



Ausarbeitung

Bewertungsprozesse in der Wissenschaft

Bewertungsprozesse in der Wissenschaft

Aktenzeichen:

WD 8 - 3000 - 051/18

Abschluss der Arbeit:

10. Oktober 2018

Fachbereich:

WD 8: Umwelt, Naturschutz, Reaktorsicherheit, Bildung und
Forschung

Die Wissenschaftlichen Dienste des Deutschen Bundestages unterstützen die Mitglieder des Deutschen Bundestages bei ihrer mandatsbezogenen Tätigkeit. Ihre Arbeiten geben nicht die Auffassung des Deutschen Bundestages, eines seiner Organe oder der Bundestagsverwaltung wieder. Vielmehr liegen sie in der fachlichen Verantwortung der Verfasserinnen und Verfasser sowie der Fachbereichsleitung. Arbeiten der Wissenschaftlichen Dienste geben nur den zum Zeitpunkt der Erstellung des Textes aktuellen Stand wieder und stellen eine individuelle Auftragsarbeit für einen Abgeordneten des Bundestages dar. Die Arbeiten können der Geheimschutzordnung des Bundestages unterliegende, geschützte oder andere nicht zur Veröffentlichung geeignete Informationen enthalten. Eine beabsichtigte Weitergabe oder Veröffentlichung ist vorab dem jeweiligen Fachbereich anzuzeigen und nur mit Angabe der Quelle zulässig. Der Fachbereich berät über die dabei zu berücksichtigenden Fragen.

Inhaltsverzeichnis

1.	Zur Entstehung von Evaluation – Benchmarking – Ranking – Rating	4
2.	Bedarf und Notwendigkeit von Bewertungen im Wissenschaftssektor	6
3.	Bewertungstypen	10
3.1.	Beispiele institutioneller Bewertungen	11
3.1.1.	Humboldt Ranking der Alexander von Humboldt-Stiftung [AvH]	12
3.1.2.	CHE University-Ranking	13
3.1.3.	Destatis-Ranking	15
3.1.4.	DFG-Förderatlas	15
3.1.5.	Academic Ranking of World Universities (Shanghai Ranking)	15
3.1.6.	QS World University Ranking	16
3.1.7.	Times Higher Education World University Rankings (THE)	17
3.1.8.	U-Multirank	18
3.1.9.	Best Global Universities	20
3.1.10.	CWTS Leiden Ranking	21
3.2.	Beispiele für Publikationsleistungs-Indikatoren	22
3.2.1.	Field weighted citation impact (FWCI)	22
3.2.2.	Citation Count and Citation per Output	23
3.2.3.	H-Index	23
3.2.4.	Impact-Faktor	23
3.2.5.	Output in Top Percentiles	24
3.2.6.	i-10 Index	24
3.2.7.	g-index	24
3.2.8.	Age-weighted citation rate (AWCR, AWCRpA), AW-Index	24
3.3.	Zur Korrelation einiger Metriken	24
4.	Zu Bewertungen in der Gesundheitsforschung/Medizin	26
4.1.	Forschung und Innovation in der Hochschulmedizin: EFI	26
4.2.	Arbeitsgemeinschaft wissenschaftlicher Medizinischer Gesellschaften: Evaluation der medizinischen Forschungsleistung	28
4.3.	Bibliometric Study in the disciplines Public Health and Epidemiology	29
4.4.	Bewertungen laut UK Research Base 2016	32

1. Zur Entstehung von Evaluation – Benchmarking – Ranking – Rating

Der Begriff „Evaluation“ wird seit einigen Jahren oftmals wenig oder gar nicht begrifflich abgegrenzt und nicht nur in der Wissenschaft, sondern auch im politischen und öffentlichen Leben benutzt. Der Bedarf an „Evaluation“ erscheint in allen Ressorts enorm. Zumeist erhofft man sich davon eine weitgehend objektive, transparente Bewertungsgrundlage, auf Basis derer eine (weitere) Finanzierung erfolgen kann. Es gibt immer wieder Ansätze, eine klare Definition für Evaluation einzuführen, so dass eine Abgrenzung von schlichten „Erfahrungsberichten“ erfolgen kann. Die Gesellschaft für Evaluation e.V. definierte 2016 in ihren „Standards für Evaluation“ dies sehr allgemein wie folgt:

„Evaluation ist die systematische Untersuchung der Güte oder des Nutzens eines Evaluationsgegenstands. Evaluation definiert sich demnach

- *durch ein nachvollziehbares systematisches Vorgehen auf Basis von empirisch gewonnenen Daten, womit sie im Gegensatz zu alltäglichen Bewertungsvorgängen steht,*
- *durch eine transparente, kriteriengeleitete Bewertung, die vor dem Hintergrund eines bestimmten Verwendungskontexts (Untersuchung des Nutzens) oder übergreifend vorgenommen wird (Untersuchung der Güte), womit sie im Gegensatz zu reinen Forschungsstudien steht, sowie*
- *durch ihre Anwendbarkeit auf unterschiedliche Gegenstände. Zu den wichtigsten Evaluationsgegenständen gehören Projekte, Maßnahmen und andere Interventionen (Programme), Organisationen, Produkte sowie Evaluationen selbst (Meta-Evaluation). Als Evaluation wird sowohl der Prozess der Evaluation selbst als auch sein Produkt im engeren Sinne, z.B. als Kurzform für „Evaluationsbericht“, bezeichnet, wobei sich die Standards explizit auf den gesamten Prozess der Evaluation beziehen, nicht nur auf seine Produkte.“¹*

In der vorliegenden Arbeit wird lediglich auf Evaluationsprozesse innerhalb der Wissenschaft eingegangen. „In der Wissenschaft sind organisierte Evaluationsprozesse im 17. Jahrhundert in der britischen Royal Society entstanden, die das Peer Review zur Prüfung wissenschaftlicher Erkenntnisansprüche und zugleich als Legitimationsbasis für die Befreiung von staatlicher Zensur etablierte und sich dabei stark an ständischen Ehrenkodizes orientierte.“² Tatsächlich ist allerdings die flächendeckende Evaluation erst im 20. Jahrhundert entstanden. Nach dem Zweiten Weltkrieg etablierten sich in den 1950er Jahren Programmevaluationen in den USA, und in den 1970er Jahren gab es die Evaluationsforschung als eigenständiges Feld der Sozialwissenschaften, ebenfalls in den USA. Dieser Prozess fand in Europa deutlich zeitverzögert statt. Grundsätzlich

¹ DeGEval, Gesellschaft für Evaluation: Standards für Evaluation, Erste Revision 2016, ISBN 978-3-941569-06-5; im Internet abrufbar unter: https://www.degeval.org/fileadmin/Publikationen/DeGEval-Standards_fuer_Evaluation.pdf [zuletzt abgerufen am 28. August 2018].

