



Center for Research in Economics, Management and the Arts

Rankings sind wissenschaftsfeindlich

Margit Osterloh und Bruno S. Frey

Frankfurter Allgemeine Zeitung, 25.06.2012, Nr. 145, S. 12
DER BETRIEBSWIRT

Beiträge zur aktuellen Wirtschaftspolitik No. 2012-15

CREMA Südstrasse 11 CH - 8008 Zürich www.crema-research.ch

Frankfurter Allgemeine Zeitung, 25.06.2012, Nr. 145, S. 12

DER BETRIEBSWIRT

Rankings sind wissenschaftsfeindlich

Für die Karriere eines Forschers der BWL zählt nur eines: die Zahl der Aufsätze in einigen Wissenschaftsjournalen. Das führt zwar zu viel Transparenz - aber zu immer weniger Wissenschaft und zu einem stärkeren Elfenbeinturm-Denken. Der wichtige Dialog mit der Praxis bleibt auf der Strecke.

Von Margit Osterloh und Bruno S. Frey

Forschungsleistung wird in den Wirtschaftswissenschaften - aber nicht nur dort - zunehmend anhand von Rankings gemessen. Diese vergeben Positionen danach, wie viel und in welchen Fachzeitschriften jemand publiziert hat. Dahinter steht der Glaube, dass durch möglichst wenige, aber klare Leistungsindikatoren ein positiver Anreiz für wissenschaftliche Leistung geschaffen werde. Außerdem soll der Öffentlichkeit sowie fachfremden Wissenschaftlern eine transparente Rechenschaft über die Verwendung von Forschungsgeldern geliefert werden.

Dennoch sind Rankings mittlerweile in die Kritik geraten. Auf der 74. Wissenschaftlichen Jahrestagung des Verbandes der Hochschullehrer für Betriebswirtschaft wurde ihr Einfluss kontrovers diskutiert (F.A.Z. vom 11. Juni). Der Deutsche Wissenschaftsrat hat jüngst empfohlen, Rankings nicht unabhängig von qualitativen Expertenurteilen zu verwenden. Dennoch kann sich kaum eine Ökonomin oder ein Ökonom erlauben, das einflussreiche Handelsblatt-Ranking zu ignorieren, welches aufgrund der Anzahl von Veröffentlichungen der Forscher in einflussreichen Journal zustande kommt. Rankings (wie das auf automatisierter Basis erstellte internationale Ranking des ISI Web of Science) spielen heute eine wichtige Rolle bei Berufungen, bei der Besoldung und bei der Zuteilung von Forschungsressourcen.

Damit folgt man der gängigen Praxis, Gegenstände in Rangfolgen zu bringen und Hitlisten über Superstars, Bücher, Restaurants oder Fussballclubs zu veröffentlichen. Aber es gibt einen Unterschied. Jedermann kann ein Restaurant- oder Bücherranking selber überprüfen, indem er das Restaurant besucht oder die Bücher liest. Diese einfache Nachprüfbarkeit ist in der Wissenschaft nicht gegeben.

Wissenschaftliche Ergebnisse müssen ständig von Fachleuten in Frage gestellt und neu

diskutiert werden. Kontroverse Debatten - wie die Kritik der Volkswirtschaftslehre nach der Finanzmarktkrise - gehören zum Kern wissenschaftlicher Arbeit. Diese kontroverse Debatte zeigt aber auch, dass Fachleute oft irren. Es gibt viele Beispiele, dass Forschungsergebnisse von späteren Nobelpreisträgern lange Zeit von den Kollegen zurückgewiesen wurden und nicht veröffentlicht werden konnten. Statistiken zeigen, dass die Korrelation zwischen Gutachten mitunter nicht höher ist als die einer Zufallsauswahl. Sie ist allerdings höher bei Ablehnungen als bei der Annahme von Arbeiten. Schlechte Arbeit wird also leichter identifiziert als Spitzenleistung. Dennoch hat die Wahl der Gutachter einen entscheidenden Einfluss auf Annahme oder Ablehnung einer Publikation. Die Hoffnung, dass durch die Aggregation vieler Gutachter-Urteile mittels Rankings ein Fehlerausgleich zustande kommt, ist trügerisch. Eine Analyse der International Mathematical Union sowie zahlreiche empirische Studien legen dar, dass der Schluss vom Ranking einer Zeitschrift auf die Qualität einzelner darin veröffentlichter Aufsätze zu einer hohen Fehlerwahrscheinlichkeit führt.

Trotz dieser fragwürdigen Grundlage verlassen sich viele Forschende auf Rankings anstatt auf ihr eigenes Urteil. Sie lassen urteilen - mit erheblichen unbeabsichtigten Nebenfolgen. Schuld daran ist nicht ihre Faulheit, sondern ein "Lock-in"-Effekt. Haben sich einmal bestimmte Rankings als Verteilschlüssel für Ressourcen und Reputation etabliert, ist es für eine Fakultät sinnvoll, nur solche Kolleginnen und Kollegen zu berufen, welche im Ranking hoch angesiedelt sind. Das eigene Urteil spielt dann keine Rolle mehr.

