiality cannot simply be abolished but instead must be adapted to demands for new decision-making formats.

The more important academic research becomes for society, and the more it permeates all realms of life, the less it will be granted its own, protected space. Ironically, its very omnipresence and growing significance have dethroned it at the same time. Even the traces of the old image of an almost monastic space for the contemplative life, with its material deprivations, rites, forms, obligations and freedoms have become blurred, and with it, the appreciation of the special conditions needed for academic work. In light of the developments sketched in Chapter 4, the suspicion is that the baby is being thrown out with the bathwater: given the emphasis placed on opening academic research to social and economic needs, its own goals and parameters are being lost from sight. The value of having a space to conduct pure research, to freely contemplate, or to test received wisdom and assumptions without constraint, rests precisely in the indeterminacy about what will result. This also means such endeavours wholly evade, or largely retreat, from economic cost-benefit calculations, short-term efficiency considerations, and from controls involving measurement. The uniqueness of the pursuit of knowledge in academic settings, with its alternation between painstaking, meticulous, detailed work, unforeseen setbacks, and sudden enthusiasms over new insights, is subject to its own motivations, rhythms, and rewards. It is barely possible to capture this using measurement tools employed in other realms – unless, of course, these tools were modelled on those used in measuring academic research.
Large companies have begun to staff and treat their R&D units differently than other parts of their enterprises. In much the same way, policymakers who finance academic research institutions, as well as the public, should understand, protect, and promote the particular parameters and unique manner in which such institutions function. The close dialogue between academic research and the economy that has triggered so much dynamism in the last decades, creating innovative accomplishments out of their cooperation, should not lead to misjudging the distinctive features of each. The increase in the autonomy of higher education institutions, meant primarily as an increase in flexibility and adaptive ability, and which has meant loosened control and less access on the part of government, should not be devalued by new forms of micromanagement. Instead, it should be in the vested interest of the public, and the public authorities, who finance academic research institutions to engage in a more intense dialogue about the contents, goals, values and products of the research which strongly influences our lives today and will continue to do so in the future. In the coming years, drafting these goals and values should be a key task of Swiss science policy.
This task strongly suggests itself for the Swiss science system for the following reasons:

1. Since science systems began to be compared, Switzerland has shown an especially strong ability to compete at the international level, and has a high density of research accomplishments and many strengths. As a result, it needs a close-woven network of controlling instrument less.
2. Its science system is far less underfinanced than many other science systems. There, advantageous resource allocations are subject to much larger needs for justification, since basic financing is not yet secured.
3. Evaluations have long been established in Switzerland. In that sense, it has reached a level of maturity in its quality assurance systems which suggests that the marginal utility of performance reviews from new evaluations will continue to decline.
4. The small size of the country allows for a density of communication that promotes trust between the public, the marketplace, and academic research, one which has repeatedly been outstandingly practised.
5. Switzerland has long fostered a culture of sustainability and, more so than many other industrially developed countries, dared to employ longer-term perspectives. Academic research here should find it easier to be understood as its own cultural practice, with its own long-term questions and problem definitions.

In that sense, it is conceivable that Switzerland could position itself internationally as a science system that strives to create optimal parameters for good research practice. To that end, a new dialogue between the sci-

ence system and society/politics/the public is needed to strengthen, and publicly emphasize, the conviction that trust is both an appropriate and efficient basis for the relationship between both worlds. The consistent *ex ante* selection of researchers, who are shown trust after having been initially scrutinized and evaluated, should replace the principle – which is built on mistrust – of evaluations that constantly control. Evaluation would then be correctly utilized, situationally rather than routinely, when it is structured in terms of the content of research, when it pursues formative goals, and when it incorporates the self-reflection that seeks out and fosters dialogue with the public and with the marketplace.

7 Conclusions

1. Over the last decade, in the perception of many scholars, the variety and frequency of research performance assessments have increased markedly. The result has been that a large, and increasing, amount of their time is now taken up with assessing or reviewing scholarly accomplishments instead of actually conducting research or transmitting it. The increase in reviewing achievements has come about in part because peer review is used in a growing share of the decisions made in the national and international competition for resources, in academic appointments and publication, as well as over matters related to reputation and visibility. It has also come about because of the ever more time-consuming reporting duties governments, funding organizations and the public demand.

2. Both types of assessment, meaning quantitative performance assessments using indicators and qualitative evaluation methods that draw on peer expertise, are used as the basis for performance-oriented steering instruments. The limitations of these methods, however, are barely mentioned in higher education policy discussions about increasing the use of achievement-oriented steering instruments. Yet, both types of assessment are heavily criticized for their methodological shortcomings, their lack of reliability in what they can assert about research quality, and their negative effects on the research behaviour of scholars. The limits of performance assessment methods are inadequately discussed and these limits are insufficiently clearly communicated.

