



WESTFÄLISCHE
WILHELMS-UNIVERSITÄT
MÜNSTER



Centrum für Interdisziplinäre Wirtschaftsforschung
Diskussionspapier

1/2013

Zur Ethik von Rankings im Hochschulwesen
Eine Betrachtung aus ökonomischer Perspektive

Harry Müller

Center for Interdisciplinary Economics
Discussion Paper

Centrum für Interdisziplinäre Wirtschaftsforschung
Diskussionspapier
1/2013

Februar 2013

ISSN 2191-4419

Zur Ethik von Rankings im Hochschulwesen
Eine Betrachtung aus ökonomischer Perspektive

Harry Müller

Zusammenfassung

In der kontrovers geführten Debatte über den Sinn und Unsinn von Rankings und über die Validität bibliometrischer Indikatoren wird immer wieder auch auf unintendierte Konsequenzen von Rankings für das Wissenschaftssystem hingewiesen. Vor diesem Hintergrund wird sowohl deren Erstellung und Publikation als auch deren Verwendung im Rahmen von hochschulinternen oder hochschulpolitischen Entscheidungsprozessen zu einem ethischen Problem. Dem Informationsbedarf von Hochschulleitungen und externen Stakeholdern sowie dem Erkenntnisinteresse von Hochschulforschern steht die Forderung der Wissenschaftler nach akademischer Freiheit, die auch als eine Freiheit von einem ständigen Controlling durch Rankings verstanden werden kann, entgegen. Nur durch einen ethisch reflektierten Umgang mit Rankings kann vermieden werden, dass sich durch sie ein unerwünschter Paradigmenwechsel im Publikationsverhalten verfestigt, der nachträglich wohl nur noch schwer zu korrigieren sein dürfte.

JEL-Codes: I23, M00, M50, Z10

On the Ethics of Rankings in Higher Education

An Analysis From a Business Administration Perspective

Abstract

In the controversial debate about the pros and cons of rankings and the validity of bibliometric indicators, attention is also drawn to the unintended consequences of rankings for the science system. Against this background, both the creation and publication of rankings and its use in institutional or political decision-making lead to ethical problems. The information needs of university managements and external stakeholders as well as the scientific interest of higher education researchers contrast the academic freedom of the professors, who claim to be independent from extensive monitoring by rankings. Therefore, rankings should be handled with ethical reflection. Otherwise they might cause an undesirable paradigm shift in publishing behaviour that might probably be hard to correct afterwards.

Im Internet unter:

http://www.wiwi.uni-muenster.de/ciw/forschen/downloads/DP-CIW_01_2013.pdf

Westfälische Wilhelms-Universität Münster
CIW – Centrum für Interdisziplinäre Wirtschaftsforschung
Scharnhorststraße 100
D-48151 Münster

Tel: +49-251/83-25329 (Sekretariat)
E-Mail: clementine.kessler@uni-muenster.de
Internet: www.wiwi.uni-muenster.de/ciw

Zur Ethik von Rankings im Hochschulwesen

Eine Betrachtung aus ökonomischer Perspektive*

1. Die Aktualität von Rankings

Die deutschsprachige Betriebswirtschaftslehre (BWL) wurde, wie auch ihre sozialwissenschaftlichen Schwesterdisziplinen, in den vergangenen Jahren Gegenstand methodologisch vielfältiger Rankings. Auch wenn das Handelsblatt-BWL-Ranking (siehe Müller/Storbeck 2009) das in jüngster Zeit am meisten diskutierte Projekt ist, existieren zahlreiche, methodologisch anders konstituierte Alternativen. Laut Dilger (2010, S. 92) sind Rankings Verfahren, mit denen „ganz allgemein Beobachtungseinheiten nach einem Kriterium in eine Reihenfolge gebracht“ werden. Entlang dieser beiden Charakteristika, nämlich (1) welche Beobachtungseinheiten (2) nach welchem Kriterium gerankt werden, lässt sich die Vielzahl der in den vergangenen Jahren diskutierten Rankingansätze systematisieren. Dabei kann zudem noch zwischen Forschungs- und Lehrrankings differenziert werden, wobei im Rahmen der folgenden Betrachtung vornehmlich erstgenannte im Fokus stehen sollen. Abbildung 1 strukturiert eine exemplarische Auswahl der in den Wirtschaftswissenschaften derzeit am meisten diskutierten Ansätze gemäß den obengenannten Dimensionen.

		Rankingkriterium				
		Lehre		Forschung		
		Lehrevaluation/ Befragung	Verdienst Absolventen	(Zeitschriften-) Publikationen	Zitationen	Drittmittel
Beobachtungseinheit	Einzelne Personen	Lehrpreis der Fakultät		Handelsblatt	Podsakoff u. a. (2008)	
	Fachbereiche	CHE-Ranking, Wirtschaftswoche	Financial Times Ranking	Wolf/Rohn/Macharzina (2006), Handelsblatt	Müller (2012b)	
	Hochschulen			Shanghai Ranking ¹	Podsakoff u. a. (2008)	DFG Förder- atlas
	Länder	CHE-Länder- ranking			Müller (2012b)	

Abb. 1: Typologie von Rankings mit ausgewählten Beispielen

* Ich danke Alexander Dilger und den Teilnehmer(inne)n des gemeinsamen Workshops der wissenschaftlichen Kommissionen „Wissenschaftstheorie und Ethik in der Wirtschaftswissenschaft“ und „Hochschulmanagement“, der vom 16.-18.2.2012 an der Freien Universität in Berlin stattfand, für wertvolle Hinweise und Kommentare.

¹ Das Shanghai-Ranking ist multidimensional und berücksichtigt neben der Anzahl der Veröffentlichungen auch deren Zitationen und weitere Kriterien.