² Stefan Hornbostel: (Forschungs-Evaluation); In Simon, D., Knie, A., Hornbostel, S., & Zimmermann, K. (Hrsg.), Handbuch Wissenschaftspolitik (S. 1-18). Wiesbaden: Springer. ISBN: 978-3-658-05677-3. Seiten 293-309.

kann eine Evaluation als Beurteilung **entweder auf Basis von „Peers“³ oder auf Basis von Indikatoren** erfolgen. Natürlich können auch beide Formen parallel angewandt werden. In der Wissenschaft erstrecken sich Evaluationsprozesse u.a. auf die Bewertung und Bewilligung von Programmen und Drittmitteln, auf Publikationsevaluation (insbesondere Peer Review) und die Evaluation ganzer Hochschulgänge. In Deutschland ergibt sich die Notwendigkeit einer Evaluation im Hochschulsektor auch bereits durch die Festschreibung im Hochschulrahmengesetz: „Die Arbeit der Hochschulen in Forschung und Lehre, bei der Förderung des wissenschaftlichen Nachwuchses sowie der Erfüllung des Gleichstellungsauftrags soll regelmäßig bewertet werden. Die Studierenden sind bei der Bewertung der Qualität der Lehre zu beteiligen. Die Ergebnisse der Bewertungen sollen veröffentlicht werden.“⁴

Neben der Evaluation sind in der Wissenschaftsbewertung Begriffe wie **Benchmarking, Rating und Ranking** von besonderer Bedeutung. Ebenso wie die Evaluation basieren sie auf einer ganz ähnlichen Informations- und Datengrundlage. Allerdings sind „Evaluationen [...] jedoch in der Regel nicht explizit vergleichend angelegt; andererseits bleiben sie in der Regel nicht dabei stehen, die Qualität oder Leistung des bewerteten Objekts festzustellen, sondern suchen nach Erklärungen, die dem betroffenen Akteur helfen, seine Leistung zu verbessern. [...] Rankings und Ratings sind summativ, Benchmarkings und Evaluationen haben in der Regel eine (trans-)formativen Komponente.“⁵ Zur begrifflichen Abgrenzung zwischen Evaluation und Ranking bemerkt der Wissenschaftsrat 2004⁶: „Rankings unterscheiden sich von Evaluationen durch die Fokussierung auf die Messung und Bewertung des Outputs – sie beinhalten keine Handlungsempfehlungen und sind nicht prozessorientiert – sowie die Vergleichsabsicht, es werden also mehrere Institutionen oder Förderprogramme nach den gleichen Maßstäben analysiert und bewertet. Unter den verschiedenen Verfahren des Leistungsvergleichs zeichnet Rankings aus, dass die Objekte einer bestimmten Menge (z. B. „Universitäten in Deutschland“) möglichst vollständig erfasst, dass die Leistungskriterien durch ein Kennzahlensystem operationalisiert, und dass die Ergebnisse der Leistungsmessung durch die Bildung von Ranglisten verdichtet werden. Die vollständige Erfassung unterscheidet Rankings von Benchmarkings, bei denen entscheidungsstützende Informationen durch den Vergleich mit gezielt ausgewählten, in der Regel besonders leistungsfähigen Referenzeinrichtungen gewonnen werden sollen; Operationalisierung durch Kennzahlen und Verdichtung in Ranglisten wiederum unterscheidet Rankings von Ratings, also der Bewertung von

3 Unter Peers versteht man Personen, die eine reflektierte Kritik zu dem zu bewertenden Objekt abgeben.

4 Hochschulrahmengesetz (HRG) in der Fassung der Bekanntmachung vom 19. Januar 1999 (BGBl. I S. 18) 1. Kapitel, § 6.

5 Rainer Lange: Benchmarking, Rankings und Ratings, in: Simon, D., Knie, A., Hornbostel, S., & Zimmermann, K. (Hrsg.), Handbuch Wissenschaftspolitik (S. 1-18). Wiesbaden: Springer. ISBN: 978-3-658-05677-3. Seiten 322-333.

6 Wissenschaftsrat: Empfehlungen zu Rankings im Wissenschaftssystem; Drs. 6285-04 Hamburg, 12. November 2004; <https://www.wissenschaftsrat.de/download/archiv/6285-04.pdf> [zuletzt abgerufen am 3. September 2018].

Institutionen nach einer vorgegebenen Notenskala, die in der Regel durch Expertengruppen vorgenommen wird. Auch Ratings können allerdings zur Bildung von Ranggruppen führen oder in Rankings eingehen.“⁷

Benchmarking, Rating und Ranking arbeiten mit Systemen, bei den vergleichend bewertende Kriterien zur Erfassung der Qualität und Leistung abgeleitet werden. Bei Rankings resultiert dies im Erstellen von Ranglisten, bei Ratings erfolgt eine Notenvergabe und „in Benchmarking-Prozessen schließlich finden sich Unternehmen oder andere organisationale Akteure mit ausgewählten Partnern zusammen, um durch die wechselseitige Offenlegung von Informationen über interne Strukturen und Abläufe Stärken und Schwächen zu identifizieren und voneinander zu lernen. In der Regel sind Benchmarking-Prozesse deshalb anders als Rankings und Ratings nicht-öffentliche.“⁸ Da letztere für die Bewertung der Wissenschaft weniger relevant sind, wird verstärkt auf die besonders häufig verwendeten „Rankings“ eingegangen. Im Allgemeinen wird das seit 1983 jährlich erscheinende Ranking von Universitäten und Colleges des U.S. News & World Report als ältestes Ranking im Hochschulbereich genannt. „In Deutschland wurden erste, nicht breit publizierte Leistungsvergleiche wissenschaftlicher Einrichtungen seit 1975 im Auftrag staatlicher Akteure erstellt. Dabei ging es zunächst nur um die Forschungsleistungen. Erst gut zehn Jahre später begannen die Medien auch in Deutschland, sich für das Thema zu interessieren. Die ersten weit verbreiteten Hochschulrankings erschienen 1987 in den Zeitschriften Capital und Wiener, es folgten ab Anfang der 1990er Jahre unter anderem Rankings in Spiegel, Manager-Magazin, Forbes, Wirtschaftswoche, Stern und Focus. 1996 veröffentlichte die Deutsche Forschungsgemeinschaft (DFG) Daten über die zehn Hochschulen, die von 1991 bis 1995 den höchsten Einwerbeerfolg bei der DFG zu verzeichnen hatten. Daraufhin wurde an sie der Wunsch herangetragen, umfassender über die Bewilligungen an die einzelnen Hochschulen zu berichten. Sie reagierte darauf im Jahr 1997 mit einer Publikation, mit der die DFG zwar nicht beanspruchte, ein Ranking vorzulegen, in der die Hochschulen aber tabellarisch nach dem Einwerbeerfolg (absolut oder je Professur) in absteigender Reihenfolge aufgeführt werden.“⁹

Im weiteren Verlauf der vorliegenden Arbeit wird unter dem Begriff „Bewertung“ auf Rankings und Evaluationen innerhalb der Wissenschaft mit besonderem Fokus auf Gesundheitswissenschaften eingegangen.