3. The use of quantitative performance indicators for allocating resources, as well as for creating influential research rankings, distorts the image of research achievements. The explanatory power of quantitative assessment cannot be genuinely realized in the absence of contextual information and an understanding of which parameters are relevant to a given research or publication practice. In addition, use of such indicators leads to undesirable behavioural modifications on the part of researchers dependent on such assessments. The incentive effects quantitative performance indicators may have on increasing productivity, effects only seen when these indicators have large consequences for the allocation of resources, should not be misunderstood as an increase in the quality of what is being produced. To the extent to which effects on research quality have been empirically investigated, they have tended to be adverse.

4. The balance found between resources which may be freely and flexibly allocated for research, which form part of the basic financing of universities, and competitively-awarded external funding for fixed-term projects, determines the long-term ability of a science system to be innovative. In the last decade, the share of fixed-term external funds has rapidly – and worrisomely – increased, to the detriment of research resources provided through basic financing that can be flexibly allocated, including over a longer term. At universities of applied sciences (Fachhochschulen), research is entirely dependent on the availability of external resources. External funds have significantly increased in importance, both for funding research projects and in calculating what share of basic funding for research should be allocated to the universities. The "bonus" external funding provides, at universities where funding allocations are based on indicators, magnifies this effect.
   External funding is important in terms of competition and in accommodating research to external needs or demands. However, a too large share of fixed-term external funding undermines the longer-term funding of research, diminishes the probability of path-breaking research, and eats up a disproportionate share of scholars' time in peer review processes.

5. Of all the forms of performance assessment, formative evaluation, in which peers engage in a dialogue about the potential for improvement and development with the unit being evaluated, is assessed the most positively not only in empirical studies but also by scholars and higher education staff. It is a recognized instrument for improving quality. However, its effectiveness begins to decline the more frequently it is used, and effectiveness is also di-

minished as formative evaluations become a routinized part of externally determined, and temporally brief, cycles of evaluation. The marginal utility of learning effects and improved quality on the one hand, and the effort of evaluation on the other, declines with increased frequency of use as well. After two decades of developing and consolidating quality assurance generally, and evaluating disciplines specifically, the utility of evaluations at many Swiss universities has become small.

6. Scholarly research practices thrive on a high degree of creative freedom and unfettered choices about the contents, goals, methods, and processes to be employed in the research. This in turn means scholars have a high degree of intrinsic motivation. For many of them, this justifies the high personal costs associated with a career in research (which includes relatively low pay relative to the level of qualifications and the time invested, and a long phase of career uncertainty as compared with other professions).
   It is precisely these basic conditions of academic life – an intrinsic motivation and a considerable willingness by researchers to make an effort and take risks – that are slowly being eroded by the ongoing obligations to report or engage in assessment and evaluation processes. Forms of performance assessment in accreditations and evaluations that have elements of control further undermine the cultural basis and attractiveness of academic careers.

7. An *ex ante* steering of academic research by training and selecting researchers, following the goals of an institution, is considerably more effective than any attempts to try to control achievement or performance in an *ex post* manner, or by using incentives. The variety of institutional goals and different realms of research can be reflected in the selection criteria a university uses rather than by employing uniform assessment standards later. Such uniformity serves to mask external and internal differentiation in, across, and between universities, which is what one actually wants to promote. The quality of the selection processes would be measured, therefore, not in terms of the highest international standards in a particular academic discipline but would instead reflect a sensibility and sensitivity to the orientation of the institution, its research units, and the demands placed on them.

8. The core idea of New Public Management in this context, namely that one can improve academic research efficiency by using performance incentives and controls, misjudges the nature of academic research processes, where unforeseen developments, detours, failed experiments, and the risk of ideas that fail or of the falsification of a hypothesis are the lifeblood of progress and a key aspect in its long-term ability to innovate. An academic life oriented to confirmation and steady production runs the longer-term risk of becoming sclerotic. Confirmation also helps promote mediocre research, and it is precisely this mediocrity that is being strengthened by the growing culture of mistrust and an ever-denser net of quality control measures.

8 Bibliography

Amabile, Teresa M. (1996). *Creativity in Context: Update to the Social Psychology of Creativity*. Boulder, CO: Westview Press.

AUBR (2010). *Assessing Europe's University-Based Research*. Expert Group on Assessment of University-Based Research, European Commission.

Bjørnkilde, Thomas, and Christian Bason (2000). *Meta Evaluation of the Evaluations of Higher Education in Denmark. Paper for Case-Study Session at the 4th EES Conference in Lausanne*. 12–14 October 2000.

Böhmer, Susan, et al. (2011). Wissenschaftler-Befragung (2010). Forschungsbedingungen von Professorinnen und Professoren an deutschen Universitäten. *iFQ Working Paper* No. 8. Bonn: iFQ.

Bornemann, Lutz, and Hans-Dieter Daniel (2003). "Begutachtung durch Fachkollegen in der Wissenschaft", in: Schwarz, Stefanie, and Ulrich Teichler (Eds), *Universität auf dem Prüfstand. Konzepte and Befunde der Hochschulforschung*. Frankfurt: Campus, pp. 211–230.