In der Orientierung an Rankings und anderen Verfahren zur Messung von Forschungs- und Lehrleistungen manifestiert sich ein allgemeiner Paradigmenwechsel im Hochschulmanagement. Während traditionell ein inputorientiertes, verwaltungsmäßiges Steuerungsmodell an Hochschulen dominierte, wurden in den letzten Jahren verstärkt Konzepte des New Public Management auf den Hochschulbereich übertragen (siehe Lange 2008). Hierbei werden zur Steuerung anstatt der Inputfaktoren der Output oder der Outcome betrachtet. Zur Begründung wird auf die Neue Institutionenökonomik verwiesen, gemäß der die untergebenen Stellen aufgrund ihres Informationsvorsprungs am besten dazu befähigt seien, die Mittel zur Erreichung der vorgegebenen Ziele (Output) zu wählen, anstatt auch diese ex ante zu definieren, wie dies bei der Inputsteuerung der Fall ist. Die Universität und ihre Beschäftigten würden durch den Wandel von der Input- zur Outputsteuerung von bürokratischen und verwaltungstechnischen Vorgaben „entfesselt“ (Müller-Böling 2000) und sollen auf diese Weise zu größeren Leistungen befähigt werden. Die Ideen des New Public Management bilden den theoretischen Hintergrund für viele hochschulpolitische Reformen der letzten Jahre wie etwa die Einführung der W-Besoldung (siehe Blomeyer 2007) und pauschalierter Budgets oder die Nutzung von Zielvereinbarungen.

Kritikerinnen und Kritiker² sehen in diesem Paradigmenwechsel jedoch auch Gefahren: Die Outputorientierung des Managements erfordere ein immer größeres Maß an Mess- und Kontrollsystemen, deren Betrieb nicht nur Ressourcen verschlingt, sondern auch das Funktionieren sozialer Systeme direkt behindere. Power (1997), der dieses Problem gesamtgesellschaftlich untersucht, sieht darin eine Entwicklung zur „Audit Society“. In Bezug auf den Hochschulkontext argumentieren Kritiker, dass die klassische Universitätsidee sowohl durch Rankings als auch durch hochschulinterne Leistungsvergleiche beschädigt werde (siehe Erne 2007).

Im Hinblick auf die Forschungsleistungen begann die Evaluation der Outputs in den englischsprachigen Naturwissenschaften, die in der Bibliometrie nach wie vor eine Vorreiterrolle einnehmen. Mit der Entwicklung des *Science Citation Index* wurde Garfield (1964) zum Pionier der Bibliometrie, die sich dann in den Folgejahren in die Geistes- und Sozialwissenschaften zunächst des englischen Sprachraums und später dann auch in die nicht genuin englischsprachigen Wissenschaftslandschaften Kontinentaleuropas ausbreitete. Aus der Perspektive der deutschsprachigen Wirtschafts- und Sozialwissenschaften sind die Natur-

² Aus Gründen der besseren Lesbarkeit wird im Folgenden das generische Maskulinum verwendet, das die weibliche Form jeweils mit berücksichtigt.

wissenschaften und der angloamerikanische Sprachraum in der beschriebenen Entwicklung also voraus, was den Vorteil hat, dass sowohl mögliche positive als auch negative Effekte der zunehmenden Outputorientierung in Bezug auf Forschungsleistungen mit einem Blick ins Ausland exemplarisch betrachtet werden können, noch bevor entsprechende Instrumente hierzulande praktisch etabliert werden. Für die empirische Untersuchung von Publikationen, Autoren und Institutionen mittels statistischer Verfahren hat sich die Bibliometrie als Spezialwissenschaft herausgebildet.

2. Rankings verändern Verhalten

Sowohl die Befürworter der beschriebenen Entwicklung als auch ihre Kritiker unterstellen, dass Rankings das Verhalten der Gerankten zum Guten oder zum Schlechten verändern. Auch wenn in vielen Fällen kein direkter Bezug zwischen den Erstellern von Rankings und den gerankten Personen oder Institutionen besteht, erzeugen Rankings bisweilen unintendierte Anreize zur Verhaltensänderung. Obwohl Rankingergebnisse auf den ersten Blick unpolitisch erscheinen und vielleicht aus ganz anderen Gründen, etwa dem Erkenntnisinteresse von Hochschulforschern, motiviert sind, können von ihnen starke Anreize ausgehen, weil sie Reputation auf die gerankten Personen oder Hochschulen verteilen. Dadurch kann ein Motivationseffekt erzeugt werden, der die Gerankten zu intensiveren Leistungen anspricht.

Demgegenüber stehen jedoch Anreize zu Verhaltensanpassungen mit negativen Folgen. So wird in der angloamerikanischen Literatur darauf hingewiesen, dass die Dominanz der Publikationsstrategie über den Inhalt eine Folge der Outputorientierung sei (siehe Lawrence 2007, Todd/Ladle 2008). Die Ergebnisse eines Forschungsprojekts würden in so viele Zeitschriftenbeiträge wie möglich zerteilt, um die Zahl der Publikationen zu maximieren. Diese auch als „Salamitaktik“ bezeichnete Strategie wirkt sich nicht nur in publikationsbasierten, sondern auch in zitationsbasierten Rankings (zur Differenzierung siehe Müller 2012a) positiv auf das Ergebnis aus, weil jeder einzelne Zeitschriftenbeitrag auch einzeln zitiert werden kann und ggf. sogar muss, wenn sich ein anderer Beitrag fundiert mit dem (aus strategischen Gründen fragmentierten) Gesamtwerk auseinandersetzen möchte. Daneben kann die Outputorientierung dazu führen, dass risikoärmere Forschungsprojekte attraktiver erscheinen als solche mit sehr ungewissem Ergebnis. Zeitschriftengutachter haben typischerweise eine Präferenz für am Status quo orientierte Arbeiten, während wirklich bahnbrechende Erkenntnisse von ihnen nur schwer als solche zu erkennen sind und oftmals zu Unrecht aussortiert werden (vgl. Benos 2007). Ebenso ist die Orientierung an Wissenschaftsmoden (siehe Crane

1969) im Hinblick auf Rankings attraktiv, da mit viel Aufmerksamkeit gerechnet werden kann, die sich in Publikationsgelegenheiten und später in Zitationen niederschlägt.