2. Bedarf und Notwendigkeit von Bewertungen im Wissenschaftssektor

Gerade der Wissenschaftssektor ist ein Bereich, in dem die Leistungsbewertung von Personen und Institutionen eine ganz herausragende Rolle spielt. Daher ist es nicht verwunderlich, dass gerade hier Evaluations- und Rankingsysteme (kontrovers) diskutiert werden. Alle existierenden Methoden sind Gegenstand kontroverser Diskussionen. Laut Einschätzung des Wissenschaftsforschers Stefan Hornbostel liegt die Ursache dafür insbesondere darin, „dass die Bewertung der im Wissenschaftssystem erbrachten Leistungen zu einem großen Teil nur innerhalb des Wissenschaftssystems erfolgen kann. [...] All diese Verfahren sind selbst nicht unumstritten, denn Ziel-

7 Ebd., Seite iv.

8 Ebd., Seite 322.

9 Ebd., Seite 323.

setzungen wie **Innovation, Originalität, Fortschritt, Forschungsqualität, internationale Sichtbarkeit, Erkenntnisgewinn und Anwendungspotential** etc. sind nur schwer zugängliche Konstrukte, die auch innerhalb der scientific community nicht unbedingt konsensual beurteilt werden. Bedingt durch diese Selbstreferenzialität und Bezugnahme auf auch innerwissenschaftlich kontroverse Positionen geraten auch die traditionellen wissenschaftsinternen Bewertungsverfahren immer wieder in die Kritik. Das Besondere am Wissenschaftssystem ist also, dass Außenkriterien, wie man sie beispielsweise für die Bewertung eines Produktes leicht finden kann, kaum existieren. Zur Bewertung wissenschaftlicher Leistungen muss daher immer auf in der Wissenschaft selbst vorhandene oder zu Evaluationszwecken generierte Bewertungen zurückgegriffen werden; entweder in Form von Beurteilungen durch Peers oder in Form von Indikatoren, die Peer Urteile oder relevantes Verhalten der scientific community abbilden. Trotz der geradezu endemischen Kritik an den wissenschaftsinternen Verfahren ist der Bedarf an Evaluation im Wissenschaftssystem massiv gestiegen.¹⁰ Da Bewertungsprozesse ganz unmittelbare Folgen für die Institutionen oder Personen haben, ist es nicht verwunderlich, dass sie heftig kritisiert werden. „Die Auseinandersetzung um Verfahren, Kriterien, Standards und Methoden sind daher nicht nur ein fachlicher Streit um Angemessenheit, sondern auch ein Streit um Definitionsmacht im jeweiligen Feld.“¹¹ Dies wird nochmals besonders verschärft durch Internationalisierung der Wissenschaft, so dass der Referenzrahmen sehr groß, nämlich weltweit ist.

Für die **Notwendigkeit von Bewertungssystemen** werden verschiedene Gründe angegeben. So ist im vergangenen Jahrhundert weltweit ein deutlicher Zuwachs bzw. eine Ausweitung des Wissenschaftssystems zu beobachten. Daraus folgt eine erhöhte Konkurrenzbreite um begrenzte Mittel. Bewertungen werden gerne als Grundlage für Berufungen, Mittelvergabe, internationalen Vergleich und politische Bewertung herangezogen. Dabei ist zu beachten, dass in der Folge aber auch das Evaluationssystem bzw. das Rankingsystem ebenso wie die eine Bewertung veranlassende Stelle in immer stärkerem Ausmaß selbst einer Bewertung unterliegt. Schließlich ist auch festzustellen, dass durch eine geeignete Beurteilung letztendlich auch eine Ausrichtung der Wissenschaft an gesellschaftlichen Vorgaben und gesellschaftlichem/politischem Bedarf möglich wird.

Neben der Wahl der Methodik (Bewertungstypen), auf die im nachfolgenden Kapitel eingegangen wird, stellt sich die Frage der **Organisation der Bewertung**. Diese kann extern¹², intern¹³ oder in einer Mischform durchgeführt werden. Problemfelder ergeben sich in der Regel durch das Ausmaß der Einflussnahme auf die Auswahl der Gutachter, durch das Ausmaß der berücksichtigten Selbstdarstellungen und durch die Festlegung der Indikatoren. Stefan Hornbostel konstatiert zur Problematik der Gutachterwahl: „Als problematisch erweist sich die Gutachterrolle im Spannungsfeld von intern und extern bei institutionellen Evaluationen. Bei einer ausschließlich externen Gutachterberufung besteht die Gefahr, dass nicht hinreichend kontextuelles Wissen für eine

¹⁰ Stefan Hornbostel: (Forschungs-Evaluation); In Simon, D., Knie, A., Hornbostel, S., & Zimmermann, K. (Hrsg.), Handbuch Wissenschaftspolitik (S. 1-18). Wiesbaden: Springer. ISBN: 978-3-658-05677-3. Seite 297.

¹¹ Ebd.

¹² Außerhalb der Organisation veranlasste und durchgeführte Bewertung.

¹³ Innerhalb der zu bewertenden Einrichtung veranlasste und durchgeführte Bewertung.

sachgerechte Beurteilung zur Verfügung steht, bei einer rein intern gesteuerten Berufung ergibt sich in der Regel hohe Konzilianz der Gutachter, was sich in deutlich positiveren Gutachtervoten niederschlägt.“¹⁴ Zumeist wird aus qualitativer Sicht das Peer Review System – ob extern oder intern – besonders geschätzt; die Verwendung von Indikatoren bringt zwar in einem gewissen Maß ein Plus an Transparenz und Objektivität (in Abhängigkeit davon, wer die Auswahl der Indikatoren veranlasst hat), geht aber zu Lasten der Qualität. So bemerkt Georg Schütte in einem Diskussionspapier der Alexander von Humboldt-Stiftung aus dem Jahr 2009: „Die peers, Kolleginnen und Kollegen, sind die einzige kompetente Kontrollinstanz, die Publikationen und Forschungsanträge begutachten kann. Peers entscheiden darüber, welche Wissensbehauptungen Eingang in wissenschaftliche Fachzeitschriften finden.“¹⁵

Schließlich sind trotz des Bedarfs und der Notwendigkeit von Bewertungen sowohl von Institutionen wie auch von individuellen Leistungen die Folgen von Bewertungen zu beachten. Auch dies ist ein Gegenstand kontroverser Debatten. Naheliegende Auswirkungen sind natürlicherweise Budgetkürzungen, Mittelvergabe oder auch Mittelkürzung, Strukturänderungen etc. Allerdings können Einrichtungen auch mit einer zunehmenden Ausrichtung auf das Erfüllen von Bewertungskriterien reagieren¹⁶, so dass die Gefahr besteht, dass Wissenschaft teilweise in Hinblick auf das Erfüllen von Indikatoransprüchen betrieben wird. Momentan existieren parallel eine Vielzahl verschiedener Evaluations- und Rankingsysteme, die für verschiedene Fachbereiche und verschiedene Institutionen von unterschiedlicher Bedeutung sind. Inzwischen gibt es eine Vielzahl an Analysen, die sich mit der Untersuchung der Kriterien der Ausrichtung auf Rankingsysteme und den **beeinflussenden Faktoren** auseinandersetzen.

