

# **Impact Faktoren**

## **Absurde Vermessung der Wissenschaft**

**Von Bruno S. Frey und Margit Osterloh**

**Zusammenfassung:** Forschende werden heute weitgehend nach Indizes wie der Anzahl ihrer Veröffentlichungen in Zeitschriften mit einem hohen Impact-Faktor beurteilt. Diese entscheiden maßgeblich über Karrieren und die Verteilung der staatlichen Mittel. Der Beitrag zum wissenschaftlichen Fortschritt hat demgegenüber wenig Bedeutung. Dieses System hat allerdings schwerwiegende Mängel. Gibt es Alternativen? Wir stellen drei Vorschläge zur Diskussion: Input- anstelle von Output-Kontrolle, stärkere Betonung von Zufall als Auswahlverfahren und das offene Post-Publication-Peer-Review-Verfahren. Alle drei Vorschläge stellen darauf ab, das qualitative Urteil der „Gelehrtenrepublik“ wieder in Stellung zu bringen.

### **1. Lässt sich wissenschaftliche Qualität messen?**

Der Rang einer wissenschaftlichen Zeitschrift wird heute weitgehend durch ihren *Impact Faktor* bestimmt. Der Impact Faktor erfasst, wie häufig im Durchschnitt in dieser Zeitschrift gedruckte Artikel im Zeitraum von zwei Jahren nah deren Veröffentlichung von anderen wissenschaftlichen Zeitschriften zitiert wurden. Der Impact Faktor wird jährlich für die Zeitschriften aktualisiert, die im Thomson Reuters' ISI Web of Science gelistet sind. Er wird als allgemeingültige und verlässliche Messung der wissenschaftlichen Leistung dargestellt: „a systematic, objective means to critically evaluate the world's leading journals based on citation data“.<sup>1</sup> Entsprechend wird eine Veröffentlichung in einer wissenschaftlichen Zeitschrift mit einem hohen Impact Faktor – einem sog. A+ oder A-Journal – als „gut“ taxiert, eine solche in einer Zeitschrift mit einem tiefem Impact Faktor als wenig bedeutend. Diese Interpretation hat sich international durchgesetzt (z.B. Archambault / Larivière 2009, Jarwal / Brion / King 2009).

Forschende werden – zumindest in den Natur- und Sozialwissenschaften – heute weitgehend danach beurteilt, wie viele ihrer Veröffentlichungen in Zeitschriften mit einem hohen Impact-Faktor veröffentlicht werden. Es gibt einige Varianten dieses Bewertungs-Verfahrens sowie weitere Indizes, die auf den Zählungen vom ISI beruhen. Am bekanntesten ist der Hirsch-Index h (vgl. zur Darstellung z.B. Helbing / Balietti 2011). Ein Forschender hat einen Hirsch-Index von h, wenn h seiner oder ihrer Publikationen mindestens h-mal zitiert wurden. Eine Person hat z.B. einen h-Index von 15, wenn 15 ihrer Publikationen wenigstens 15-mal zitiert werden. Manche Wissenschaftlerinnen und Wissenschaftler geben sogar in ihrem Lebenslauf an, wie hoch ihr Hirsch-Index oder – wie mitunter in der Biomedizin – wie hoch die Kumulation der Impact Faktoren der Zeitschriften ist, in denen sie veröffentlicht haben (Haustein / Larivière 2015).

Damit hat sich eine beinahe totale *Vermessung der Wissenschaft* durchgesetzt. Der Inhalt einer Publikation hat demgegenüber wenig Bedeutung. Wurde früher bewundernd von jemandem gesagt, er oder sie hätte einen wichtigen Beitrag zu einem bestimmten wissenschaftlichen Thema geliefert, gilt heute eine hohe Anzahl von Veröffentlichungen in A-Journals oder ein hoher Hirsch-Index als bemerkenswert.

