

Absurde Mess-Manie

Der fragwürdige Impact des Impact-Faktors

| MARGIT OSTERLOH | BRUNO S. FREY | Bei der Leistungsbeurteilung von Wissenschaftlerinnen und Wissenschaftlern spielt der Impact-Faktor eine essenzielle Rolle. Das ist jedoch unsinnig, sagen die Autoren des folgenden Beitrags.*

Im September 2017 wurde erneut das Handelsblatt-Ranking der deutschsprachigen Ökonomen und Ökonomen veröffentlicht, das in der Zunft und in den Medien große Aufmerksamkeit findet. Es misst die Forschungsleistung einzelner Wissenschaftler anhand der Anzahl von Publikationen in Top-Journals. Es geht damit ähnlich vor wie die heute übliche Leistungsbeurteilung von Forschern in den Natur- und Sozialwissenschaften, nämlich anhand der Anzahl von Veröffentlichungen in Journals mit einem hohen Impact-Faktor. Dieses Verfahren ist aber unsinnig: Es produziert erstens in

einem hohen Ausmaß akademische Trittbrettfahrer. Zweitens – und weitaus schlimmer – führt es zu fragwürdigen Kriterien bei der Stellenbesetzung von Professuren und akademischen Positionen. Die International Mathematical Union nennt dies „atemberaubend naiv“

»75 Prozent der in Nature 2013/2014 veröffentlichten Beiträge wurden weniger als 38 mal – dem Impact-Faktor von Nature – zitiert.«

und in den meisten Fällen sogar falsch. Die Wissenschaft missachtet somit in ihren eigenen Angelegenheiten Kriterien, die sie sonst als unwissenschaftlich bezeichnen würde. Wie konnte das geschehen?

schungsinstitutionen festzustellen. Impact-Faktoren sind für die Reputation entscheidend und müssen deshalb von Universitätsrektoren beobachtet werden. Sie sind international an vielen Universitäten ausschlaggebend für die Besetzung von Professuren und die Vergabe von Forschungsressourcen. Vielfach ist eine Promotion oder Habilitation Formsache, wenn man mehrere Publikationen in einem sog. A-Journal aufweisen kann. Einige Universitäten zahlen auch noch Geldbeträge für Publikationen in solchen Journals.

Die Impact-Logik unterstellt, dass ein in einem A-Journal veröffentlichter Artikel auch

eine „gute Publikation“ darstellt, auf jeden Fall eine bessere als in einem B- oder C-Journal. Dies, weil Publikationen in einem A-Journal *im Durchschnitt* häufig zitiert worden sind.

AUTOREN



Margit Osterloh ist ständige Gastprofessorin an der Universität Basel und Professorin (em.) an der Universität Zürich.



Bruno S. Frey ist ständiger Gastprofessor an der Universität Basel. Er steht im VWL-Ranking des Handelsblatts (2017) an erster Stelle der

Gesamtforschungsleistung (Lebenswerk).

Beide sind Forschungsdirektoren von CREMA – Center for Research in Economics, Management and the Arts, Switzerland.

Impact-Faktor

Der Impact-Faktor gibt an, wie oft alle Artikel eines wissenschaftlichen Journals in den vorangegangenen zwei Jahren *im Durchschnitt* zitiert worden sind. Er wurde ursprünglich entwickelt, um Bibliotheken eine Hilfestellung zu geben, welche Journals die größte Beachtung finden und deshalb regelmäßig geführt werden sollen. Mittlerweile hat sich der Impact-Faktor jedoch zu einem Bewertungsinstrument von einzelnen Wissenschaftlerinnen und Wissenschaftlern und in der Folge von ganzen Institutionen entwickelt. Komitees beurteilen die Qualität von Forschern und Forschungs-Anträgen mit Hilfe von Impact-Faktoren (oder einer seiner zahlreichen Varianten). Regierungen benutzen diese, um die Leistung von For-