Brennan, John, Tarla Shah, and Ruth Williams (1996). *Quality Assessment and Quality Improvement: An Analysis of the Recommendations Made by HEFCE Assessors*. Bristol: Higher Education Funding Council for England (HEFCE).

Brennan, John, Mark M. H. Frederiks, and Tarla Shah (1997). *Improving the Quality of Education: The Impact of Quality Assessment on Institutions*. Bristol: Higher Education Funding Council for England (HEFCE).

Bundesgesetz vom 8. Oktober 1999 über die Förderung der Universitäten und über die Zusammenarbeit im Hochschulbereich (Universitätsförderungsgesetz, UFG).

Csikszentmihalyi, Mihaly (1996). *Creativity. Flow and the Psychology of Discovery and Invention*. New York: Harper Collins.

Curaj, Adrian, et al. (2012). *European Higher Education at the Crossroads. Between the Bologna Process and National Reforms. Part V: Higher Education Governance in the European Higher Education Area*. Heidelberg: Springer.

Daniel, Hans-Dieter (1993). *Guardians of Science. Fairness and Reliability of Peer Review*. Weinheim: VCH.

Daniel, Hans-Dieter, Sandra Mittag, and Lutz Bornemann (2007). "The Potential and Problems of Peer Evaluation in Higher Education and Research", in: Cavalli, Alessandro (Ed.), *Quality Assessment for Higher Education in Europe*. London: Portland Press, pp. 71–82.

Deci, Edward, Richard Koestner, and Richard Ryan (1999). "A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation", *Psychological Bulletin* 125 (6), pp. 627–668.

Deci, Edward, and Richard Ryan (1985). *Intrinsic Motivation and Self-determination in Human Behaviour*. New York: Plenum.

"Der Forschungsindex. Die Top-Institute der deutschen Forschung", *Bild der Wissenschaft* 5 (1993–1994).

"Die besten deutschen Universitäten", *Focus* 39 (1993). http://www.focus.de/magazin/archiv/jahrgang_1993/ausgabe_39 (accessed on 31 July 2015).

EUA (2008). *Financially Sustainable Universities. Towards Full Costing in European Universities. An EUA Report*. Brussels: Publications of the European University Association.

EUA (2009). *Improving Quality, Enhancing Creativity: Change Processes in European Higher Education Institutions. Final Report of the Quality Assurance for the Higher Education Change Agenda (QAHECA) Project*. Brussels: Publications of the European University Association.

EUA (2010). *Trends 2010: A Decade of Change in European Higher Education*. Brussels: Publications of the European University Association.

EUROHORCs (2006). *Strengthening European Research Cooperation through Partnership. EuroHORCs Perspective on Networking and Coordination of National Programmes*. http://www.eurohorcs.org/SiteCollectionDocuments/Networking_and_Coordination_of_National_Programmes.pdf (accessed on 27 May 2013).

Evidence Ltd (2005). *Impact of Selective Funding of Research in England and the Specific Outcome of HEFCE Research Funding*. Report to HEFCE and the Department for Education and Skills.

Fehr, Ernst, and Armin Falk (1999). "Wage Rigidity in a Competitive Incomplete Contract Market", *Journal of Political Economy* 107 (1), pp. 106–134.

Ferlie, Ewan, Christine Musselin, and Gianluca Andresani (2009). "The Governance of Higher Education Systems: A Public Management Perspective", in: Paradeise, Catherine, et al. (Eds), *University Governance. Western European Comparative Perspectives*. Dordrecht: Springer, pp. 1–19.

Ferlie, Ewan (2012). "Concluding discussion: Paradigms and instruments of public management reform – the question of agency", in: Teelken, Christine, Ewan Ferlie, and Mike Dent (Eds), *Leadership in the public sector: Promises and pitfalls*. New York, NY: Routledge, pp. 237–251.

Fischer, Klaus (2003). "Soziale and kognitive Aspekte des Peer Review-Verfahrens", in: Fischer, Klaus, and Heinrich Parthey (Eds), *Evaluation wissenschaftlicher Institutionen: Wissenschaftsforschung Jahrbuch 2003*. Berlin: Gesellschaft für Wissenschaftsforschung, pp. 23–62.

Frederiks, Mark M. H., Don F. Westerheijden, and Peter J. M. Weusthof (1994). "Effects of Quality Assessment in Dutch Higher Education", *European Journal of Education* 29 (2), pp. 181–199.

Frey, Bruno S. (1997a). *Markt und Motivation: Wie Ökonomische Anreize die (Arbeits-) Moral verdrängen*. München: Vahlen.

Frey, Bruno S. (1997b). "Moral und ökonomische Anreize: Der Verdrängungseffekt", in: Hegselmann, Rainer, and Hartmut Kliemt (Eds), *Moral und Interesse. Zur interdisziplinären Erneuerung der Moralwissenschaften*. München, Wien: R. Oldenbourg Verlag, pp. 111–132.