Die Vermessung der Wissenschaft anhand von Impact Faktoren oder anderen Indizes könnte als Spielerei abgetan werden, wenn nicht diese Qualitätsbewertung heute massgeblich über die Verteilung der staatlichen Mittel und über die Karrieren von Wissenschaftlerin-

1 <http://thomsonreuters.com/journal-citation-reports/>.

nen und Wissenschaftlern entscheiden würde. Bei Habilitationen und Berufungen werden an manchen Universitäten die Forscherinnen und Forscher mit Punkten bewertet, die sich aus den Impact-Faktoren ergeben. So gibt zum Beispiel eine A+ Publikation sieben Punkte, hingegen eine C Publikation nur einen Punkt. Einige Universitäten zahlen Geldbeträge für Publikationen in mittels Impact Faktoren bewerteten „guten“ Zeitschriften.

## 2. Schwerwiegende Mängel

Die Vorstellung, eine Publikation in einer „guten“ Zeitschrift sei auch eine „gute“ Publikation, ist verfehlt. Zum ersten werden im Impact Faktor wegen des viel zu kurzen Zeithorizonts von zwei Jahren viele Zitierungen nicht erfasst. Gerade besonders originelle und ungewöhnliche Beiträge brauchen Zeit, bis sie von andern Wissenschaftlern zur Kenntnis genommen werden. Es gibt Beispiele, dass es ein Dutzend Jahre und mehr dauern kann (siehe Helbing / Balietti 2011).

Zum zweiten werden nur Zitate in gelisteten Zeitschriften, nicht aber in Büchern, Blogs oder anderen sozialen Medien erfasst, welche in manchen Bereichen der wissenschaftlichen Kommunikation eine grosse Rolle spielen und darüber hinaus für den interdisziplinären Diskurs von großer Bedeutung sind.<sup>2</sup>

Zum dritten entstehen Fehler, wenn die zitierenden und zitierten Wissenschaftler in Übereinstimmung gebracht werden sollen. Thomson Reuters' Web of Science wird vorgeworfen, dass solche Fehler häufig auftreten und darüber hinaus schwer überprüfbar sind (Monastersky 2005).<sup>3</sup> Darüber hinaus werden Mehrfach- und Einzelautorschaften nicht unterschieden. Geringe Unterschiede in der Messmethodik führen zu großen Unterschieden (Rost / Frey 2011).

Zum vierten gibt es einen Bias zugunsten englischsprachiger bzw. US-Journals und zugunsten der Sozial- und Kulturwissenschaften, welche im Web of Science unterrepräsentiert sind (Hornbostel et al. 2009; Schmoch 2015). Auch gibt es zwischen Disziplinen und sogar Teildisziplinen unterschiedliche Zitationsgewohnheiten, sodass entsprechende Vergleiche unmöglich sind (Kieser 2012).

Zum fünften sind hohe Zitationsraten und Impact Faktoren keineswegs immer ein Maßstab für wissenschaftliche Bedeutung. Es wird dabei nicht berücksichtigt, ob die zitierten Publikationen Übersichtsartikel oder Originalarbeiten darstellen, ob es sich dabei um die Behandlung bloß modischer Themen handelt oder ob „Zitationskartelle“ am Werk sind. Darüber hinaus werden manche Artikel nur deshalb zitiert, weil sie als falsch oder zumindest als mangelhaft betrachtet werden. Es soll sogar Autoren geben, die bewusst Lücken oder Fehler in ihre Aufsätze einbauen, damit möglichst viele Kollegen und Kolleginnen veranlasst werden, sie zu zitieren.

Zum sechsten sind Zitierungen einzelner Artikel in einer Zeitschrift extrem ungleich. Dieses Problem wiegt besonders schwer. Die anhand von Zitaten gemessene Bedeutung eines wissenschaftlichen Organs wird maßgeblich durch *einige wenige Beiträge* bestimmt. Etwa 20% der Publikationen vereinigen 80-90% aller Zitationen auf sich (Hornbostel et al. 2009; Haustein / Larivière 2015). Die meisten Artikel, die in einer aufgrund ihres Impact Faktors als „gut“ eingestuften Zeitschrift publiziert werden, profitieren von den wenigen Artikeln, die eine große Zahl von Zitaten auf sich ziehen (Moed 2002; Baum 2010). Die überwiegende Anzahl von Artikeln wird hingegen kaum oder überhaupt nie zitiert (Rousseauw

<sup>2</sup> Vgl. jedoch neuerdings die sogenannten Altmetrics, z.B. Weller 2015.

<sup>3</sup> Mittels ResearcherID ([www.researcherid.com](http://www.researcherid.com)) können aber neuerdings Autoren ihre Daten überprüfen.