In einem gewissen Maße wäre auch eine sachlich unbegründete Arbitrage hin zu Forschungsfeldern mit besser gerankten Zeitschriften oder höheren Zitationsraten möglich: Zwar wird deswegen beispielsweise kein Historiker zu medizinischen Themen publizieren; einer medizinhistorischen Forschungsarbeit wird aber in auf Impact-Faktoren basierenden Rankings ein viel höheres Gewicht beigemessen, wenn sie statt in einer historischen in einer medizinischen Fachzeitschrift erscheint, weil die Impact-Faktoren in den Naturwissenschaften aufgrund der unterschiedlichen Publikations- und Zitationskulturen bis zu zehnmal höher sind als in den Geisteswissenschaften (vgl. Althouse u. a. 2009). Dieser Bewertungsunterschied wäre nicht durch einen Leistungsunterschied begründet, sondern lediglich das Ergebnis strategischen Publizierens. Eine solche interdisziplinäre Arbitrage dürfte zwar nur in Ausnahmefällen wie dem hier skizzierten möglich sein, innerhalb eines Fachs ist das Problem jedoch ungleich relevanter: Beispielsweise existieren auch innerhalb der Betriebswirtschaftslehre unterschiedliche Publikations- und Zitationskulturen in den jeweiligen Teilfächern (vgl. Dilger/Müller 2012). Weil die Betriebswirte oftmals in mehr als einem Teilfach tätig sind, besteht hier weit mehr Spielraum für mögliche Fehlanreize.³

Wenn strategische Überlegungen bzgl. der Publikationsmedien die Inhalte eines Forschungsprojekts dominieren, handelt es sich um unethisches Verhalten, weil das Streben nach Wissen und Wahrheit nicht mehr primäres Ziel ist, sondern durch die strategische Orientierung an Rankings eingeschränkt wird. Davon abzugrenzen ist Betrug, das vorsätzliche Überschreiten juristischer Grenzen. Diese Grenzen müssen nicht immer durch Gesetze definiert sein, wie am Beispiel des Plagiats deutlich wird. Es verstößt nicht nur gegen die Praxis guter Wissenschaft, sondern auch gegen die Promotions- oder Habilitationsordnungen der jeweiligen Hochschulen, i. d. R. gegen die vertraglichen Vereinbarungen zwischen Autor und Verlag und z. T. sogar gegen das Urheberrecht. Ähnliches gilt für fingierte Daten oder manipulierte Statistiken bei empirischen Arbeiten, die ebenfalls als Betrug zu bezeichnen sind. Beide Phänomene wurden in der jüngsten Vergangenheit auch im Zusammenhang mit Rankings öffentlich: So

³ Diese ergeben sich jedoch nicht nur bei zitationsbasierten Rankings, sondern insbesondere auch aus dem Jourqual-2 Zeitschriftenranking des Verbands der Hochschullehrer für Betriebswirtschaft (VHB), das beispielsweise im Teilfach Hochschulmanagement kein mit A+ oder A bewertetes Publikationsorgan ausweist und nur eine einzige mit B gerankte Zeitschrift enthält, während im Marketing vier Zeitschriften mit A+, sechs mit A und elf mit B bewertet sind. Da das Jourqual-Ranking nicht nur Reputation zuweist, sondern auch im hochschulinternen Controlling bzw. in kumulativen Promotions- und Habilitationsverfahren Verwendung findet, setzt es einen inhaltlich unbegründeten (Fehl-)Anreiz, z. B. Themen des Hochschulmanagements zu meiden, weil damit keine Spitzenplatzierungen erreichbar sind.

wurden einem im Handelsblatt-VWL-Ranking prominent platzierten Ökonomen wiederholt Selbstplagiate nachgewiesen, während ein vom Handelsblatt gefeierter Betriebswirt inzwischen mehrere Veröffentlichungen widerrufen musste, nachdem ihm zumindest Unrichtigkeiten in der zugrunde liegenden Empirie nachgewiesen worden waren. Derartiges Fehlverhalten wird durch Rankings belohnt, sofern die Unregelmäßigkeiten, wie in den beiden genannten Beispielen, nicht später öffentlich angeprangert werden. Jedenfalls dürfte es kein Zufall sein, dass die beiden Genannten im Ranking so prominent platziert sind, erleichtern doch die ihnen vorgeworfenen Arbeitsmethoden die Produktion einer Vielzahl von Veröffentlichungen. Auf diese Weise setzen Rankings nicht nur Fehlanreize im Hinblick auf die Themenfindung und auf bestimmte Publikationsstrategien, sondern reizen auch zu betrügerischem Verhalten an, auch wenn dies ihrem eigentlichem Ziel, Forschungsleistungen mess- und vergleichbar zu machen, vollkommen zuwiderläuft. Die Situation ist vergleichbar mit Wettbewerben im Sport, die ebenfalls Spitzenleistungen hervorbringen sollen, damit aber zugleich Anreize für Doping schaffen.

Wenn Rankings das Wissenschaftssystem in unvorteilhafter Weise verändern können, weil sie potenziell schädliche Verhaltensanreize setzen, folgt daraus die Frage nach ethischen Prinzipien bei ihrer Erstellung und im Umgang mit ihren Ergebnissen. Weil Rankingergebnisse auf die Wissensproduktion, die sie abbilden sollen, zurückwirken, besteht seitens der Ersteller oder Hochschulforscher hier eine Verantwortung für die möglichen Folgen. Aus diesem Grund sollen im nächsten Abschnitt zunächst die divergierenden Interessen der verschiedenen Stakeholder in diesem Prozess herausgearbeitet werden.