Frey, Bruno S. (2006). "Evaluitis – Eine neue Krankheit", *IEW Working Papers* No. 293, Institute for Empirical Research in Economics, University of Zurich. http://www.fh-dortmund.de/de/studi/fb/2/personen/lehr/wilharm/medien/evalu_doku/Evaluitis_BrunoFrey.pdf (accessed on 27 May 2013).

Frey, Bruno S. (2007). "Evaluierungen, Evaluierungen … Evaluitis", *Perspektiven der Wirtschaftspolitik* 8 (3), pp. 207–220.

Frey, Bruno, and Margit Osterloh (2006). "Evaluations: Hidden Costs, Questionable Benefits and Superior Alternatives", *IEW Working Papers* No. 32, Institute for Empirical Research in Economics, University of Zurich. http://www.iew.unizh.ch/wp/iewwp302.pdf (accessed on 17 December 2012).

Henkel, Mary (1999). "The Modernisation of Research Evaluation: The Case of the UK", *Higher Education* 38, pp. 105–122.

Hornbostel, Stefan (1997). *Wissenschaftsindikatoren. Bewertungen in der Wissenschaft*. Opladen: Westdeutscher Verlag.

Hornbostel, Stefan (2008). "Neue Evaluationsregime? Von der Inquisition zur Evaluation", in: Matthies, Hildegard, and Dagmar Simon (Eds), *Wissenschaft unter Beobachtung. Effekt and Defekt von Evaluationen. Leviathan: Zeitschrift für Sozialwissenschaft*. Sonderheft 24/2007. Wiesbaden: Verlag für Sozialwissenschaften, pp. 59–82.

Hornbostel, Stefan, and Anna Schelling (Eds) (2011). *Evaluation: New Balance of Power. iFQ Working Paper* No. 9. Bonn: iFQ.

Horrobin, David F. (1990). "The Philosophical Basis of Peer Review and the Suppression of Innovation", *Journal of the American Medical Association* 263, pp. 1438–1441.

Hulpiau, Veerle, and Kim Waeytens (2003). "Improving quality of education: What makes it actually work? A case study", in: Prichard, Craig, and Paul Trowler (Eds), *Realizing qualitative research into higher education*. Aldershot: Ashgate Publishing, pp. 145–169.

Jacob, Anna-Katharina, and Ulrich Teichler (2011). *Der Wandel des Hochschullehrerberufs im internationalen Vergleich. Ergebnisse einer Befragung in den Jahren 2007/08*. Bonn, Berlin: BMBF. http://www.uni-kassel.de/wz1/pdf/BMBF_Hochschullehrerstudie2011_Druck.pdf (accessed on 20 December 2012).

Jeliazkova, Margarita (2002). "Running the Maze: Interpreting External Review Recommendations", *Quality in Higher Education* 8 (1), pp. 89–96.

Kieser, Alfred (2010). "Unternehmen Wissenschaft?", *Leviathan. Berliner Zeitschrift für Sozialwissenschaft* 38, pp. 347–376.

Kohn, Alfie (1993). *Punished by Rewards*. New York: Plenum Press.

Krull, Wilhelm (2011). "Bewertung, Begutachtung and Evaluation in Wissenschaft and Forschung", in: Hornbostel, Stefan, and Anna Schelling (Eds), *Evaluation: New Balance of Power. iFQ Working Paper* No. 9, pp. 15–23. Bonn: iFQ.

Lamont, Michèle (2009). *How Professors Think: Inside the Curious World of Academic Judgment*. Cambridge: Harvard University Press.

Lamont, Michèle (2011). "Pragmatic Fairness: Production of the Sacred while Observing the Rules", in: Hornbostel, Stefan, and Anna Schelling (Eds), *Evaluation: New Balance of Power. iFQ Working Paper* No. 9, pp. 47–56. Bonn: iFQ.

Langfeldt, Liv (2004). "Expert Panels Evaluating Research: Decision-making and Sources of Bias", *Research Evaluation* 13 (1), pp. 51–62.

League of European Research Universities (LERU) (2012). *Research Universities and Research Assessment*. Brussels: Publications of the League of Research Universities.

Lee, Frederic S., and Sandra Harley (1998). "Economics divided: the limitations of peer review in a paradigm bound social science", in: Jary, David, and Marin Parker, (Eds), *The New Higher Education: Issues and Directions for the Post-Dearing University*. Stoke-on-Trent: Staffordshire University Press, pp. 185–204.

Lee, Frederic (2007). "Research Assessment Exercise, the State and the Dominance of Mainstream Economics in British Universities". *Cambridge Journal of Economics* 31, pp. 309–325.

Lepper, Mark, and David Greene (Eds) (1978). *The Hidden Costs of Reward: New Perspectives on the Psychology of Human Motivation*. New York: John Wiley.

Loukkola, Tia, and Thérèse Zhang (2010). *Examining Quality Culture: Part 1 – Quality Assurance Processes in Higher Education Institutions*. Brussels: Publications of the European University Association.

Matthies, Hildegard, and Dagmar Simon (Eds) (2008). *Wissenschaft unter Beobachtung. Effekt und Defekt von Evaluationen. Leviathan: Zeitschrift für Sozialwissenschaft*. Sonderheft 24/2007. Wiesbaden: Verlag für Sozialwissenschaften.