Irreführung durch den Impact-Faktor

Die Bewertung von Publikationen – und in der Summe die Qualität eines Wissenschaftlers oder einer Fakultät – anhand von Bewertungen der Journals, in welchen sie veröffentlicht wurden, ist jedoch unsinnig: Die Einstufung eines Journals sagt nichts darüber aus, wie stark ein *einzelner* Artikel in diesem Journal in der Wissenschaft Beachtung fand. Eine solche kommt allenfalls in der Anzahl der Zitierungen der jeweiligen einzelnen Publikation zum Ausdruck. Diese Anzahl ist zwar ein sehr unvollkommenes Maß für die Qualität, gibt aber immerhin die wissenschaftliche Relevanz eines Artikels annähernd wieder. Zitierungen von Beiträgen in akademischen Zeitschriften sind aber äußerst schief verteilt: Einige wenige Aufsätze

werden oft zitiert, die meisten Artikel werden deutlich weniger häufig zitiert, als dies der Impact-Faktor suggeriert.

75 Prozent der in *Nature* 2013/2014 veröffentlichten Beiträge wurden weniger als 38 mal – dem Impact-Faktor von *Nature* – zitiert, wohingegen das erfolgreichste Papier 905 Zitierungen erreichte. Bei *Science* wurden 76 Prozent der Aufsätze weniger zitiert als der Impact-Faktor von 35, während der erfolgreichste Artikel 694 mal zitiert wurde (<http://www.biorxiv.org/content/early/2016/09/11/062109>). Auch in wichtigen Management-Zeitschriften zeigt sich, dass die Bewertung einzelner Aufsätze anhand des Impact-Faktors des Journals unsinnig ist: Im Jahr 2008 erbrachten 20 Prozent der Artikel 50 bis 60 Prozent der jährlichen Zitierungen. Einige wenige Artikel werden 100 mal und häufiger zitiert, die Mehrheit einmal oder sogar gar nicht. In ökonomischen Zeitschriften sieht es ähnlich aus. Gemäß einer Analyse von zwei A-Zeitschriften (*American Economic Review*; *Econometrica*) wurden nach 25 Jahren fünf dort veröffentlichte Beiträge kein einziges Mal zitiert. Hingegen wurden im gleichen Zeitraum in vier B-Journals

die meist zitierten Aufsätze 50 bis 199 mal zitiert. Es ist absurd anzunehmen, dass diese Beiträge schlechter sind als diejenigen, die zwar in einem A-Journal veröffentlicht, aber nie zitiert wurden. Die Zitierungen von Artikeln in einem „weniger guten“ Journal sollten sogar höher als die in einem A-Journal bewertet werden, weil sie eine geringere Chance haben, zur Kenntnis genommen zu werden.

Falsche akademische Karriere-Entscheidungen

Bei einer stark schiefen Verteilung darf aus dem Durchschnittswert kein Schluss auf Einzelfälle gezogen werden. Das sollte jeder wissen, der eine Grundausbildung in Statistik genossen hat. Gleichwohl verwenden Wissenschaftler diese Art der Qualitätsbewertung bei der Entscheidung über Karrieren und Ressourcen! Das widerspricht nicht nur statistischen Grundkenntnissen, sondern es werden dabei zu einem hohen Prozentsatz *Trittbrettfahrer-Karrieren* produziert. Dies gilt für die zahlreichen Autoren, die vom Impact-Faktor einer Zeitschrift profitieren, deren Zitierungen aber weit unter dem Impact-Faktor

liegen. Wird der Impact-Faktor eines Journals als Gütekriterium für einzelne Publikationen in diesem Journal verwendet, führt dies zu falschen Kriterien bei der Besetzung akademischer Positionen und der Verteilung von Ressourcen.

Das System wurde inzwischen nicht nur von der *International Mathematical Union* kritisiert, sondern von vielen einflussreichen Wissenschaftlern und Journal-Editoren. Zahlreiche führende Wissenschafts-Gremien haben die DORA-Erklärung (Declaration on Research Assessment) unterschrieben, welche sich gegen diese unsinnige Praxis richtet. Einige US-amerikanische Spitzen-Universitäten bezeichnen sich heute als „reading schools“. Sie verpflichten sich, bei Berufungen und Beförderungen nicht auf Indikatoren abzustellen, sondern Veröffentlichungen *sorgfältig zu lesen*. Gleichwohl dominiert die Impact-Logik heute an vielen Universitäten und hat darüber hinaus noch weitere negative Auswirkungen.