1991). Nur 50% der in referierten Zeitschriften veröffentlichten Artikel werden von anderen als den Autoren und den Gutachtern gelesen. 90% der Artikel werden niemals zitiert (Meho 2006). Laband und Tollison (2003) bezeichnen sie als „dry holes“. Bei einer stark schießen Verteilung darf aber keinesfalls auf Eigenschaften einzelner Werte geschlossen werden. Dennoch werden fälschlicherweise „gut“ publizieren und „gute“ Publikationen weitgehend gleich gesetzt (vgl. Frey / Osterloh 2013).

Eine beträchtliche Zahl von Beiträgen von hoher Qualität – in diesem Fall gemessen an den Zitierungen – wurde in weniger guten Zeitschriften, und von solchen mit geringer Qualität in besonders guten Zeitschriften veröffentlicht (Oswald 2007). Aus dem Impact Faktor einer Zeitschrift lässt sich deshalb nicht auf die Qualität eines einzelnen Beitrages schließen. Das Abstellen auf Impact Faktoren wird deshalb in einer Studie der *International Mathematical Union* (2008: 14) als „breathtakingly naive“ bezeichnet. 2012 wurde in San Francisco die „Declaration of Research Assessment (DORA) verabschiedet und von über 400 Organisationen – unter anderem dem Schweizer Nationalfonds – unterzeichnet. Danach darf die Qualität eines Aufsatzes nicht nach dem Impact Faktor der veröffentlichten Zeitschrift bewertet werden (DORA, 2012).

### 3. Falsche Anreize in der Wissenschaft

Trotz der aufgezeigten schwerwiegenden Mängel wird heute der Impact Faktor als Maßstab für wissenschaftliche Leistung verwendet. Diese als absurd einzustufende Vermessung der Wissenschaft geht nicht auf eine unzureichende Messung zurück, die sich mit besseren Daten und sorgfältigerer Aufbereitung korrigieren ließe. Die Fehler sind vielmehr grundsätzlich.

Die Wissenschaft wird durch die Verwendung von Impact Faktoren in mehrfacher Hinsicht geschädigt. Die Auswahl der zukünftigen Forscherinnen und Forscher wird systematisch verzerrt. Das Abstellen auf solche Indizes kann in keiner Weise eine sorgfältige *inhaltliche Analyse* der einzelnen Beiträge ersetzen. Überdies werden den Forschenden falsche Anreize gesetzt: es wird ihnen deutlich gemacht, dass nicht der Beitrag zur Erkenntnis zählt, sondern die Reputation des Journals, in dem die Erkenntnisse veröffentlicht werden.

Zum ersten vergrößern Leistungsbeurteilungssysteme, welche auf Indizes abstellen, die Unterschiede in den Positionen zwischen einzelnen Wissenschaftlern unverhältnismäßig (Espeland / Sauder 2007). Auf diese Weise werden die vielfältigen Kriterien, nach denen eine komplexe Leistung beurteilt werden müsste, durch Zahlen zunehmend in eine einfältige, hierarchische Rangordnung gebracht (Heintz 2008). Diese untergräbt den „institutionalisierten Skeptizismus“ (Merton 1973), der gute Forschung ausmacht.

Zweitens fördert das Abstellen auf Indizes die Neigungen, die Regeln zu manipulieren, um das „System zu schlagen“, was zu „ranking games“ führt (Osterloh / Frey 2014). Dies ist umso stärker zu erwarten, desto mehr die intrinsische Motivation eines „taste for science“ (Merton 1973) durch extrinsische Belohnungen aufgrund von quantitativen Indikatoren zerstört ist.

Drittens werden die aufgezeigten Probleme auf institutioneller Ebene durch „lock-in“-Effekte verstärkt. Einzelne Personen oder Institutionen können sich ihnen nur schwer entziehen, auch wenn sie deren Schädlichkeit erkennen. Wissenschaftlerinnen und Wissenschaftler werden immer stärker dem Druck ausgesetzt, sich dem Diktat des „publish or perish“ zu beugen. Das Ergebnis ist „publishing as prostitution“ (Frey 2003).