May, Robert (2004). "The Pathologies of the RAE", *The Journal of the Foundation for Science and Technology* 18 (4), pp. 3–4.

Mittag, Sandra, Lutz Bornemann, and Hans-Dieter Daniel (2006). "Qualitätssicherung and -verbesserung von Studium and Lehre durch Evaluation. Akzeptanz und Folgen mehrstufiger Evaluationsverfahren", *Beiträge zur Hochschulforschung* 28 (2), pp. 6–27.

Moed, Henk, Marc Luwel, and Anton Nederhof (2002). "Towards Research Performance in the Humanities", *Library Trends* 50, pp. 498–520.

Moscati, Roberto (2012). "University Governance in Changing European Systems of Higher Education", in: Curaj, Adrian, Peter Scott, Lazar Vlasceanu and Lesley Wilson (Eds), *European Higher Education at the Crossroads: Between the Bologna Process and National Reforms*. Dordrecht: Springer, pp. 599–611.

Moxham, Harriet, and Joe Anderson (1992). "Peer Review. A View from the Inside", *Science and Technology Policy*, pp. 7–15.

Müller-Böling, Detlef (Ed.) (1995). *Qualitätssicherung in Hochschulen. Forschung – Lehre – Management*. Gütersloh: CHE der Bertelsmann-Stiftung.

Neave, Guy (2012). *The Evaluative State, Institutional Autonomy and Re-engineering Higher Education in Western Europe: The Prince and his Pleasure*. Basingstoke: Palgrave Macmillan.

Niedermann, Christoph, Urs Hugentobler, and Konrad Osterwalder (2003). *Handbuch zum Qualitätsmanagement von Lehre, Forschung und Dienstleistungen an der ETH Zürich*, Zurich: ETH Zürich.

Nowotny, Helga, Peter Scott, and Michael Gibbons (2001). *Re-thinking Science. Knowledge and the Public in an Age of Uncertainty*. Cambridge: Blackwell.

Olbrecht, Meike, and Lutz Bornemann (2010). "Panel Peer Review of Grant Applications: What Do We Know from Research in Social Psychology on Judgment and Decision-Making in Groups?", *Research Evaluation* 19 (4), pp. 293–304.

Olbrecht, Meike, and Thamar Klein (2011). "SFB-Begutachtung: Entscheidungsfindung in Gruppen", in: Hornbostel, Stefan, and Anna Schelling (Eds), *Evaluation: New Balance of Power. iFQ Working Paper* No. 9, pp. 33–45. Bonn: iFQ.

Overbeke, John, and Elizabeth Wager (2003). "The State of the Evidence: What We Know and What We Don't Know about Journal Peer Review", in: Godlee, Fiona, and Tom Jefferson (Eds), *Peer Review in Health Sciences*. London: BMJ Books, pp. 45–61.

PA Consulting Group (2009). *RAE 2008 Accountability Review*. HEFCE.

Paradeise, Catherine, et al. (Eds) (2009). *University Governance. Western European Comparative Perspectives*. Dordrecht: Springer.

Paradeise, Catherine (2011). "La réception des démarches d'évaluation", Paper presented at the CRUS Colloquium *Mesurer les performances de la recherche: enjeux et perspectives*, Lausanne, 7 October 2011.

Paradeise, Catherine (2012). "Tools and Implementation for a New Governance of Universities: Understanding Variability Between and Within Countries", in: Curaj, Adrian, et al., *European Higher Education at the Crossroads. Between the Bologna Process and National Reforms. Part V: Higher Education Governance in the European Higher Education Area*. Heidelberg: Springer, pp. 573–598.

Paradeise, Catherine, Emanuela Reale, and Gaële Goastellec (2012). "A Comparative Approach to Higher Education Reforms in Western European Countries", in: Paradeise, Catherine, et al. (Eds), *University Governance. Western European Comparative Perspectives*. Dordrecht: Springer, pp. 197–226.

Pechar, Hans (2012). "The decline of an academic oligarchy. The Bologna Process and 'Humboldt's last warriors", in: Curaj, Adrian, Peter Scott, Lazar Vlasceanu and Lesley Wilson (Eds), *European Higher Education at the Crossroads: Between the Bologna Process and National Reforms*. Dordrecht: Springer, pp. 613–630.

Peters, Douglas, and Stephen Ceci (1982). "Peer-review Practices of Psychological Journals: The Fate of Published Articles, Submitted Again", *The Behavioral and Brain Sciences* 5 (2), pp. 187–195.

Power, Michael (1997). *The Audit Society. Rituals of Verification*. Oxford: Oxford University Press.

Power, Michael (2008). "Research Evaluation in the Audit Society", in: Matthies, Hildegard, and Dagmar Simon (Eds), *Wissenschaft unter Beobachtung. Effekt and Defekt von Evaluationen. Leviathan: Zeitschrift für Sozialwissenschaft*, Sonderheft 24/2007. Wiesbaden: Verlag für Sozialwissenschaften, pp. 15–24.