Es tritt das Paradox der Leistungsmessung ein. Das Paradox besagt, dass alle Leistungsindikatoren mit der Zeit die Fähigkeit verlieren, gute von schlechten Leistungen zu unterscheiden, und dies umso mehr, je eindeutiger sie sind. Diese Wirkung entfaltet sich idealtypisch in drei Schritten. Leistungsindikatoren rufen erstens einen positiven Lerneffekt hervor. Sie machen deutlich, dass für gute Wissenschaft gute Publikationen wichtig sind. In der Tat wird heute viel mehr als in der vergangenen Forschergeneration in begutachteten Zeitschriften publiziert; so viel, dass nur noch die Hälfte aller Aufsätze in wissenschaftlichen Zeitschriften von anderen als den Gutachtern gelesen werden. Jedoch treten zweitens auch perverse Lerneffekte auf, welche allerdings bei komplexen Sachverhalten schwer von den positiven Lerneffekten unterschieden werden können. Perverse Lerneffekte treten dann auf, wenn man den Fokus auf den Leistungsindikator legt und nicht auf das, was er messen soll. Menschen sind besonders kreativ, wenn es darum geht, bei Kennzahlen gut abzuschneiden, ohne die tatsächlich relevante Leistung zu erhöhen. Beispiele in der Wissenschaft finden sich zahlreich, etwa die Vernachlässigung der Lehre und der akademischen Selbstverwaltung, die Auswahl von Forschungsthemen danach, ob leicht erreichbare Daten vorliegen oder die

Vernachlässigung von Buchpublikationen, weil diese nicht in Rankings gezählt werden.

Der dritte Schritt besteht darin, dass die Indikatoren wegen dieser Effekte ständig angepasst werden müssen. Dies kann aber nur durch erfahrene und erfolgreiche Fachleute vorgenommen werden, die selber im betreffenden Gebiet tätig sind. Diese haben allerdings eher einen Anreiz, eine solche Anpassung zu verhindern. Sie würden sonst die Kriterien unterminieren, nach denen sie Erfolg errungen haben. Im Ergebnis führt bei komplexen Sachverhalten wie der Forschung die Messung der Leistung mittels eindimensionaler Indikatoren tendenziell zu einer Verringerung der Aussagekraft der Leistungsmessung - paradoxerweise um so mehr, je mehr man dem Ruf nach "wenigen und klaren Indikatoren" nachkommt.

Für die rankinggesteuerte Wirtschaftswissenschaft hat dies nachteilige Folgen. Die erste und schädlichste Folge besteht darin, dass der wissenschaftlich Diskurs und der organisierte Skeptizismus - das Herzstück allen wissenschaftlichen Arbeitens - unterminiert werden. Statt besserer Argumente zählt die Position im Ranking.

Zweitens werden Interdisziplinarität, Vielfalt von Forschungsansätzen sowie der Dialog mit der Öffentlichkeit und der Praxis behindert. Wissenschaftlerinnen und Wissenschaftler konzentrieren sich auf das, was in den Publikationsrankings zählt. Da dabei Artikel in populärwissenschaftlichen Organen nicht berücksichtigt sind, werden nur noch Artikel in Fachzeitschriften geschrieben, welche auf eine spezialisierte Leserschaft abstellen. Überdies lassen sich unorthodoxe Beiträge schlechter in führenden Zeitschriften unterbringen, was zusätzlich zu einer Homogenisierung der Forschung führt. Homogenität der Forschung führt jedoch zu Forschungs-Silos und zu einer Überschätzung der eigenen Modelle und Ansätze. Eine solche Überschätzung wurde von zahlreichen Kritikern als eine Ursache der Finanzmarktkrise analysiert.

Drittens wird die für gute Forschung unabdingbare intrinsisch motivierte Neugier ("taste for science") durch ein wissenschaftsfremdes Ziel verdrängt, die Maximierung von Publikationen in Top-Zeitschriften ("taste for publications"). Das geht bis zur akademischen Prostitution. Wie empirische Befunde zeigen, passen Autoren zu einem beachtlichen Prozentsatz ihre Papiere den Gutachter-Vorstellungen an, auch wenn sie überzeugt sind, dass diese falsch sind.

Viertens werden Forschungsmittel zugunsten rankingrelevanter Aktivitäten umverteilt, weil kleine Unterschiede eine große Bedeutung erhalten. Dies ist ablesbar an Titeln wie "Mannheimer Volkswirte drängen Zürich von Platz 1" oder "Bei den Volkswirten geht es zu wie in der Fußball-Champions-League". Deshalb fließt neuerdings viel Geld in den

Einkauf von Stars - nicht selten zu Lasten des Nachwuchses. Dadurch treten Mathhäus-Effekte auf im Sinne des "Wer hat, dem wird gegeben". Dies erklärt, warum vergangene Rankings die stärkste Vorhersagekraft für gegenwärtige Rankings haben. Darüber hinaus entsteht die Gefahr, dass Forschungsimperien entstehen, für welche zusätzliche Ressourcen mit einem abnehmenden Grenznutzen einhergehen.