3. Der „Markt“ für Forschung

Aus der Sicht der ökonomischen Theorie sind Forschungsleistungen keine marktgängigen Güter, weil ihnen die typischen Eigenschaften von privaten Gütern (Teilbarkeit, Rivalität im Konsum etc.) in der Regel fehlen (vgl. Grossekketter 2007). Eine Analyse mit Hilfe des klassischen Marktmodells ist daher nicht möglich, schon weil es hier, außer im Falle der Auftragsforschung, keinen Preis gibt. Trotzdem lässt sich auch der wissenschaftliche Austausch als Interaktion von Anbietern und Nachfragern beschreiben. Die Autoren bieten ihre Forschungsergebnisse in ihren Publikationen zum Lesen an, andere Wissenschaftler rezipieren die Ergebnisse und nutzen sie für ihre eigene Forschung. Diese Interaktion wird durch das wissenschaftliche Zitat des rezipierten Texts dokumentiert.

Betrachtet man die Struktur der Interaktionsbeziehungen näher, so können Aussagen im Hinblick auf die ggf. konkurrierenden Interessen der beteiligten Stakeholder gemacht werden. Vergleichsweise eindeutig ist der Fall der Auftragsforschung, bei dem die Antwort auf eine vom Auftraggeber gestellte Frage direkt entgolten wird. Der Nutzen fällt beim Auftraggeber direkt an und spiegelt sich in dessen Zahlungsbereitschaft. Ein wie auch immer geartetes Forschungsleistungsranking wäre in einem solchen Fall unnötig, weil die Bewertung durch den Markt vorgenommen wird. Die Qualität der Auftragsforschung liegt in der Nützlichkeit für den Auftraggeber, so dass eine externe Kontrolle durch Kollegen und auch die Veröffentlichung grundsätzlich verzichtbar sind.

Verglichen damit sind die Beziehungen im Bereich der Grundlagenforschung oder bei der Erforschung von Fragestellungen von allgemein-gesellschaftlichem Interesse deutlich komplexer. Bei wissenschaftlichen Publikationen sind die Nachfrager oft ihrerseits wiederum Produzenten, die an Bestehendes anknüpfen und existierende Erkenntnisse ausbauen oder verwerfen. Der gesellschaftliche Nutzen der Wissenschaft fällt im Gegensatz zur Auftragsforschung extern an. Der Wissenszuwachs als Outcome des Forschungsprozesses kommt einer Vielzahl unterschiedlicher Akteure zugute, die sich zudem ex ante, also bevor das entsprechende Wissen erzeugt wurde, nicht bestimmen lassen. Der Wissenszuwachs hat also Kollektivgutcharakter, was in der volkswirtschaftlichen Theorie als Argument für dessen staatliche Finanzierung angeführt wird.

Aus der Sicht von Staat und Gesellschaft als Financiers der Forschung existiert ein Interesse an der Sichtbarmachung des Produktionsprozesses. Die Frage, wer wie viel zum Kollektivgut Forschung beiträgt, kann für die Akteure der Hochschulpolitik in den Ministerien wie auch für private Drittmittelgeber eine entscheidungsrelevante Information darstellen. Publikationsdaten, Rankings und Leistungsvergleiche können, sofern sie sachgerecht angewandt und interpretiert werden, in dieser Hinsicht nützliche Informationen bereitstellen. Daneben hat auch die Hochschulforschung ein genuin wissenschaftliches Interesse an der Erforschung der Forschenden.

Dieses Interesse kann jedoch im Konflikt mit der akademischen Freiheit der Forscher stehen. Akademische Freiheit kann nämlich auch als Freiheit von externer Rechenschaftspflicht verstanden werden (vgl. Wicks 2004, Nkomo 2009). Auch wenn Rankings und bibliometrische Untersuchungen nicht zur direkten, monetären Steuerung einzelner Wissenschaftler durch die Hochschulleitungen zum Einsatz kommen, können auch durch Dritte erstellte Rankings, die keine unmittelbaren materiellen Auswirkungen auf die Gerankten haben, deren intrinsische

Motivation verdrängen (vgl. hierzu Frey 2003) oder Fehlanreize im Hinblick auf die zukünftige Publikationstätigkeit setzen (siehe oben).

Dieser Interessenkonflikt lässt sich nicht ohne Weiteres auflösen, denn auch wenn man etwa akademische Freiheit gegenüber den Aufsichtsinteressen der Hochschulpolitik strikt superior bewertete, bliebe der Konflikt mit der Hochschulforschung bestehen: Ein Verzicht auf Rankings und bibliometrische Untersuchungen wäre aus der Perspektive ihrer Kritiker ein Gewinn an akademischer Freiheit (vgl. Kieser 2010, 2012), für Teile der Hochschulforschung jedoch eine Form der Zensur. Das Argument, ein Verzicht auf Rankings führe zu einem Mehr an akademischer Freiheit, wäre also nicht ohne Einschränkung haltbar. Dieser fundamentale Interessenkonflikt zwischen der bibliometrischen Hochschulforschung und ihrem Erkenntnisobjekt lässt sich daher nicht logisch auflösen, sondern muss durch verantwortungsvolles Handeln stets neu austariert werden. Die Erstellung von Rankings und der Umgang damit werden auf diese Weise auch zu einer ethischen Frage.