Rankings als Geschäft

Im Forschungsbetrieb ist eine ganze Ranking-Bürokratie entstanden, welche

mittlerweile viele Arbeitsplätze umfasst und einen beträchtlichen Teil des Budgets und der Energie bindet. Eher harmlos sind bezahlte Lektoren-Dienste, die dazu verhelfen sollen, Papiere stromlinienförmiger zu schreiben, um die Eingangshürden von Top-Journals besser zu überwinden. Weniger harmlos sind an Rankings ausgerichtete interne Berichts-, Evaluations- und Entlohnungssysteme.

Schwerwiegend sind die folgenden Maßnahmen: Hochschulen errichten Kommunikationsabteilungen, welche ein aktives „Ranking-Management“ betreiben. Von der King Saud University in Riad wurden prominenten Forschern Honorare um die 50 000 Euro angeboten, damit sie als Gegenleistung diese Universität auf ihren Publikationen angeben, obwohl sie kaum Zeit dort verbringen. Die King Saud Universität hat ihre Position im Shanghai Ranking auf diese Weise um *mehrere* hundert(!) Positionen verbessert. Höchst fragwürdig sind auch Forschungsprojekte, die auf die Verbesserung internationaler Ranking-Ergebnisse zielen.

Matthäus-Effekte

Rankings – beruhen sie auf noch so irreführenden Kriterien – lösen Matthäus-Effekte aus: Wer hat, dem wird gegeben. Publikationen in A-Journals werden im Durchschnitt häufiger zitiert als solche in B-Journals. Damit wird nicht nur der Impact-Faktor eines Journals von der Anzahl der Zitierungen darin veröffentlichter Aufsätze bestimmt, sondern umgekehrt die Anzahl der Zitierungen vom Impact-Faktor. Ursprünglich geringe Unterschiede werden dadurch uneinholbar verstärkt. Auch Universitätsrektoren müssen Rankings beachten, sind sie doch Teil der Selbstdarstellung und Grundlage für die Einwerbung von Ressourcen. Dies, auch wenn sie ansonsten Rankings als verfehlt ansehen.

Mess-Manie

Angetrieben wird diese Entwicklung durch die Popularisierung von Konzepten der Transparenz und Evidenzbasierung. Diese haben durchaus Vorteile. Experten und Experten-Organisationen müssen Außenstehenden Rechenschaft ablegen. „Old boys networks“ verlieren an Bedeutung. Das schlägt aber in einen Nachteil um, wenn die verwendeten Kriterien undeutlich, interessengeleitet oder gar falsch sind. Außenstehende

werden irregeleitet. Die beabsichtigte Rechenschaftspflicht und Transparenz wird ad absurdum geführt.

Rankings sorgen angeblich für Übersichtlichkeit in einer immer komplexeren Welt. Merkwürdig ist nur das naive Vertrauen in Ranglisten, deren Datenbasis und Konstruktionslogik meist nicht interessieren und die darüber hinaus keineswegs immer transparent sind. Dennoch gelten Rankings als „Objektivitätsgeneratoren“.

»Rankings – beruhen sie auf noch so irreführenden Kriterien – lösen Matthäus-Effekte aus: Wer hat, dem wird gegeben.«

Darüber hinaus haben Rankings die Doppelfunktion der *Darstellung* und *Herstellung* der Wirklichkeit: Die Realität passt sich den Rankings an. Diejenigen schieben sich an die Spitze, die am besten mit dem System spielen können. Sie werden sich gegen jede Veränderung wehren. Wer hingegen hohe wissenschaftliche Kompetenz, aber wenig Freude an dieser Art Wettbewerb aufweist, hat das Nachsehen.