In einer Publikation aus dem Jahre 2017 wurden Wissenschaftler der Psychologie in einer multivariaten Analyse dahingehend untersucht, welche Faktoren die Publikationsrate (als Indikator für Produktivität) besonders beeinflusste.¹⁷ Als besonders wichtige Faktoren werden dabei das Geschlecht (Männer publizieren häufiger) und die Lehrstuhlgöße (an größeren Lehrstühlen wird mehr publiziert) beobachtet. Auch die Bundeslandzugehörigkeit hatte einen Einfluss.

14 Stefan Hornbostel: Forschungs-Evaluation; In: Simon, D., Knie, A., Hornbostel, S., & Zimmermann, K. (Hrsg.), Handbuch Wissenschaftspolitik (S. 1-18). Wiesbaden: Springer. ISBN: 978-3-658-05677-3. Seite 302.

15 Alexander von Humboldt-Stiftung: Publikationsverhalten in unterschiedlichen wissenschaftlichen Disziplinen Beiträge zur Beurteilung von Forschungsleistungen; Diskussionspapiere der Alexander von Humboldt-Stiftung; 12/2009; im Internet abrufbar unter: https://www.humboldt-foundation.de/pls/web/docs/F13905/12_disk_papier_publikationsverhalten2_kompr.pdf [zuletzt abgerufen am 3. September 2018].

16 Die Wirtschaftswissenschaftler Margit Osterloh und Bruno S. Frey treten in verschiedenen Artikeln gegen die immer größer werdende Bedeutung von Rankings und Impact-Faktoren als Qualitäts-Kriterium in der Wissenschaft ein. In einem Artikel in der Frankfurter Allgemeine Zeitung, 25.06.2012, Nr. 145, S. 12, der den Titel trägt: „Rankings sind wissenschaftsfeindlich“, argumentieren sie, dass Ranking negative Auswirkungen auf die Wissenschaft haben.

17 Justus Rathmann, Sabrina Mayer: Was beeinflusst die Produktivität von Professorinnen und Professoren? Eine Untersuchung individueller und organisatorischer Einflussfaktoren in der Psychologie., Beiträge zur Hochschulforschung, 39. Jahrgang, 2/2017.

Von Oktober 2013 bis September 2017 wurde ein BMBF-Projekt mit dem Titel „Nichtintendierte Effekte der Leistungsbewertung in der Wissenschaft“¹⁸ an der Universität Bamberg gefördert. Dabei konzentrierte man sich auf die Betrachtung der Fächer Chemie, Soziologie und Geschichte. Aus diesem Projekt sind zahlreiche Publikationen hervorgegangen. Unter anderem wurde am Beispiel des britischen Research Assessment Exercise (RAE) und dessen Nachfolger, das Research Excellence Framework (REF), untersucht, welche Auswirkung die Einführung von Leistungsindikatoren auf das Wissenschaftssystem hat.¹⁹ Hierbei stellen die Autoren kritisch fest: „Die vorliegenden empirischen Ergebnisse deuten an, dass es im Zuge des RAE/REF eine Konzentration von Ressourcen an Spitzenuniversitäten in Großbritannien gibt. Dabei ist zunächst der Anstieg der Ungleichheit zwischen den Universitäten bemerkenswert [...]. Wenn die beobachtete Gruppe hinsichtlich ihrer Zuordnung zu den Leistungsklassen homogener wird, so steigt dennoch die Ungleichverteilung von Ressourcen, Publikationen und Anerkennung in Form von Zitationen zwischen ihnen an. [...] Dass es sich dabei vor allem um Chemieinstitute handelt, die einen großen Zuwachs von Professoren verzeichnen, liegt eventuell daran, dass gezielt Wissenschaftler rekrutiert werden, die den Leistungskriterien des RAE Rechnung tragen. [...] Diese Selbstselektion der Professoren sowie der Zuwachs des wissenschaftlichen Personals an symbolisch aufgeladenen Standorten (hohe Leistungsklassen und Rangstufen) entsprechen [bereits publizierten Befunden]. Dabei bleibt fraglich, ob das RAE/ REF die Forschungseffizienz der Universitäten sowohl der unteren Leistungsklassen, als auch des gesamten, akademischen Feldes gesteigert hat.“²⁰

Jährlich findet eine internationale Konferenz zu Wissenschafts- und Technologieindikatoren (STI; International Conference on Science and Technology Indicators) statt.²¹ Neben der Analyse und Etablierung neuer Indikatorsets sind Themen wie open science, open innovation, Kooperationsprojekte, die soziale Akteure einbeziehen, Crowdsourcing in großen digitalen Plattformen, engere Verbindungen zwischen Nutzern und Herstellern wissenschaftlicher und technologischer Geräte, die zunehmende Digitalisierung, das Gleichgewicht zwischen produktiven und "Marktzugang"-Aktivitäten etc. ein Fokus der Konferenzreihe. Zudem sind die Diskussion der unterschiedlichen Auswirkungen von Indikatoranwendungen auf weitreichende, auch nicht wissenschaftliche Sektoren wie der Tourismus, Kultur, das Gesundheitssystem, Altern und die Lebensmittelversorgung wichtige Punkte der STI.

18 Weitere Details siehe: <https://www.uni-bamberg.de/soztheorie/forschung/projekte/nichtintendierte-effekte-der-leistungsbewertung-in-der-wissenschaft/> [zuletzt abgerufen am 3. September 2018].

19 Oliver Wieczorek und Len Ole Schäfer: Verwaltungspraktiken Die Auswirkung von Leistungsindikatoren am Beispiel des britischen Research Assessment Exercise; In: Bauer, Nina, Christina Besio, Maria Norkus and Grit Petschick (Hrsg.): Wissen – Organisation – Forschungspraxis. Der Makro – Meso – Mikro – Link in der Wissenschaft [Knowledge – Organization – Research Practice. The Macro – Meso – Micro – Link in academia]. VS-Verlag: Wiesbaden, 510 – 551.