4. Qualitätskriterien von Rankings

Generell lassen sich Rankings nicht abstrakt, sondern nur mit Bezug zu ihren z. T. selbst definierten Zielen bewerten (vgl. Dilger 2010). Dabei kann zunächst auf die allgemeinen Gütekriterien der Messtheorie zurückgegriffen werden, nämlich Objektivität, Reliabilität und Validität (vgl. Lienert/Raatz 1998). Ein Ranking sollte zunächst also unabhängig von der Person/Position des Erstellers sein. Zweitens darf die Messung nicht durch Zufallsereignisse beeinflusst werden, muss also reproduzierbar sein. Validität schließlich bedeutet, dass das Ranking tatsächlich das misst, was gemessen werden soll. Sie setzt Objektivität und Reliabilität voraus und ist das kritische der drei Gütekriterien, da es danach fragt, ob die im Ranking verwendeten Kennzahlen auch wirklich das messen, was gemessen werden soll. In diesem Sinne weist das Validitätserfordernis über den rein technischen Bereich der Messtheorie hinaus und definiert auch konzeptionelle Ansprüche an Rankings, nämlich die Wahl von geeigneten Kennzahlen. Dabei ist zu berücksichtigen, dass die Kennzahlen nicht mit der zu messenden Leistung identisch sind, sondern mehr oder weniger gute Indikatoren darstellen. Ob die Produktivität eines Wissenschaftlers durch die Zahl seiner Zeitschriftenaufsätze approximierbar ist oder ob zu diesem Zweck andere Kennzahlen geeigneter erscheinen, wäre eine Fragestellung, anhand derer verschiedene Rankingansätze hinsichtlich ihrer Validität miteinander verglichen werden könnten.

Die Erfordernisse der Messtheorie stellen im Hinblick auf die Qualität von Rankings ein notwendiges, jedoch kein hinreichendes Kriterium dar. Aus ökonomischer Perspektive tritt die Anreizkompatibilität als weiteres Kriterium hinzu (zum Konzept siehe Milgrom/Roberts 1992). Sie bezieht sich auf die Auswirkungen des Rankings auf die Gerankten und fokussiert damit die oben beschriebenen Fehlanreize in Bezug auf das Publikationsverhalten (siehe Abschnitt 2). Nur wenn ein Ranking neben den eher technischen Kriterien der Messtheorie auch anreizkompatibel in Bezug auf alle Beteiligten ist, genügt es den wissenschaftsethischen Erfordernissen in einem strengen Verständnis derselben. Das Qualitätskriterium der Anreizkompatibilität soll mithin sicherstellen, dass Rankings keine negativen Auswirkungen auf das Erkenntnisobjekt haben.

Die Überlegungen machen deutlich, dass es Qualitätskriterien bedarf, um negative Auswirkungen von Rankings zu vermeiden oder zumindest einzugrenzen. Auch wenn ein umfassendes Bewusstsein die Folgen und Probleme von Rankings betreffend noch nicht zu existieren scheint, legte der Wissenschaftsrat (2004) konkrete Empfehlungen zum Umgang mit Rankings vor. Viele Punkte der Empfehlungen erscheinen unmittelbar einsichtig, sie adressieren jedoch nicht die Fehlanreize von Rankings. Die Empfehlungen lassen sich wie folgt zusammenfassen:

- Zielsetzung angeben
- Entscheidungsrelevante Objektdefinition wählen
- Leistungsdimensionen getrennt bewerten
- Leistungskriterien offen legen
- Indikatoren nach Leistungskriterien auswählen
- Hohe Datenqualität und optimale Auswertbarkeit anstreben
- Unabhängige Trägerschaft und Bewertung sicherstellen
- Wettbewerb und Autonomie wahren

Auch wenn die genannten Punkte in den Empfehlungen des Wissenschaftsrats wortreich erläutert werden, so lassen sie sich letztlich auf die drei Forderungen der allgemeinen Messtheorie zurückführen: Das Offenlegen von Zielsetzung und Kriterien, unabhängige Bewertung und Wahrung der Autonomie lassen sich im Punkt Objektivität zusammenfassen. Die Forderung nach Datenqualität und Auswertbarkeit ist Bestandteil der Reliabilität, während die

Trennung der Leistungsdimensionen und die Auswahl der passenden Indikatoren die Validität der Messung erhöhen sollen.

Neben den bereits genannten Punkten empfiehlt der Wissenschaftsrat, bei der Präsentation der Ergebnisse Ranggruppen statt Ranglisten auszugeben und das Produkt dann Rating anstatt Ranking zu nennen. Beide Punkte zielen offenkundig auf die Rezeption von Rankings und müssen daher vor dem Hintergrund der implizierten Anreizwirkungen diskutiert werden. Dabei fällt zunächst auf, dass die Forderung nach Ranggruppen im Widerspruch zur ebenfalls geforderten hohen Datenqualität und der optimalen Auswertbarkeit steht und zudem die Nachprüfbarkeit des Ratings durch Außenstehende unmöglich macht. Schließlich zieht der Verzicht auf Rangplätze notwendigerweise auch einen Verzicht auf die Publikation der rangbildenden Indikatoren nach sich. Die Methodologie kann dann sowohl von Außenstehenden als auch von den Gerankten selbst nicht überprüft werden, das Rating würde gegen Kritik immunisiert. Dabei entbehrt es nicht einer gewissen Ironie, dass zur Begründung dieser Empfehlung auf die jeder Messmethode innewohnenden Ungenauigkeiten und Messfehler verwiesen wird, die eine exakte Zuweisung von Rangplätzen nach Ansicht des Wissenschaftsrats nicht vertretbar machten. Ist es doch gerade der Verzicht auf die Publikation der Einzelwerte, der die Methodik jeglicher Nachprüfbarkeit entzieht und auf diese Weise etwaige Messfehler und Ungenauigkeiten verschleiert.