Rauhvargers, Andrejs (2011). *Global University Rankings and their Impact*. Brussels: EUA Publications.

Röbbecke, Martina (2007). "Evaluation als neue Form der 'Disziplinierung' – ein nicht intendierter Effekt?" in: Matthies, Hildegard, and Dagmar Simon (Eds), *Wissenschaft unter Beobachtung. Effekt and Defekt von Evaluationen. Leviathan: Zeitschrift für Sozialwissenschaft.* Sonderheft 24/2007. Wiesbaden: Verlag für Sozialwissenschaften, pp. 161–177.

Röbbecke, Martina, and Dagmar Simon (2001). *Reflexive Evaluation. Ziele, Verfahren and Instrumente der Bewertung von Forschungsinstituten.* Berlin: edition sigma.

Schweizerische Universitätskonferenz (SUK) (2006). *Richtlinien für die Qualitätssicherung an den schweizerischen universitären Hochschulen (Qualitätssicherungs-Richtlinien) vom 7. Dezember 2006.*

The Danish Evaluation Institute (2004). *Effect Study. Fact Finding Study*. Kopenhagen.

Thune, Christian, and Dorte Kristoffersen (1999). *Guarding the Guardian: The Evaluation of the Danish Centre for Quality Assurance and Evaluation of Higher Education.* Paper presented on INQAAHE Conference in Chile, May 1999.

Torka, Marc (2011). "Institutioneller gleich handlungspraktischer Wandel? Das Beispiel von Begutachtungspraktiken bei der Evaluation wissenschaftlicher Einrichtungen", in: Hornbostel, Stefan, and Anna Schelling (Eds), *Evaluation: New Balance of Power. iFQ Working Paper* No. 9, pp. 69–81. Bonn: iFQ.

van Raan, Anthony (1995). "Bewertung von Forschungsleistungen: Fortgeschrittene bibliometrische Verfahren", in: Müller-Böling, Detlef (Ed.), *Qualitätssicherung in Hochschulen. Forschung – Lehre – Management.* Gütersloh: CHE der Bertelsmann-Stiftung, pp. 85–102.

van Raan, Anthony (2005). "Measurement of Central Aspects of Scientific Research: Performance, Interdisciplinarity, Structure", *Measurement* 3 (1), pp. 1–19.

Weingart, Peter (1995). "Forschungsindikatoren: Instrumente politischer Legitimierung oder organisatorischen Lernens", in: Müller-Böling, Detlef (Ed.), *Qualitätssicherung in Hochschulen. Forschung – Lehre – Management.* Gütersloh: CHE der Bertelsmann-Stiftung, pp. 73–84.

Weingart, Peter (2005a). "Impact of Bibliometrics upon the Science System: Inadvertent Consequences?", *Scientometrics* 62 (1), pp. 117–131.

Weingart, Peter (2005b). "Das Ritual der Evaluierung und die Verführbarkeit der Zahlen", in: Weingart, Peter (Ed.), *Die Wissenschaft der Öffentlichkeit: Essays zum Verhältnis von Wissenschaft, Medien and Öffentlichkeit.* Weilerwist: Velbrück, pp. 102–122.

"Welche Uni ist die beste? SPIEGEL-Rangliste der deutschen Hochschulen", *Der Spiegel* 16 (1993). http://www.spiegel.de/spiegel/print/index-1993-16.html (accessed on 31 July 2015).

Wenneras, Christine, and Agnes Wold (1999). "Bias in Peer Review of Research Proposals", in: Godlee, Fiona, and Tom Jefferson (Eds), *Peer Review in Health Sciences.* London: BMJ Books, pp. 79–89.

Whitley, Richard, Lars Engwall, and Jochen Glaser (Eds) (2010). *Reconfiguring Knowledge Production: Changing Authority Relationships in the Sciences and their Consequences for Intellectual Innovation.* Oxford: Oxford University Press.

Wiesenthal, Helmut (2008). "Evaluation als Organisationslernen", in: Matthies, Hildegard, and Dagmar Simon (Eds), *Wissenschaft unter Beobachtung. Effekt and Defekt von Evaluationen. Leviathan: Zeitschrift für Sozialwissenschaft.* Sonderheft 24/2007. Wiesbaden: Verlag für Sozialwissenschaften, pp. 315–327.