20 Ebd. Seite 384 ff.

21 Die diesjährige Konferenz (2018) fand vom 12. bis 14. September 2018 in Leiden, Niederlande, statt. Im Internet sind Details unter: <http://sti2018.cwts.nl/> abrufbar. [zuletzt abgerufen am 3. September 2018].

3. Bewertungstypen

Mittlerweile existiert eine kaum überschaubare Fülle an Indikatoren und Bewertungssystemen. Für die Durchführung einer Bewertung muss u.a. zunächst einmal grundlegend entschieden werden: (a) Soll eine Evaluation oder ein vergleichendes Ranking erfolgen? (b) Was ist Gegenstand der Bewertung? (c) Stützt sich die Bewertung auf ein Peer-System oder ein Indikatorssystem und aus welchen Gründen? (d) Wird die Bewertung (gestützt auf Peers) intern oder extern durchgeführt? Hinzu kommen Fragen der konkreten Durchführung (insbesondere die Ableitung der Indikatoren, auf die in den nachfolgenden Unterkapiteln 3.1 und 3.2 detaillierter eingegangen wird).

Zu (a): Ein Ranking – wie bereits erwähnt – orientiert sich im Wesentlichen an Output-Variablen, weniger an Prozessabläufen, und ist vergleichend angelegt. Gerade letzteres erfordert, dass möglichst eine **annähernd vollständige oder mindestens repräsentative Darstellung** erfolgt (eine ausreichende Anzahl von Institutionen muss verglichen werden, die das jeweilige Feld ausreichend verlässlich abbildet, da sonst ein verzerrtes Bild entsteht). Es werden **keine Empfehlungen** abgegeben. Bei Evaluationen hingegen sollte die **Abgrenzung zu Erfahrungsberichten** und deren Auswertung erkennbar sein. Sie basieren auf Daten, sind transparente und kriteriengeleitete Bewertungen mit einem **konkreten Verwendungsweck**.

Zu (b): Jegliche Bewertungen können sich auf sehr unterschiedliche (entweder eines allein oder mehrere) Bereiche beziehen. So findet man individuelle Bewertungssysteme (einzelne Personen), die Bewertung der Lehre (oder Forschung, oder klinischen Leistung...), der Institution als Ganzes, einzelner Programmsysteme (etwa Projektevaluationen). „Querschnittsevaluationen befassen sich mit fachübergreifenden Forschungsschwerpunkten oder disziplinären Forschungsleistungen und -bedingungen. Dabei kommen Institutionen, Förderprogramme, Kooperationen und internationale Vergleiche zum Tragen. Systemevaluationen haben die institutionellen Strukturen des Wissenschaftssystems im Blick und legen neben der Leistungsfähigkeit einzelner Institutionen das Schwergewicht auf die Gesamtwirkungen im Wissenschaftssystem, also das Zusammenwirken verschiedener Akteure. Auch Evaluationen des nationalen Wissenschaftssystems lassen sich hier einordnen. Strukturevaluationen betrachten bestimmte Funktionsbereiche wie etwa das Begutachtungswesen, Qualitätssicherungs- oder Steuerungsverfahren.“²²

Zu (c): Die Vor- und Nachteile von Peer Review versus indikatorgestützter Bewertung sind seit Langem Gegenstand kontroverser Debatten. Insgesamt sprechen sich Wissenschaftler aus qualitativen Gründen in den unterschiedlichsten Bewertungsprozessen immer wieder für die Verwendung von Peer Review aus. Dabei wird insbesondere bei der Bewertung von Publikationen ein Peer Review als qualitativ dem indikatorgestützten System überlegen dargestellt.²³ Demgegenüber steht die Forderung nach Transparenz und stärkerer Objektivität durch Indikatoren.

²² Stefan Hornbostel: Forschungs-Evaluation; In Simon, D., Knie, A., Hornbostel, S., & Zimmermann, K. (Hrsg.), Handbuch Wissenschaftspolitik (S. 1-18). Wiesbaden: Springer. ISBN: 978-3-658-05677-3. Seite 302.

²³ Siehe beispielsweise: Alexander von Humboldt-Stiftung (AvH): Publikationsverhalten in unterschiedlichen wissenschaftlichen Disziplinen, Diskussionspapier der AvH 12/2009; https://www.humboldt-foundation.de/pls/web/docs/F13905/12_disk_papier_publikationsverhalten2_kompr.pdf [zuletzt abgerufen am 3. September 2018].

Zu (d): Unter interner Evaluation wird gemeinhin eine Bewertung, veranlasst durch die zu evaluierende Einheit selbst, angesehen. „Diese [führt sie dann] entweder vollständig [selbst durch] oder [zieht] nach eigenen Auswahlkriterien [bestimmte] externe Expertisen hinzu.“²⁴ „Als externe Evaluation wird hingegen eine von Außen veranlasste und organisierte Evaluation verstanden. Je nach Verfahren haben dabei die Evaluierten mehr oder weniger großen Einfluss auf die Auswahl der Gutachter und über Selbstdarstellungsberichte, Anlass und Gelegenheit, die interne Situation zu reflektieren und darzulegen.“²⁵ Beide Formen haben Vor- und Nachteile. Diese liegen insbesondere darin, dass ein „nicht hinreichend kontextuelles Wissen“²⁶ eine sinnvolle Bewertung erschwert oder unmöglich macht. Demgegenüber stehen natürlicherweise zu einem gewissen Ausmaß Befangenheit und Eigeninteresse bei interner Begutachtung. Oftmals findet man Mischformen.

3.1. Beispiele institutioneller Bewertungen

Eine kritische Haltung zu Rankings (bezogen auf Deutschland) findet sich einleitend in einem Beitrag des Statistischen Bundesamtes „Statistik und Wissenschaft, Bd. 11/2007“. Hier werden die Grenzen dessen, was von einem Ranking zu erwarten ist, zusammengefasst:

„Rankings – auch methodisch anspruchslose – können in einer qualitativ stark differenzierten Hochschullandschaft zwischen oben und unten unterscheiden und sei es nur in Gestalt sich selbst stabilisierender Reputationsurteile. Ungleich schwieriger wird es, wenn die Differenzen zwischen den Bewertungseinheiten weniger groß sind. Die Anforderungen an die methodische Solidität eines Rankings und damit auch die Kosten steigen enorm. Diese Situation haben wir in Deutschland. Ausgehend von einem formal gleichrangigen Institutionenensemble haben sich in den letzten Jahrzehnten immer stärker Stichworte wie Wettbewerb, Exzellenz, Profilbildung und öffentlich verhandelte Leistungsunterschiede in den Vordergrund geschoben. [...] Dieser Informationsbedarf lässt sich mit einfach gestrickten Rankings nicht befriedigen. Der in internationalen Rankings nach wie vor vorherrschende Trend, ganze Hochschulen mit einem Index, der aus unterschiedlich gewichteten, einfach zu erhebenden Einzelindikatoren besteht, zu bewerten, gehört in den Bereich des „Edutainments“. Auch die extreme Informationsverdichtung – wie sie für Rankings typisch ist – wird diesem Informationsbedarf immer weniger gerecht. Zukünftig wird sich wohl eher eine Arbeitsteilung herausbilden: einfache Rankings als Schlagzeilen, die eigentlich nur den Zweck haben, Akademiker auf ein Presseorgan aufmerksam zu machen und auf der anderen Seite methodisch solide Detailanalysen, die auch von der Darstellungsebene sich eher an

24 Stefan Hornbostel: Forschungs-Evaluation; In Simon, D., Knie, A., Hornbostel, S., & Zimmermann, K. (Hrsg.), Handbuch Wissenschaftspolitik (S. 1-18). Wiesbaden: Springer. ISBN: 978-3-658-05677-3. Seite 302.