Abgesehen davon erscheint die Forderung nach Ranggruppen statt Rangplätzen vor dem Hintergrund der in Wissenschaftsrangings typischerweise auftretenden Pareto-Verteilung (Pareto 1896, siehe auch Hardy 2010) gleich doppelt unpassend. Die Pareto-Verteilung der Messergebnisse hat zur Folge, dass die Messwerte nur im Bereich der Spitzengruppe signifikant voneinander abweichen und die Differenzen zwischen den Rangplätzen zunehmend geringer werden. Bei der Verwendung von Ranggruppen werden in der Spitzengruppe, in der zwischen den einzelnen Plätzen Unterschiede messbar wären, Informationen verschenkt, während im Verlauf der Verteilung, wo sich die einzelnen Messwerte kaum noch unterscheiden, Differenzierungsmerkmale behauptet werden, wo keine sind. Nur bei einer Publikation der einzelnen Messwerte kann ein Leser deren zunehmend insignifikanten Unterschiede im Verlauf der Verteilung erkennen, während Ranggruppen an dieser Stelle einen anderen Eindruck erwecken können.

Die Verwendung von Ranggruppen ist schließlich auch nicht geeignet, die Folgen von Messfehlern oder Ungenauigkeiten auf der Ebene der Datenbasis abzumildern oder einzudämmen, wie an einem einfachen Beispiel zu erkennen ist: Werden etwa Ranggruppen in Zehner-

schritten gebildet, so verlöre ein Messfehler, der zur Vertauschung der Ränge 5 und 6 geführt hätte, an Bedeutung, denn beide Objekte wären dennoch korrekt der Ranggruppe zugeordnet. Eine Vertauschung der Plätze 10 und 11 jedoch führte zu umso größerem Schaden, da im Falle der Zehnerranggruppen der Messfehler zehn Rangplätze statt nur eines Platzes betrüge. Im Erwartungswert ergibt sich daraus keine Differenz, denn der Verzicht auf mehrere kleine Fehler wird durch das Auftreten seltenerer, aber dafür größerer Abweichungen erkauft. Die Verwendung von Ranggruppen statt Rangplätzen ist also nicht geeignet, die Auswirkungen von Messfehlern oder Ungenauigkeiten abzumildern.

Die vom Wissenschaftsrat propagierte Sprachregelung Rating statt Ranking erweist sich schließlich ebenfalls als problematisch. Ratings bezeichnen eine Beurteilung ohne Rangfolgenbildung (vgl. Everling 2010), insbesondere im Finanzwesen hinsichtlich der Kreditwürdigkeit von Schuldern. Dieser Ansatz ist jedoch mit dem Konzept der Ranggruppen nur sehr bedingt vereinbar, denn während Ranggruppen ebenso wie Rangplätze einem Objekt jeweils eine relative Position zuweisen, beschreibt ein Rating eine absolut definierte Qualitätseinstufung und setzt eine vorab definierte Skala voraus. Eine solche Skala, wie sie für Kreditratings beispielsweise von AAA (ausfallsicher) bis D (Zahlungsausfall) definiert ist, sehen die Empfehlungen des Wissenschaftsrats jedoch nicht vor. Die Verwendung von Ranggruppen ersetzt eine solche Skala nicht, da sie stets eine relative Position beschreiben, während bei Ratings nicht notwendigerweise ein Objekt mit der besten und eines mit der schlechtesten Note bewertet werden muss. Vor diesem Hintergrund erscheint die Verwendung des Begriffs Rating als eine reine Sprachregelung, die anscheinend der Attribution von Qualität dienen soll. Auch dies ist jedoch inzwischen einigermaßen zweifelhaft, denn während der Begriff Rating im Erscheinungsjahr der Empfehlungen des Wissenschaftsrats (2004) noch mit Reputation verknüpft war, so haben die Ratingagenturen und ihre Kreditbeurteilungen im Zuge der Finanzkrise deutlich an Reputation eingebüßt. Ein Ranking, das begrifflich hier anknüpft, erschiene heute wohl weniger opportun.

Abgesehen von den hier kritisch beurteilten Ranggruppen und der Benennungsfrage enthalten die Empfehlungen des Wissenschaftsrats jedoch viele wichtige Hinweise für eine Qualitätsverbesserung bei der Messung von Forschungsleistungen. Das Problem der unintendierten negativen Konsequenzen in Form von Fehlanreizen wird jedoch noch nicht in ausreichendem Maße adressiert. Auf der Grundlage der Empfehlungen des Wissenschaftsrats wäre eine breite Debatte wünschenswert, in die sich insbesondere die Ersteller von Rankings einbringen

sollten, um den Kriterienkatalog für eine sachgerechte Forschungsleistungsmessung in dieser Dimension zu erweitern.

Derartige Überlegungen sollten sich zunächst am Horizont der Adressaten orientieren und die Möglichkeit einer Falsch- oder Überinterpretation der Ergebnisse bereits ex ante berücksichtigen. Je forschungsferner der Adressatenkreis, desto größer ist diese Gefahr einzuschätzen. Ein internes Bewerberranking im Rahmen eines Berufungsverfahrens wäre von diesem Einwand beispielsweise weniger betroffen als ein öffentliches Ranking wie das Handelsblatt-Ranking, das weit über die Grenzen des Fachs und der Wissenschaft hinaus rezipiert wird. An dieser Stelle ist eine Selbstbeschränkung hinsichtlich der publizierten Informationstiefe sinnvoll und in der Regel auch bereits Praxis. Daher veröffentlicht das Handelsblatt lediglich die Namen der 250 bestplatzierten Personen. Andere Autoren von Rankings schränken die Informationstiefe jedoch noch weit stärker ein. So nennen die Personenrankings von Dilger (2010) und Dilger/Müller (2012) jeweils nur die zehn Erstplatzierten. Das erscheint aus zwei Gründen sachgerecht: Erstens sinkt die Signifikanz der Rangdifferenzen im Verlauf der Verteilung (s. o.), sodass dem genauen Rangplatz außerhalb der Spitzengruppe ohnehin kaum Informationswert zukommt. Da dieser Umstand aber vielen, insbesondere außerwissenschaftlichen Rezipienten kaum bekannt sein dürfte, verhindert der Verzicht auf die übrigen Rangplätze Fehlinterpretationen. Zweitens birgt die Namensnennung auf den mittleren und hinteren Plätzen die Gefahr eines negativen Motivationseffekts für die Genannten. Eine gewisse Bescheidenheit im Hinblick auf die Informationstiefe erscheint aus den genannten Gründen ethisch geboten, auch wenn dies notwendigerweise einen Verzicht auf die maximale Publicitywirkung bedeutet. Ebenso selbstverständlich sollte es für die Ersteller von Rankings sein, auf Limitationen hinsichtlich der Validität des eigenen Ansatzes hinzuweisen. Das erscheint schon aus dem Grunde angemessen, weil sich ein Wissenschaftsranking an den Qualitätsstandards seines Erkenntnisobjekts orientieren sollte, um im Urteil der Gerankten bestehen zu können.