#### 4. Drei Vorschläge

Wie lässt sich angesichts der geschilderten Situation dennoch eine Leistungsbewertung in der Wissenschaft vornehmen, ohne auf Impact Faktoren und andere Indikatoren abzustellen? Ganz generell muss die Lösung darin gesucht werden, wieder das qualitative Urteil der „Gelehrtenrepublik“ oder der „republic of science“ (Polanyi [1962]2002) in Stellung zu bringen. Dieses Urteil ist zwar fehlerhaft und wenig reliabel (vgl. Starbuck 2005; Reinhart 2012; Osterloh / Frey 2015; Nicolai / Schmal / Schuster 2015), hat aber gewichtige Vorteile, nämlich Vieldimensionalität, Dezentralität und Vielfalt. Vielfalt von Meinungen ist eine wichtige Voraussetzung eines lebendigen wissenschaftlichen Diskurses. Dieser wird durch eine von Indizes hergestellte hierarchische Ordnung behindert. Wir stellen drei Vorschläge zur Diskussion:

Der *erste Vorschlag* besteht darin, die Anlässe für Evaluationen auf wenige karriererelevante Entscheidungen zu reduzieren, z.B. bei der Bewerbung um eine Stelle oder bei der Beantragung von zusätzlichen Forschungsmitteln. Die sogenannte Input-Kontrolle (anstelle einer Output-Kontrolle) dient dazu, das Potential von Forschenden und ihre intrinsische Motivation für wissenschaftliches Arbeiten zu überprüfen (Osterloh 2010; Frey / Osterloh 2012; Frey / Homberg / Osterloh 2013; Osterloh / Frey 2014). Eine rigorose Eingangskontrolle, wie sie z.B. an der Harvard Universität praktiziert wird,<sup>4</sup> macht dauernde Evaluierungen unnötig. Sie ermöglicht weitgehende Autonomie und fördert dadurch den „taste for science“ anstelle eines „taste for publications“ oder „taste for citations“. Die Schwächen der Begutachtung, insbesondere die Schwächen der doppelt blinden Begutachtung von Zeitschriften-Artikeln (Osterloh / Kieser 2015), werden gemildert, weil Begutachtungen auf wenige Anlässe beschränkt werden und deshalb sorgfältiger ausfallen können.

Ein *zweiter Vorschlag* geht von der Beobachtung aus, dass gemäß empirischen Befunden Gutachterurteile verlässlicher sind, soweit es sich um schlechte Beiträge handelt (Cicchetti 1991; Moed 2007). Es besteht mehr Einigkeit unter den Gutachtenden, welche Beiträge abzulehnen sind. Bei guten oder mittleren Beiträgen unterscheiden sich die Einschätzungen oft diametral. Gutachterurteile grenzen hier oft an Zufall (Rothwell / Martyn 2000). Deshalb könnte – zumindest in einem zeitlich beschränkten Experiment – gleich eine *Zufallsauswahl* der zur Publikation anzunehmenden Aufsätze dort durchgeführt werden, wo gegensätzliche Voten vorliegen. Zufall scheint im ersten Moment eine unzureichende Methode zu sein, um wissenschaftliche Beiträge auszuwählen. Eine randomisierte Entscheidung hat jedoch große Vorteile. Sachfremde Erwägungen oder „Vetternwirtschaft“ aufgrund der häufig nicht wirklich doppelt-blinden Begutachtung werden ausgeschlossen.<sup>5</sup> Neue, ungewöhnliche Ideen abseits des „mainstreams“, die es bei etablierten Gutachtenden häufig schwer haben, hätten eine bessere Chance. Nach einigen Jahren ließe sich empirisch ermitteln, ob die zufällig ausgewählten Beiträge den wissenschaftlichen Diskurs besser oder schlechter befriedet haben als die begutachteten Artikel.