25 Ebd.

26 Ebd.

Verortung in mehrdimensionalen Räumen und Netzwerkvisualisierungen orientieren. Das Sieger-treppchen aus dem Sport ist für die Abbildung von Lehr- und Forschungsleistungen unter solchen Bedingungen ein Anachronismus.“²⁷

Nachfolgend werden einige auch in Deutschland geläufige Rankings vorgestellt und auf ihre kritische Auseinandersetzung eingegangen.

3.1.1. Humboldt Ranking²⁸ der Alexander von Humboldt-Stiftung [AvH]

„Das Humboldt-Ranking basiert auf einer gewichteten²⁹ Auszählung³⁰ der **Anzahl der [von der AvH geförderten] Gastwissenschaftler pro Hochschule**. Da aber jede Einrichtung ihre spezifischen Stärken und Schwächen in den verschiedenen Fachgebieten hat, differenziert das Humboldt-Ranking auch nach den vier großen Wissenschaftsbereichen: den Geistes- und Sozialwissenschaften, den Lebenswissenschaften, den Naturwissenschaften und den Ingenieurwissenschaften. Die wissenschaftlichen Hochschulen werden in diesen Tabellen nach absoluten Rangplätzen sortiert, der gewichtete Rangplatz kann einer zusätzlichen Spalte entnommen werden. Die außeruniversitären Einrichtungen werden sowohl in der Gesamtstatistik als auch in der Differenzierung nach Wissenschaftsbereichen ausschließlich nach der absoluten Anzahl von Gastwissenschaftlern gerankt.“

27 Quelle: https://www.destatis.de/DE/Publikationen/StatistikWissenschaft/Band11_Hochschulen1030811079004.pdf?blob=publicationFile [zuletzt abgerufen am 3. September 2018].

28 Einleitende Informationen sind auf der Internetseite der Stiftung abrufbar: <https://www.humboldt-foundation.de/web/humboldt-ranking-2017.html> [zuletzt abgerufen am 12. September 2018]. Es ist zu beachten, dass die Methodik in jedem Ranking leicht angepasst wird.

29 Ebd., Angaben zum Humboldt Ranking 2017 (Methoden werden jedes Mal leicht angepasst): Basis für die Gewichtung sind Daten des Statistischen Bundesamts zur Anzahl der W3/C4- Professoren und -Professorinnen an den einzelnen Hochschulen. „Von den universitären Einrichtungen wurden nur die Institutionen bei der Gewichtung berücksichtigt, die mindestens fünf W3/C4-Professoren bzw. -Professorinnen haben. Da aus Datenschutzgründen die Personalangaben des Statistischen Bundesamts bei einigen Universitäten in bestimmten Wissenschaftsbereichen geschwärzt wurden, konnten diese bei der Gewichtung nicht berücksichtigt werden. Die absoluten Rangplätze wurden über alle Institutionen berechnet. Insgesamt ist zu beachten, dass sich der Gewichtungsfaktor auf einen bestimmten Stichtag eines Jahres bezieht, während die Gastaufenthalte über einen Fünfjahreszeitraum gezählt werden.“

30 Ebd. Angaben zum Humboldt Ranking 2017 (Methoden werden jedes Mal leicht angepasst): „Um eine sachgerechte vergleichende Betrachtung zu ermöglichen, werden im Humboldt-Ranking alle bedeutsamen Aufenthalte gezählt, das heißt auch die Aufenthalte eines Humboldtianers an mehreren Gastinstitutionen. Berücksichtigt werden ebenfalls wiederholte Aufenthalte im Rahmen der Alumniförderung. Geförderte werden aber nur einmal pro Gastinstitution gezählt; bei der Auswahl unterschiedlicher Einrichtungen gehen sie jedoch mehrmals in die Statistik ein. Dabei werden alle Aufenthalte ab einer Dauer von einem Monat berücksichtigt. Das Ranking bezieht dabei nur Aufenthalte im Zusammenhang mit Forschungsstipendien und Forschungspreisen in die Statistik ein. Aufenthalte im Rahmen von Programmen, in denen vordergründig die Durchführung eines Projektes steht (Bundeskanzler-Stipendienprogramm, Internationales Klimaschutz-Stipendienprogramm), oder Partnerschaftsprogramme werden nicht berücksichtigt.“

Auf den Internetseiten der AvH-Stiftung sind Humboldt Rankings der Jahre 2009 (Bezugszeitraum 2004 bis 2008), 2012 (Bezugszeitraum 2007 bis 2011), 2014 (Bezugszeitraum 2009 bis 2013) und 2017 (Bezugszeitraum 2012 bis 2016) abrufbar.

Das Humboldt Ranking ist aufgrund seiner Zielgruppe ein **sehr spezifisches Ranking**. Internationale Sichtbarkeit und internationales Renommee lassen sich zu einem gewissen Grad ablesen. Umfassende Bewertungsschlüsse sind allerdings nicht möglich. Aufgrund seiner eingeschränkten Auswahl an Personen an einer Einrichtung (AvH Geförderte und Preisträger) ist es auch schwierlich in andere Indikatoren integrierbar.³¹ Es wurde auch eine Rangfolge nach Fächern, weniger nach Fachgruppen angeregt, um eine bessere Vergleichbarkeit mit anderen Indikatoren herzustellen.

3.1.2. CHE University-Ranking

Das CHE (Centrum für Hochschulentwicklung)-Hochschulranking beurteilt die Studienbedingungen (Fakten zu Studium, Lehre, Ausstattung, Forschung, Urteile von Studierenden über die Studienbedingungen, sowie für einige Fächer die Reputation der Fachbereiche unter den Professoren) in derzeit 39 Fächern. Die Ergebnisse werden in Zusammenarbeit mit Zeit Online veröffentlicht. Das CHE Ranking ist nach eigenen Angaben der „umfassendste und detaillierte“ Hochschulvergleich Deutschlands. Die Ergebnisse des Rankings sind im Internet kostenlos verfügbar.