5. Ausblick

Zu diskutieren ist schließlich, ob ein zentral (ggf. vom Wissenschaftsrat) erstelltes, gewissermaßen offizielles Ranking sinnvoll wäre. Ein solches Projekt könnte sich fachübergreifend an den Best Practices orientieren und die technischen Kompetenzen des Wissenschaftsmanagements bzw. der Bibliometrie mit der inhaltlichen Expertise der jeweiligen Fächer verbinden. Abgesehen davon, dass ein solches Vorgehen nicht gegen den Widerstand der

Fachvertreter durchsetzbar ist – man denke z. B. an die Ablehnung von Rankings durch die Fachverbände der Soziologen und Historiker – und in einem solchen Fall mangels interessierter Adressaten auch vollends überflüssig wäre, ist zu berücksichtigen, dass ein offziöses Ranking angesichts des derzeitigen Entwicklungsstandes der Methodologie durchaus Gefahren birgt. Wenn die Validität der angewandten Messinstrumente noch zu diskutieren ist, erzeugt Uniformität keinen Fortschritt, sondern verhindert eine produktive Debatte.

Die Akzeptanz eines Rankings kann seine mangelnde Validität nicht ersetzen. Vor diesem Hintergrund erscheinen vielmehr die Entwicklung von alternativen Verfahren und die Kritik daran geeignet zu sein, die Qualität von Rankings zu verbessern. Das gilt insbesondere, wenn man berücksichtigt, dass erfolgreiche Rankings einen Paradigmenwechsel fortifizieren können, der zunehmend irreversibel wird. Sobald sich die Gerankten bei ihrer Leistungserbringung an den Kriterien der Rankings orientieren, trägt dies zu einer Stabilisierung des Rankings und der Verdrängung von alternativen Messverfahren bei. Weil dieser Prozess allein durch die Akzeptanz der Rankings in der Zielgruppe getrieben wird und diese nicht notwendigerweise mit der Qualität des Verfahrens zusammenfallen muss, erscheint ein zu frühes Festlegen eines zentralen Standards wenig sinnvoll zu sein. Langfristig können Rankings die Realität, die sie messen wollen, selbst erzeugen, wenn sich die Gerankten mehr und mehr an den vom Ranking postulierten Leistungskriterien orientieren. Angesichts dessen besteht zum aktuellen Zeitpunkt wenig Bedarf nach einem offiziellen, allgemein akzeptierten Ranking; vielmehr sollte das Thema noch eine Weile Experimentierfeld der Hochschulforschung bleiben, bis sich – im wissenschaftlichen Wettbewerb – solide Qualitätsstandards und wissenschaftsethische Empfehlungen zum Umgang mit Rankings herausgebildet haben.

Literatur

- Althouse, Benjamin M./West, Jevin D./Bergstrom, Carl T./Bergstrom, Ted C. (2009): „Differences in Impact Factor Across Fields and Over Time“, *Journal of the American Society for Information Science and Technology* 60, S. 27-34.
- Benos, Dale J. (2007): „The Ups and Downs of Peer Review“, *Advances in Physiology Education* 31, S. 145.
- Blomeyer, Christian (2007): „Professorenbesoldung“, Hamburg.
- Crane, Diana (1969): „Fashion in Science: Does It Exist?“, *Social Problems* 16, S. 433-441.

- Dilger, Alexander (2010): „Rankings von Zeitschriften und Personen in der BWL“, Zeitschrift für Management 5, S. 91-102.
- Dilger, Alexander/Müller, Harry (2012): „Ein Forschungsleistungsranking auf der Grundlage von Google Scholar“, Zeitschrift für Betriebswirtschaft 82, S. 1089-1105.
- Erne, Roland (2007): „On the Use and Abuse of Bibliometric Performance Indicators: A Critique of Hix's 'Global Ranking of Political Science Departments'“, European Political Science 6, S. 306-314.
- Everling, Oliver (2010): „Credit Rating“, Gabler Wirtschaftslexikon, 17. Aufl., Wiesbaden.
- Frey, Bruno S. (2003): „Publishing as Prostitution? Choosing Between One's Own Ideas and Academic Success“, Public Choice 116, S. 205-223.
- Garfield, Eugene (1964): „Science Citation Index - A New Dimension in Indexing“, Science 144, S. 649-654.
- Grossekettler, Heinz (2007): „Öffentliche Finanzen“, in Bender, Dieter u. a. (Hrsg.): „Vahlens Kompendium der Wirtschaftstheorie und Wirtschaftspolitik“, München, S. 561-721.
- Hardy, Michael (2010): „Pareto's Law“, Mathematical Intelligencer 32, S. 38-43.
- Kieser, Alfred (2012): „JOURQUAL - der Gebrauch, nicht der Missbrauch, ist das Problem“, Die Betriebswirtschaft 72, S. 93-110.
- Kieser, Alfred (2010): „Unternehmen Wissenschaft?“, Leviathan 38, S. 347-367.
- Lange, Stefan (2008): „New Public Management und die Governance der Universitäten“, der moderne Staat – Zeitschrift für Public Policy, Recht und Management 1, S. 237-250.
- Lawrence, Peter A. (2007): „The Mismeasurement of Science“, Current Biology 17, S. R583-R585.
- Lienert, Gustav A./Raatz, Ulrich (1998): „Testaufbau und Testanalyse“, Weinheim.
- Milgrom, Paul/Roberts, John (1992): „Economics, Organization and Management“, Englewood Cliffs, NJ.
- Müller-Böling, Detlef (2000): „Die entfesselte Hochschule“, Gütersloh.
- Müller, Anja/Storbeck, Olaf (2009): „Ranking: Die BWL erfindet sich neu“, Handelsblatt vom 20.5.2009.
- Müller, Harry (2012a): „Zitationen als Grundlage von Forschungsleistungsrankings - Konzeptionelle Überlegungen am Beispiel der Betriebswirtschaftslehre“, Beiträge zur Hochschulforschung 34, S. 68-92.
- Müller, Harry (2012b): „Towards Objectivity in Rankings: Using Citation Data to Assess the Research Performance of Higher Education Institutions“, Journal of the World Universities Forum 5, S. 49-58.
- Nkomo, Stella M. (2009): „The Seductive Power of Academic Journal Rankings: Challenges of Searching for the Otherwise“, Academy of Management Learning & Education 8, S. 106-112.
- Pareto, Vilfredo (1896): „Cours d'economie politique“, Lausanne.
- Podsakoff, Philip M./MacKenzie, Scott B./Podsakoff, Nathan P./Bachrach, Daniel G. (2008): „Scholarly Influence in the Field of Management: A Bibliometric Analysis of the Determinants of University and Author Impact in the Management Literature in the Past Quarter Century“, Journal of Management 34, S. 641-720.