Der *dritte Vorschlag* sieht widersprüchliche Gutachten nicht als Problem, sondern als ein Zeichen solider und produktiver Wissenschaft. Kontroversen zwischen Gutachten bieten Anlass für die Fortentwicklung der Wissenschaft. Dies ist allerdings nur dann der Fall, wenn es einen offenen wissenschaftlichen Diskurs unter Begutachteten und Gutachtenden gibt. Dies ist bei der derzeitigen Doppel-Blind-Begutachtung nicht möglich. Im *offenen Post-Publication-Peer-Review-Verfahren* (Kriegeskorte 2012; Frey / Osterloh 2014; Osterloh /

4 Online abrufbar unter: <http://www.fas.harvard.edu/research/greybook/principles.html>, letztes Abrufdatum: 9.6.2015.

5 Vgl zur ausführlichen Diskussion von Zufallsverfahren Duxbury 2008; Buchstein 2009; Zeitoun / Osterloh / Frey 2014.

Kieser 2015) würden Forschende einen etablierten Kollegen oder eine Kollegin als „Editor“ anfragen, ob er oder sie Kommentare einholt. Diese würden auf einer gemeinsamen Plattform veröffentlicht werden. Der „Editor“ würde genannt werden und dadurch – ähnlich heutigen Zeitschriften-Editoren – Reputation gewinnen. Die Kommentare sollten – anders als die heutigen Gutachten – namentlich gekennzeichnet werden, damit sie zitiert werden können und ein Anreiz besteht gehaltvolle Beiträge zu schreiben. Die Verfasser des ursprünglichen Artikels können auf derselben Plattform antworten. Sind die Kommentare oberflächlich oder gar feindselig (wie dies bei anonymen Gutachten mitunter der Fall ist), schädigt dies die Reputation des Gutachtenden. Wenn ein lebendiger Diskurs zustande kommt, können nach einiger Zeit die Ergebnisse als „State of the Art“ präsentiert werden.

Dieses neue System wäre nicht nur erheblich billiger und schneller als das derzeitige System. Es würde vor allem dem offenen wissenschaftlichen Diskurs in der „Republic of Science“ die Bedeutung geben, welche für wissenschaftlichen Fortschritt unabdingbar ist.

Gegen diese Vorschläge werden diejenigen Protest anmelden, welche mit Hilfe des Systems von Impact Faktoren Gewinne erzielt und Einfluss errungen haben. Dazu gehören neben Verlagen auch das Wissenschafts-Management, das in Universitäten und Forschungseinrichtungen durch eine „Governance by Numbers“ (Heintz 2008) gegenüber den Forschenden an Einfluss gewonnen hat, obwohl dies der wissenschaftlichen Leistung wenig nützt (Goodall 2009). Auch werden Übergangs-Probleme auftreten. Aber die Probleme des heutigen Systems sind so riesig (vgl. The Economist 2013), dass dringend Alternativen aufgezeigt und ernsthaft diskutiert werden müssen.

## Referenzen

- Archambault, É. / V. Larivière (2009): History of the journal impact factor: contingencies and consequences, in: *Scientometrics*, 79 / 3, S. 639-653.
- Baum, J. A.C. (2010): Free-Riding on Power Laws: questioning the validity of the Impact Factor as a measure of research quality in organization studies, in: *Organization*, 18 / 4, S. 449-466.
- Buchstein, H. (2009): Demokratie und Lotterie, Frankfurt / Main.
- Cicchetti, D.V. (1991): The reliability of peer review for manuscript and grant submissions: a crossdisciplinary investigation, in: *Behavioral and Brain Sciences* 14 / 01, S. 119-135.
- DORA (San Francisco Declaration on Research Assessment) (2012), online abrufbar unter:  
<http://www.ascb.org/dora-old/files/SFDeclarationFINAL.pdf>, letztes Abrufdatum: 16.12.2014.
- Duxbury, N. (2008): Random Justice, Oxford.
- Espeland, W.N. / M. Sauder (2007): Rankings and Reactivity: How Public measures Recreate Social Worlds. *American Journal of Sociology* 113 / 1, S. 1-40.
- Frey, B.S. (2003): Publishing as prostitution? – Choosing between one's own ideas and academic success, in: *Public Choice* 116, S. 205-223.
- Frey, B.S. / F. Homberg / M. Osterloh (2013): Organizational Control Systems and Pay-for-Performance in the Public Service, in: *Organization Studies* 34 / 7, S. 949-972.
- Frey, B.S. / M. Osterloh (2012): Rankings: Unbeabsichtigte Nebenwirkungen und Alternativen, in: Ökonomenstimme vom 17.2.2012, online abrufbar unter: <http://www.oekonomenstimme.org/artikel/2012/02/rankings-unbeabsichtigte-nebenwirkungen-und-alternativen/>, letztes Abrufdatum: 1.5.2015.
- Frey, B.S. / M. Osterloh (2013): Gut publizieren = gute Publikation?, in: Ökonomenstimme vom 16.5.2013, online abrufbar unter: <http://www.oekonomenstimme.org/artikel/2013/05/gut-publizieren--gute-publikation/>, letztes Abrufdatum: 1.5.2015.