Das CHE-Ranking wurde verschiedentlich heftig kritisiert³². Der Verband der Historiker und Historikerinnen Deutschlands lehnt seit 2009 die weitere Unterstützung des CHE-Rankings ab.³³ Zahlreiche Historische Institute und Seminare an deutschen Hochschulen haben sich dem Boykott angeschlossen. In einer Stellungnahme vom Juni 2012 hatte sich die Deutschen Gesellschaft für Soziologie (DGS) dahingehend geäußert, das CHE-Ranking weise zum einen gravierende methodische Schwächen und empirische Lücken auf und lade zum anderen durch die summarische Bewertungspraxis und die spezifischen Publikationsformate des Rankings systematisch zu Fehldeutungen ein. In einem offenen Brief des Konvents der Mathematisch-Naturwissenschaftlichen Fakultät der Universität zu Kiel informiert der Dekan 2009 darüber, dass sie aus Gründen gravierender methodischer Mängel sich „zukünftig nicht mehr an Hochschulrankings durch kommerzielle Unternehmen oder wirtschaftsnahe Einrichtungen wie etwa dem ‘Centrum für Hochschulentwicklung (CHE)’“ beteiligen.³⁴

Einzelne Beispiele der Kritikpunkte der vergangenen Jahre waren:

31 Vgl. hierzu: <http://www.che-ranking.de/cms/?getObject=320&GetName=AvH-Ranking&getLang=de> [zuletzt abgerufen am 18. September 2018].

32 Siehe hierzu beispielsweise: <https://www.studis-online.de/HoPo/art-1486-che-ranking-kritik.php> [zuletzt abgerufen am 18. September 2018].

33 Quelle: <https://www.historikerverband.de/verband/geschichtswissenschaft/che-ranking.html> [zuletzt abgerufen am 24. September 2018].

34 Quelle: https://www.wissenschaftsmanagement-online.de/sites/www.wissenschaftsmanagement-online.de/files/migrated_wimoarticle/kommerzielleRankings_07-2009-15.pdf [zuletzt abgerufen am 24. September 2018].

(1) Veröffentlichung des Datensatzes sowie detaillierte wissenschaftliche methodische Angabe:

Es wurde verschiedentlich kritisiert, dass Methoden und insbesondere Daten nicht in ausreichend detailliertem Maße publiziert würden. Das CHE selbst entgegnet dieser Kritik, indem sie auf ihre Methodenseiten verweisen.³⁵ Rohdaten würden allerdings aus Datenschutzgründen nicht veröffentlicht.

(2) Aussagekraft einzelner Indikatoren

Am Beispiel von „Stellen aus dem Hochschulpakt“ weist die DGS auf Mängel bei der Berechnung von Forschungsleistungsparametern hin: „Ersatzweise berechnet das CHE nun zur Abbildung der Forschungsleistung die je (auf Haushaltsstellen) beschäftigten Wissenschaftler eingeworbenen Drittmittel. Dabei werden ohne weitere Differenzierung beispielsweise auch die Stellen des Hochschulpaktes – die ausdrücklich nicht für die Forschung, sondern zur Bewältigung von Lehrlasten geschaffen wurden – zum Divisor der Drittmittelgrößen gemacht. Faktisch bedeutet dies, dass sich mit steigender Lehrbelastung eines Instituts (etwa in der Lehrerbildung) rein rechnerisch dessen Forschungsleistung pro Kopf, die das CHE zu „messen“ beansprucht, verschlechtert. Dass hiervon insbesondere jene Universitäten betroffen sind, die aufgrund ihrer regionalen Lage im Zuge des Hochschulpaktes sehr viele Studierende aufgenommen haben, liegt auf der Hand.“³⁶

(3) Erhebliche Datenlücken³⁷

(4) Prüfungskriterien, Gewichtungen und Abstufungen sowie Parameterfestlegung

- a. Für den Indikator "Forschungsreputation" würden lediglich Dozenten befragt, welchen Standort sie empfehlen, so der Dekan der Fakultät für Sozial- und Verhaltenswissenschaften der Friedrich-Schiller-Universität Jena, Stephan Lessenich.
- b. In die Ermittlung der Qualität in der Lehre gingen nur Befragungsbögen von ca 20% der Studierenden ein. Zudem sei die Selektion der Befragten unklar.³⁸
- c. Die Ergebnisse führten zu einer Abstufung in Spitzengruppe, Mittelgruppe und Schlussgruppe (grüner, gelber, blauer Punkt), was eine zu verknappete Darstellung sei.
- d. Essentiell wichtige Studiensituationsparameter würden nicht in die Analyse (und Bewertung) einbezogen: die jeweiligen Betreuungsrelationen, die damit verbundenen rechnerischen (und realen) Lehrveranstaltungsgrößen und die Leistungsfähigkeit von Prüfungsämtern.³⁹

35 Quelle: <http://www.che-ranking.de/cms/?getObject=64&getLang=de> [zuletzt abgerufen am 18. September 2018].

36 Seite 2 in https://www.soziologie.de/uploads/media/Stellungnahme_DGS_zum_CHE-Ranking_Langfassung.pdf [zuletzt abgerufen am 24. September 2018].

37 Quelle: <http://fudder.de/meine-meinung-diese-ganzen-uni-rankings-sind-total-willkuerlich--140873798.html> [zuletzt abgerufen am 24. September 2018].

38 Seite 3 in https://www.soziologie.de/uploads/media/Stellungnahme_DGS_zum_CHE-Ranking_Langfassung.pdf [zuletzt abgerufen am 24. September 2018].

39 Seite 3 in https://www.soziologie.de/uploads/media/Stellungnahme_DGS_zum_CHE-Ranking_Langfassung.pdf [zuletzt abgerufen am 24. September 2018].

Diesen Kritikpunkten entgegnet das CHE, 20% Befragte seien tatsächlich ein guter Schnitt, ein Fachbeirat aus Wissenschaftlern habe die Indikatoren hergeleitet und diese würden ständig angepasst.