Wissenschaftsrat (2011). *Neuere Entwicklungen der Hochschulfinanzierung in Deutschland. Bericht des Vorsitzenden zu aktuellen Tendenzen im Wissenschaftssystem vom 8. Juli 2011.*

Appendix

Overview of Research Assessment by User Groups and Goals

Source: European Commission, Expert Group on Assessment of University-Based Research, "Assessing Europe's University-Based Research" (AUBR Report), 2010, pp. 30–34

User Group	Why is Research Assessment Data Required?	What Research Assessment Data Is Required?
HE Management and Governance		
Governing Bodies/ Councils	• Policy and planning • Strategic positioning • Research strategy development/management • Investor confidence/value-for-money and efficiency • Quality assurance	• Institutional and discipline/field data re. level of intensity, expertise, quality and competence • Benchmarking against peer institutions, nationally and worldwide • Efficiency level: how much output vis-à-vis funding • Quality of academic staff and PhD students • Attraction capacity: recruitment of students, academics and researchers from outside region and internationally
HE Executives/ Management	• Policy and planning • Strategic positioning • Research strategy development/management • Investor confidence/value-for-money and efficiency • Quality assurance • Publicity • Student and academic recruitment • Improve and benchmark performance and quality	• Institutional and discipline/field data re. level of intensity, expertise, quality and competence • Benchmarking against peer institutions, nationally and worldwide • Efficiency level: how much output vis-à-vis funding • Quality of academic staff and PhD students • Attraction capacity: recruitment of students, academics and researchers from outside region and internationally • Identification of partnerships (academic, public/private sector, NGOs, research organizations, etc.)
HE Research Groups	• Strategic positioning • Research strategy development/management • Investor confidence/value-for-money and efficiency • Student and academic recruitment	• Discipline data re. level of intensity, expertise, quality and competence benchmarked against peer institutions • Quality of academic staff and PhD students • Attraction capacity: recruitment of students, academics and researchers from outside region and internationally • Identification of partnerships (academic, public/private sector, NGOs, research organizations, etc.)
Governments and Government Agencies		
EU and National Governments	• Define policy and inform decisions about HE system and HEIs • Determine national/international competitiveness • Quality, sustainability, relevance and impact of research activity • Investor confidence/value-for-money and efficiency • Improve performance and quality • Improve system functionality	• System and institutional data re. level of intensity, expertise, quality and competence • Performance of HE system and individual institutions • Benchmarking between nationally and worldwide • Indicator of national competitiveness • Attraction capacity: recruitment of students, academics and researchers from outside region and internationally • Quality of academic staff and PhD students • Efficiency level: how much output vis-à-vis funding • Research infrastructure: level of use and efficiency
Ministries of Education/ Higher Education or Enterprise and Employment	• Policy and planning • Strategic positioning of HE institutions • Quality, sustainability, relevance and impact of research activity • Research strategy development/management • Investor confidence/value-for-money and efficiency • Quality assurance	• Institutional and discipline/field data re. level of intensity, expertise, quality and competence • Benchmarking against peer institutions, nationally and worldwide • Indicator of national competitiveness • Performance of HE system and individual institutions • Attraction capacity: recruitment of students, academics and researchers from outside region and internationally • Efficiency level: how much output vis-à-vis funding • Research infrastructure: level of use and efficiency

Local and Regional Governments	• Define local/regional policy and competitiveness • Quality, sustainability, relevance and impact of research activity • Improve integration/collaboration between universities, government and private sector • Improve attraction capacity	• Benchmarking performance and quality of HE system/institutions nationally and worldwide • Indicator of national competitiveness • Attraction capacity: recruitment of students, academics and researchers from outside region and internationally • Efficiency level: how much output vis-à-vis funding
HE Agencies	• Define policy and inform decisions about HE system and HEIs • Quality, sustainability, relevance and impact of research activity • Determine national/international competitiveness • Investor confidence/value-for-money and efficiency • Improve performance and quality • Improve system functionality	• System and institutional data re. level of intensity, expertise, quality and competence • Performance of HE system and individual institutions • Benchmarking between nationally and worldwide • Indicator of national competitiveness • Attraction capacity: recruitment of students, academics and researchers from outside region and internationally • Quality of academic staff and PhD students • Efficiency level: how much output vis-à-vis funding • Research infrastructure: level of use and efficiency
Other Government Agencies	• Improve and benchmark performance and quality • Aid resource allocation • Investor confidence/value-for-money and efficiency	• Benchmarking performance and quality of HE system institutions nationally and worldwide
Academic Organizations and Academies		
	• Benchmark professional and academic performance and quality • Student and academic recruitment	• Academic and discipline/field data re. level of intensity, expertise, quality and competence • Benchmarking against peer institutions, nationally and worldwide • Quality of academic staff and PhD students
Individuals		
Academics and Researchers	• Identify career opportunities • Identify research partners • Identify best research infrastructure and support for research	• Institutional and field data re. level of intensity, expertise, quality, competence and sustainability • Performance of individual institution benchmarked against peers in field of interest • Employment conditions • Impact of research on teaching, staff/student ratio • Institutional research support
Students	• Inform choice of HEI • Identify career opportunities	• Institutional and field data re. level of intensity, expertise, quality, competence and sustainability • Performance of individual institution benchmarked against peers in field of interest • Research capacity of institution and research team, e.g. graduate students/academic ratio, age of PhD students, time to completion, structure/characteristics of PhD programme and support • Graduate career and employment trends • Quality of the research infrastructure • Staff/student ratio