Verquere Impact-Arithmetik

Warum unterstützen jedoch auch große Teile derjenigen das System der Impact-Faktoren, die diese Entwicklung durchaus kritisch sehen und die möglicherweise niemals in den oberen Rängen auftauchen? Zum *Ersten* ist die Revolte gegen ein System schwierig, wenn man befürchtet, der Unfähigkeit, des Neides oder der Misgunst gegen die „Stars“ bezichtigt zu werden. Zum *Zweiten* profitiert paradoxerweise die Mehrheit der Autoren von der schiefen Verteilung der Zitierungen, wenn sich der Impact-Faktor erst einmal als Qualitätskriterium etabliert hat. Die Impact-Arithmetik bewirkt, dass zwei Drittel bis drei Viertel aller Publikationen falsch – nämlich zu gut – bewertet sind. Ein ebenso großer Anteil aller Autorinnen und Autoren ist – beabsichtigt oder nicht – Trittbrettfahrer dieses Systems. Spielt bei Lehrstuhl-Besetzungen der Impact-Faktor eine ausschlaggebende Rolle, hat die Mehrheit der nach diesem Kriterium ausgewählten Professoren kein Interesse, dieses System abzuschaffen.

Alternativen zu Impact-Faktoren

Angesichts dieser Interessenlage wird es schwer sein, das System der Impact-

Faktoren von innen heraus zu reformieren. Doch es steht *erstens* die Glaubwürdigkeit des wissenschaftlichen Bewertungssystems auf dem Spiel. Diese ist deshalb so wichtig, weil es in der Wissenschaft ein systematisches Marktversagen gibt, weshalb der Markt durch ein funktionierendes internes Bewertungssystem ersetzt werden muss. *Zweitens* bewirkt die Impact-Manie, dass insbesondere der wissenschaftliche Nachwuchs heute auf Gedeih und Verderb auf Veröffentlichungen in Top-Zeitschriften angewiesen ist. Nicht nur ist dieses Kriterium äußerst fragwürdig, sondern das Interesse an den Inhalten tritt zurück, ebenso wie das

Interesse an der Lehre. Es gewinnen diejenigen, die am besten die Regeln des Impact-Faktor-Spiels beherrschen. Wie kann man diese Regeln unterlaufen, wenn man davon ausgehen muss, dass quantitative Indikatoren aus dem Werkzeugkasten des Wissenschafts-Managements kaum mehr eliminiert werden können?

Der *erste* Möglichkeit wäre, nur noch auf Zitierungen zu schauen und nicht auf Impact-Faktoren. Das ist zwar nicht befriedigend, aber besser als das Abstellen auf Impact-Faktoren. Eine *zweite* Möglichkeit wurde von Herausgebern und führenden Mitarbeitern wichtiger wissenschaftlicher Publikationsorgane in die Diskussion gebracht. Sie wäre auch für das Handelsblatt-Ranking anwendbar. Der Vorschlag sieht vor, dass zugleich mit der Veröffentlichung der Journal-Rankings deren *Zitierungs-Verteilungen* offengelegt werden. Damit würde gezeigt, dass es zwischen A-, B- oder C-Journals bezüglich der Zitierungen – die ja die Grundlage der Impact-Logik sind – starke Überschneidungen gibt. Es würde deutlich, dass jede *einzelne Publikation* für sich bewertet werden muss. Dieses Vorgehen wird bereits heute von einigen Verlagen praktiziert. Es würde dem Impact-Faktor längerfristig seine Bedeutung nehmen.

Reformen sind dringend notwendig, will die Wissenschaft nicht ihre Glaubwürdigkeit dadurch verlieren, dass sie auf ein fragwürdiges und weitgehend unwissenschaftliches Bewertungs- und Karriere-System setzt.

* Der Artikel ist eine erweiterte Fassung einer Publikation in der „Oekonomenstimme“: <http://www.oekonomenstimme.org/artikel/2017/06/rankings-auf-dem-falschen-dampfer/> Dort finden sich auch Literaturhinweise.