- Frey, B. S. / M. Osterloh (2014): Schlechte Behandlung des wissenschaftlichen Nachwuchses und wie man das ändern könnte, in: Ökonomenstimme vom 28.10.2014, online abrufbar unter: <http://www.oekonomenstimme.org/artikel/2014/10/schlechte-behandlung-des-wissenschaftlichen-nachwuchses-und-wie-man-das-aendern-koennte/>, letztes Abrufdatum: 1.5.2015.
- Goodall, A.H. (2009): Highly cited leaders and the performance of research universities, *Research Policy* 38, S. 1070-1092.
- Haustein, S. / V. Larivière (2015): The Use of Bibliometrics for Assessing Research: Possibilities, Limitations and Adverse Effects, in: I. Welpe / J. Wollersheim / S. Ringelhan / M. Osterloh (Hrsg.), *Incentives and Performance – Governance of Research Organization*, Heidelberg, S. 121-139.
- Heintz, B. (2008): Governance by Numbers. Zum Zusammenhang von Quantifizierung und Globalisierung am Beispiel der Hochschulpolitik, in: G.F. Schuppert / A. Voßkuhl (Hrsg.), *Governance von und durch Wissen*, Nomos, S. 110-304.
- Helbing, D. / S. Balietti (2011): How to create an innovation accelerator, in: EPJ Special Topics 195, S. 101-136.
- Hornbostel, S. / B. Klingsporn / M. von Ins (2009): Messung von Forschungsleistungen – eine Vermessensheit?, in: Alexander von Humboldt Stiftung (Hrsg.), *Publikationsverhalten in unterschiedlichen Disziplinen. Beiträge zur Beurteilung von Forschungsleistungen*, 2. erweiterte Auflage, Berlin, S. 14-35.
- International Mathematical Union (IMU) (2008): Citation Statistics. A report, Corrected version, 16.12.08.
- Jarwal, S.D. / A.M. Brion / M.L. King (2009): Measuring research quality using the journal impact factor, citations and ‚Ranked Journals‘: blunt instruments or inspired metrics?, in: *Journal of Higher Education Policy and Management* 31 / 4, S. 289-300.
- Kieser, A. (2012): JOURQUAL – der Gebrauch, nicht der Missbrauch, ist das Problem. Oder: Warum Wirtschaftsinformatik die beste deutschsprachige betriebswirtschaftliche Zeitschrift ist, in: *Die Betriebswirtschaft* 72, S. 93-110.
- Kriegeskorte, N. (2012): Open evaluation: a vision for entirely transparent post-publication peer review and rating for science, in: *Frontiers in Computational Neuroscience* 6, S. 1-18.
- Laband, D.N. / R.D. Tollison (2003): Dry holes in economic research, in: *Kyklos* 56, S. 161-174.
- Meho, L.I. (2006): The Rise and Fall of Citation Analysis, Preprint physics, online abrufbar unter: <http://arxiv.org/pdf/physics/0701012>, letztes Abrufdatum: 1.5.2015.
- Merton, R.K. (1973): The sociology of science: Theoretical and empirical investigation, Chicago / IL.
- Moed, H.F. (2002): The impact-factors debate: the ISI's uses and limits, in: *Nature* 415 / 6873, S. 731-732.
- Moed, H.F. (2007): The future of research evaluation rests with an intelligent combination of advanced metrics and transparent peer review, in: *Science and Public Policy* 34, S. 575-583.
- Monastersky, R. (2005): The number that's devouring science, in: *Chronicle of Higher Education* 14.10.2005, online abrufbar unter: <http://www3.nd.edu/~pkamat/citations/chronicle.pdf>, letztes Abrufdatum: 1.5.2015.
- Nicolai, A.T. / S. Schmal / C. Schuster (2015): Interrater Reliability of the Peer Review Process in Management Journals, in: I. Welpe / J. Wollersheim / S. Ringelhan / M. Osterloh (Hrsg.), *Incentives and Performance – Governance of Research Organization*, Heidelberg, S. 107-120.
- Osterloh, M. (2010): Governance by Numbers. Does it Really Work in research?, *Analyse & Kritik* 32, S. 267-283.
- Osterloh, M. / B.S. Frey (2014): Ranking Games, in: *Evaluation Review* 23.3.2014, online abrufbar unter: <http://erx.sagepub.com/content/39/1/102>, letztes Abrufdatum: 1.5.2015.
- Osterloh, M. / B.S. Frey (2015): Rankings und der Preis der Wissenschaft, in: *Zeitschrift für Kulturwissenschaft* (erscheint demnächst).