Peer HEIs		
	• Identify peer HEIs and best research partners	• Institutional and field data re. level of intensity, expertise, quality, competence and sustainability • Performance of individual institutions and researchers benchmarked against peers in field of interest • Research capacity of institution and research team • Potential for partnership
Industry Partner Organizations		
Private Firms and Entrepreneurs	• Quality, sustainability, relevance and impact of research activity • Identify potential partners and expertise • Identify consultancy, technology transfer and knowledge transfer partners and expertise • Identify potential employees	• Institutional and field data re. level of intensity, expertise, quality, competence and sustainability • Performance of individual institution benchmarked against peers in field of interest • Competitive positioning of institution and researchers • Trends in graduate employment and competence • Quality of HE programme, and link between research and teaching
Public Organizations	• Quality, sustainability, relevance and impact of research activity • Identify potential partners and expertise • Identify consultancy, technology transfer and knowledge transfer partners and expertise • Identify potential employees	• Institutional and field data re. level of intensity, expertise, quality, competence and sustainability • Performance of individual institution benchmarked against peers in field of interest • Competitive positioning of institution and researchers • Trends in graduate employment and competence • Quality of HE programme, and link between research and teaching
Employers	• Quality, sustainability, relevance and impact of research activity • Identify potential partners and expertise • Identify consultancy, technology transfer and knowledge transfer partners and expertise • Identify potential employees	• Institutional and field data re. level of intensity, expertise, quality, competence and sustainability • Performance of individual institution benchmarked against peers in field of interest • Competitive positioning of institution and researchers • Trends in graduate employment and competence • Quality of HE programme, and link between research and teaching
Civic Society and Civic Organizations		
	• Identify specific expertise and information • Identify potential collaborator • Identify consultancy, technology transfer and knowledge transfer partners	• Institutional and field data re. expertise, quality and competence • Peer esteem indicators
Ministries of Higher Education in Developing Countries		
	• To help determine which foreign higher education institutions are applicable for overseas scholarships studies • To help determine research partnerships for knowledge and technology transfer	• Institutional and discipline/field data re. level of intensity, expertise, quality and competence • Competitive positioning of institution and researchers • Trends in graduate employment and competence • Quality of academic staff and PhD students

Sponsors and Private Investors		
Benefactors/ Philanthropists	• Determine institutional performance vis-à-vis national and international competitors • Investor confidence/value-for-money and efficiency • Quality, sustainability, relevance and impact of research activity • Quality of academic staff and PhD student • Contributor to own brand image	• Institutional data re. level of quality and international competitiveness • Benchmarking between nationally and worldwide • Quality of academic staff and PhD students
Alumni	• Determine institutional performance vis-à-vis national and international competitors • Investor confidence/value-for-money and efficiency • Quality of academic staff and PhD students activity • Quality of academic staff and PhD student • Reflect pride and career aspirations/reputation	• Institutional data re. level of quality and international competitiveness • Benchmarking between nationally and worldwide • Quality, sustainability, relevance and impact of research
Public Opinion		
	• Determine institutional performance vis-à-vis national and international competitors • Quality, sustainability, relevance and impact of research activity • Student choice and career opportunities • Investor/parental confidence and value-for-money	• Institutional data re. level of intensity, expertise and competence • Benchmarking against peer institutions, nationally and worldwide • Indicator of national competitiveness • Performance of HE system and individual institutions • Efficiency level: how much output vis-à-vis funding

Abbreviations

AHELO	Assessment of Higher Education Learning Outcomes (OECD)
ARWU	Academic Ranking of World Universities
AUBR	Assessment of University–Based Research
BMBF	Federal Ministry of Education and Research (Germany)
CHE	Centre for Higher Education (Germany)
CRUS	Conference of Swiss University Rectors
DFG	German Research Foundation
ETH	Federal Institute of Technology (Switzerland)
EUA	European University Association
EUROHORCs	European Heads of Research Councils
HE	Higher Education
HEFCE	Higher Education Funding Council for England
HEI	Higher Education Institution
iFQ	Institute for Research Information and Quality Assurance (Berlin)
LERU	League of European Research Universities
OECD	Organization for Economic Co–operation and Development
RAE	Research Assessment Exercise
REF	Research Excellence Framework
SFB	Collaborative Research Centres
SUK	Swiss University Conference
SSTC	Swiss Science and Technology Council
UFG	Federal Law on Financial Aid to Universities (Switzerland)
WKN	Scientific Commission of Lower Saxony (Germany)

Imprint

Swiss Science and Innovation Council SSIC
Einsteinstrasse 2
CH–3003 Bern
T 0041 (0)58 463 00 48
F 0041 (0)58 463 95 47
swir@swir.admin.ch
www.swir.ch

ISBN 978–3–906113–27–2
Bern 2015

Copy–editing: Doris Tranter
Layout: VischerVettiger, Basel
Cover photograph: Mélanie Rouiller

Swiss Science and Innovation Council SSIC
Einsteinstrasse 2
CH–3003 Bern

T 0041 (0)58 463 00 48
F 0041 (0)58 463 95 47
swir@swir.admin.ch
www.swir.ch