Was kann man angesichts einer solchen Situation tun? Methodische Verbesserungen mögen die technische Qualität von Rankings heben. Jedoch können sie weder die Eindimensionalität und Abstraktion vom Kontext reduzieren, die per Definition zu jedem Ranking gehören, noch können sie die unbeabsichtigten, perversen Nebenwirkungen verhindern. Ein Boykott durch Individuen ist schwierig. Insbesondere junge Forschende können sich der "Lock-in"-Situation von Rankings nicht entziehen, wenn sie Karriere machen wollen. Vielmehr muss auf mehrfachen Ebenen jenseits von Individuen angesetzt werden.

Erstens muss von den Wissenschaftsorganisationen dafür gesorgt werden, dass einzelne Rankings keine dominante Rolle erhalten. Zwar kann man Zeitungen oder anderen kommerziellen Ranking-Produzenten ihre Tätigkeit nicht verbieten. Aber wissenschaftliche Gemeinschaften wie der Wissenschaftsrat oder die Vereinigungen von Hochschulprofessoren dürfen sich weder daran beteiligen noch eigene Rankings durchführen. Sie tragen sonst mit einem "amtlichen Gütesiegel" zur Dominanz einzelner Rankings bei.

Zweitens dürfen Rankings nie ohne eine Einbettung in qualitative Expertenurteile verwendet werden. Dies fordert auch der Deutsche Wissenschaftsrat. In sogenannte "informed peer reviews" können im Gegensatz zu Rankings viele unterschiedliche Gesichtspunkte und Kontextfaktoren einfließen. Das Paradox der Leistungsmessung kann vermieden werden.

Drittens sollten "informed peer reviews" auf wenige karriererelevante Entscheidungen beschränkt sein. Hierbei spielen Berufungen eine überragende Rolle. Tritt an die Stelle einer kontinuierlichen Bewertung durch Rankings ein sorgfältiges Auswahlverfahren mit gelegentlichen Zwischenevaluationen - wie dies den Prinzipien an der Harvard-Universität entspricht - dann wird der Spielraum für unorthodoxe Forschung erweitert. Schließt das Auswahlverfahren auch die Einschätzung der wissenschaftlichen Neugier ein, dann kann mit einer dauerhaften Kreativität anstelle bloßer Maximierung der Publikationen gerechnet werden.

Diese Vorschläge haben den Nachteil, dass sie Qualität der Forschung für die

Öffentlichkeit nicht leicht erkennbar machen. Diesem Zweck dient ein weiterer Vorschlag, dem verstärkten Einsatz von Auszeichnungen. Sie befriedigen das Bedürfnis nach Sichtbarkeit außerhalb der engen Fachgemeinschaft, erfordern aber eine mehrdimensionale Würdigung einer Leistung nach kontextspezifischen Kriterien. Werden allerdings - wie leider üblich geworden - Auszeichnungen an Forschende vor allem deshalb vergeben, weil ihre Arbeit in Top-Zeitschriften an der Spitze von Rankings veröffentlicht wurde, dann werden die geschilderte kontraproduktiven Wirkungen eher verstärkt.

Alle diese Vorschläge tragen der Tatsache Rechnung, dass Wissenschaft nicht wie eine Hitparade funktioniert. Sie haben den Vorteil, dass sie zur dezentralen und diskursiven Beurteilung von Forschungsleistungen veranlassen. Sie sind zwar alles andere als fehlerfrei, vermeiden aber homogene Kriterien und blinde Flecken in der Forschung eher als eindimensionale Rankings. Sie fördern hingegen die dringend nötige Vielfalt gerade deshalb, weil sie nicht beanspruchen, objektiv und präzise zu sein. Wissenschaft floriert, wenn organisierter Skeptizismus und lebendiger Diskurs anstelle von vorgeblicher Gewissheit floriieren.

Margit Osterloh ist Professorin an der Warwick Business School, University of Warwick, Großbritannien. Sie war bis 2011 Mitglied des Deutschen Wissenschaftsrates. Bruno S. Frey ist Distinguished Professor of Behavioral Science an der Warwick Business School, University of Warwick. Beide sind Forschungsdirektoren von CREMA (Center for Research in Economics, Management and the Arts, Zürich).

Kastentext:

Rankings - gemessen an der Zahl der Veröffentlichungen - spielen heute eine wichtige Rolle bei Berufungen, bei der Besoldung und bei der Zuteilung von Forschungsressourcen.

Die für gute Forschung unabdingbare intrinsisch motivierte Neugier wird durch ein wissenschaftsfremdes Ziel verdrängt, die Maximierung von Publikationen in Top-Zeitschriften.

Auszeichnungen befriedigen das Bedürfnis nach Sichtbarkeit außerhalb der engen Fachgemeinschaft, erfordern aber eine umfassende Würdigung der Wissenschaftsleistung.

Bildunterschrift: Illustration Peter von Tresckow

Alle Rechte vorbehalten © Frankfurter Allgemeine Zeitung GmbH, Frankfurt am Main
Vervielfältigungs- und Nutzungsrechte für F.A.Z.-Inhalte erwerben Sie auf www.faz-rechte.de