- Osterloh, M. / A. Kieser (2015): Double-Blind Peer Review: How to Slaughter a Sacred Cow, in: I. Welpe / J. Wollersheim / S. Ringelhan / M. Osterloh (Hrsg.), Incentives and Performance – Governance of Research Organization, Heidelberg, S. 307-324.
- Oswald, A.J. (2007): An examination of the reliability of prestigious scholarly journals: Evidence and implications for decision-makers, in: *Economica* 74, S. 21-31.
- Polanyi, M. ([1962]2002): The republic of science: Its political and economic theory, in: *Minerva* 1, S. 54-73, wieder abgedruckt in: P Mirowski / E.M. Sent (2002), *Science Bought and Sold. Essays in the Economics of Science*, Chicago / IL, S. 465–485.
- Reinhart, M. (2012): Soziologie und Epistemologie des Peer Reviews, Baden-Baden.
- Rost, K. / B.S. Frey (2011): Quantitative and qualitative rankings of scholars, in: Schmalenbach Business Review 63, S. 63-91.
- Rothwell, P.M. / C.N. Martyn (2000): Reproducibility of peer review in clinical neuroscience. Is agreement between reviewers any greater than would be expected by chance alone?, *Brain* 123, S. 1964-1969.
- Rousseeuw, P.J. (1991): A diagnostic plot for regression outliers and leverage points, in: Computational Statistics & Data Analysis 11 / 1, S. 127-129.
- Schmoch, U. (2015): The Informative Value of International University Rankings: Some Methodological Remarks, in: I. Welpe / J. Wollersheim / S. Ringelhan / M. Osterloh (Hrsg.), Incentives and Performance – Governance of Research Organization, Heidelberg, S. 141-154.
- Starbuck, W.H. (2005): How much better are the most prestigious journals? The statistics of academic publication, in: *Organization Science* 16, S. 180-200.
- The Economist (2013): How Science goes Wrong, 19.10.2013, abrufbar unter: <http://www.economist.com/news/leaders/21588069-scientific-research-has-changed-world-now-it-needs-change-itself-how-science-goes-wrong>, letztes Abrufdatum: 1.5.2015.
- Weller, K. (2015): Social Media and Altmetrics: An Overview of Current Alternative Approaches to Measuring Scholarly Impact, in: I. Welpe / J. Wollersheim / S. Ringelhan / M. Osterloh (Hrsg.), Incentives and Performance – Governance of Research Organization, Heidelberg, S. 261-277.
- Zeitoun, H. / M. Osterloh / B.S. Frey (2014): Learning from Ancient Athens: Demarchy and Corporate Governance, in: *Academy of Management Perspectives* 28, S. 1-14.

Prof. Dr. Dr. h.c. mult. Bruno S. Frey  
Universität Basel  
Center for Research in Economics and Well-Being (CREW)  
Peter Merian-Weg 6  
4002 Basel, Schweiz  
und  
CREMA – Center for Research in Economics, Management and the Arts  
Südstrasse 11  
8008 Zürich, Schweiz  
[bruno.frey@econ.uzh.ch](mailto:bruno.frey@econ.uzh.ch)

Prof. Dr. Dr. h.c. Margit Osterloh  
Universität Zürich  
und  
Universität Basel  
Center for Research in Economics and Well-Being (CREW)  
Peter Merian-Weg 6  
4002 Basel, Schweiz  
und  
CREMA – Center for Research in Economics, Management and the Arts  
Südstrasse 11  
8008 Zürich, Schweiz  
[margit.osterloh@business.uzh.ch](mailto:margit.osterloh@business.uzh.ch)