- Power, Michael (1997): „The Audit Society: Rituals of Verification“, New York.
- Todd, Peter A./Ladle, Richard J. (2008): „Hidden Dangers of a 'Citation Culture'“ Ethics in Science and Environmental Politics 8, S. 13-16.
- Wicks, David (2004): „The Institution of Tenure: Freedom or Discipline?“, Management Decision 42, S.619-627
- Wissenschaftsrat (2004): „Empfehlungen zu Rankings im Wissenschaftssystem“, Hamburg.
- Wolf, Joachim/Rohn, Anne/Macharzina, Klaus (2006): „Forschungsleistung in der deutschsprachigen Betriebswirtschaftslehre: Konzeption und Befunde einer empirischen Untersuchung“, Wiesbaden.

Bisher erschienen:

**Diskussionspapiere des
Centrums für Interdisziplinäre Wirtschaftsforschung**

- DP-CIW 1/2011:** Die Gemeinschaft der Lehrenden und Lernenden: Festvortrag zur Promotionsfeier der Wirtschaftswissenschaftlichen Fakultät am 24. November 2010 in der Aula des Schlosses
Alexander Dilger
Januar 2011
- DP-CIW 2/2011:** Plädoyer für eine problemorientierte, lerntheoretisch und fachlich fundierte ökonomische Bildung
Gerd-Jan Krol, Dirk Loerwald und Christian Müller
Februar 2011
- DP-CIW 3/2011:** Gefangen im Dilemma? Ein strategischer Ansatz der Wahl- und Revolutionsteilnahme
Marie Möller
April 2011
- DP-CIW 4/2011:** Overconfidence and Team-Performance: An Analysis of NBA-Players' Self-Perception
Hannah Geyer, Hanke Wickhorst
April 2011
- DP-CIW 5/2011:** Kompetenzziele für das allgemein bildende Fach „Wirtschaft/Ökonomie“ in der Sekundarstufe I
AGOEB – Arbeitsgruppe Ökonomische Bildung
Mai 2011
- DP-CIW 6/2011:** Coping with Unpleasant Surprises in a Complex World: Is Rational Choice Possible in a World with Positive Information Costs?
Roger D. Congleton
Juni 2011
- DP-CIW 7/2011:** Warum der Baseler Ausschuss für Bankenaufsicht mit seinem antizyklischen Kapitalpuffer falsch liegt
Björn Ludwig
Juli 2011
- DP-CIW 8/2011:** Bestimmungsgründe für die Beschäftigung und Rekrutierung von Älteren sowie für das Angebot an altersspezifischen Personalmaßnahmen
Christian Lehmann
August 2011

- DP-CIW 9/2011:** Das „Bruttonationalglück“ als Leitlinie der Politik in Bhutan
– eine ordnungspolitische Analyse
Tobias Pfaff
September 2011
- DP-CIW 10/2011:** Economic Voting and Economic Revolutionizing?
The Economics of Incumbency Changes in European
Democracies and Revolutionary Events in the Arab World
Marie Möller
Oktober 2011
- DP-CIW 11/2011:** Geschlechtsspezifische Verdienstunterschiede und
Diskriminierung am Arbeitsmarkt – Eine Untersuchung unter
Berücksichtigung von Voll- und Teilzeitarbeit
Nele Franz
November 2011
- DP-CIW 1/2012:** Toward a More General Approach to Political Stability in
Comparative Political Systems
Thomas Apolte
Januar 2012
- DP-CIW 2/2012:** An Empirical Study of the Limits and Perspectives of
Institutional Transfers
Marie Möller
Februar 2012
- DP-CIW 3/2012:** Wie (un-) fair sind Ökonomen? Neue empirische Evidenz zur
Marktbewertung und Rationalität
René Ruske, Johannes Suttner
September 2012
- DP-CIW 1/2013:** Zur Ethik von Rankings im Hochschulwesen
Eine Betrachtung aus ökonomischer Perspektive
Harry Müller
Februar 2013



Herausgeber:
Westfälische Wilhelms-Universität Münster
CIW – Centrum für Interdisziplinäre Wirtschaftsforschung
Scharnhorststr. 100
D-48151 Münster

Tel: +49-251/83-25329

Fax: +49-251/83-28429

www.wiwi.uni-muenster.de/ciw