3.1.3. Destatis-Ranking

Das Statistische Bundesamt (destatis) veröffentlicht **Drittmitteleinnahmen**, die deutsche Universitäten (ohne medizinische Einrichtungen) nach den jeweils höchsten Drittmitteleinnahmen pro Professor ordnen. Hierbei sind technologische Universitäten besonders erfolgreich. Dieses Ranking ist wiederum ein sehr spezifisches Ranking, da es lediglich Drittmitteleinwerbungen betrachtet, die in verschiedenen Fächern unterschiedlich starkes Gewicht für die Forschung haben.⁴⁰

3.1.4. DFG-Förderatlas

Der DFG-Förderatlas veröffentlicht seit 1997 im Abstand von drei Jahren Eckdaten zur **öffentlicht geförderten Forschung** in Deutschland, um ein Gesamtbild der Forschungslandschaft, ihrer finanziellen Unterstützung und den Beitrag der in Deutschland sehr unterschiedlichen Forschungsförderer zu vermitteln. Dabei wurde bei der Auswahl der statistischen Größen besonderer Wert darauf gelegt, dass internationale Aspekte der Forschung besonders gewürdigt werden. Daher zielt dieses Ranking auf ein „internationales Publikum“. Der aktuelle Förderatlas liegt seit dem 5. Juli 2018 vor.⁴¹ Am Förderatlas 2015 war beispielsweise Kritik zur Bewertung von Grundlagen und transferorientierter Forschung geübt worden. So kritisierte der Vorsitzende der Hochschulallianz für den Mittelstand und Präsident der Hochschule Niederrhein: „In ihren Strukturen und Auswahlverfahren ist die DFG traditionell auf universitäre Grundlagenforschung geeicht. Das wird dem Charakter der transferorientierten Anwendungsforschung nicht gerecht. Interdisziplinäre anwendungsnahe Forschungsprojekte wie an Hochschulen für angewandte Wissenschaften üblich, fallen durch das DFG-Prüfungs raster.“⁴²

3.1.5. Academic Ranking of World Universities (Shanghai Ranking)

Das Academic Ranking of World Universities (ARWU) - auch bekannt als Shanghai Ranking - ist eine globale Hochschul-Rangliste, die die **Forschungsleistung** in den Vordergrund stellt. Es wird jährlich aktualisiert und von der Shanghai Ranking Consultancy veröffentlicht. Die Leistung internationaler Hochschulen wird anhand von Kriterien wie der Anzahl der Alumni und Mitarbeiter, die Nobelpreise oder andere bedeutende Forschungspreise erhalten haben, der Anzahl der häufig zitierten Forscher, der Anzahl der veröffentlichten und indexierten Artikel sowie der akademischen Leistung pro Kopf im Verhältnis zur Größe der Einrichtung bewertet. Die ARWU führt

40 Quelle: <https://www.research-in-germany.org/en/research-landscape/research-ranking.html> [zuletzt abgerufen am 24. September 2018].

41 Quelle: <http://www.dfg.de/sites/foerderatlas2018/> [zuletzt abgerufen am 24. September 2018].

42 Quelle: <https://www.hs-niederrhein.de/news/news-detailseite/hochschulallianz-kritisiert-neuen-dfg-foerderatlas-geldtopf-fest-in-uni-hand-11587/> [zuletzt abgerufen am 24. September 2018].

über 1.300 Universitäten auf. Die besten 500 werden in einem online verfügbaren Ranking veröffentlicht.

Auch an diesem Ranking wird verschiedentlich Kritik geübt. Unter dem Titel „A Critical Review on World University Ranking in Terms of Top Four Ranking Systems“ wurden 2015 die vier Rankings „the Academic Ranking of World Universities (ARWU)“, „the THE World University Rankings QS“, „World University Rankings“ und „Webometrics Ranking“ miteinander verglichen. Die Autoren stellen fest, dass das ARWU Ranking sich vorrangig auf Naturwissenschaften (im Vergleich zu Sozial- oder Geisteswissenschaften) stütze, so dass dies natürlicherweise Kritik mit sich bringe. Keines der Rankings stelle ausreichend Informationen zu Daten und Datenquellen zur Verfügung, so dass sich die Objektivität der Ranking-Parameter nicht nachvollziehen lasse. Zudem benutze jedes Ranking unterschiedliche Gewichtungswerte, was dazu führe, dass unterschiedliche Institutionen jeweils bevorzugt würden. Spezifisch für das ARWU Ranking seien insgesamt sechs Zählparameter-Kriterien (wie Preise und Medaillen, viel zitierte Forscher, spezifische Publikationsgrößen). Daher sei es natürlich, dass diese Größen stark mit der Größe der Einrichtung korrelierten.⁴³

3.1.6. QS World University Ranking

Die QS (Quacquarelli Symonds) World University Rankings richten sich in erster Linie an Studierende, sind aber aufgrund ihrer Popularität auch eine häufig genutzte Informationsquelle für politische Entscheidungsträger und andere Akteure der Hochschulbildung. Diese Rankings basieren nicht nur stark auf der Reputation einer Institution bei Akademikern und Arbeitgebern, sondern berücksichtigen auch bibliometrische Indikatoren, Studenten-Personalquoten sowie mehrere Indikatoren zur Messung von Internationalisierung und Lehrqualität. Dieser Vergleich bietet eine Rangliste der 400 besten Universitäten und zeigt die darauf folgenden 550 Universitäten in Ranglisten an.⁴⁴

Auch dieses Ranking wurde in dem bereits zitierten Papier⁴⁵ kritisiert. Laut der Autoren ist der umstrittenste Teil der QS World University Rankings die Verwendung einer Meinungsumfrage, die als „Academic Peer Review“ bezeichnet werde. In dieser Umfrage würden aktive Akademiker

43 Farzana Anowar, Mustakim A. Helal, Saida Afroz, Sumaiya Sultana, Farhana Sarker und Khondaker A. Mamun: A Critical Review on World University Ranking in Terms of Top Four Ranking Systems; New Trends in Networking, Computing, E-learning, Systems Sciences, and Engineering, Lecture Notes in Electrical Engineering 312, DOI 10.1007/978-3-319-06764-3_72 (2015); im Internet abrufbar unter: https://www.researchgate.net/publication/278656267_A_Critical_Review_on_World_University_Ranking_in_Terms_of_Top_Four_Ranking_Systems [zuletzt abgerufen am 24. September 2018].

44 Quelle: <https://www.topuniversities.com/qs-world-university-rankings> [zuletzt abgerufen am 24. September 2018].

45 Farzana Anowar, Mustakim A. Helal, Saida Afroz, Sumaiya Sultana, Farhana Sarker und Khondaker A. Mamun: A Critical Review on World University Ranking in Terms of Top Four Ranking Systems; New Trends in Networking, Computing, E-learning, Systems Sciences, and Engineering, Lecture Notes in Electrical Engineering 312, DOI 10.1007/978-3-319-06764-3_72 (2015); im Internet abrufbar unter: https://www.researchgate.net/publication/278656267_A_Critical_Review_on_World_University_Ranking_in_Terms_of_Top_Four_Ranking_Systems [zuletzt abgerufen am 24. September 2018].

auf der ganzen Welt befragt und vermittelten so in der Tat ein **Bild von der universitären Situation als Arbeitsplatz und akademischer Sektor**. Zur gleichen Zeit seien allerdings auch andere Indikatoren wichtig und sollten entsprechend gewichtet werden. Globale Leistungsfaktoren wie internationale Zusammenarbeit zwischen Universitäten oder Wissenschaftlern sollten stärker berücksichtigt werden. Länder des “Commonwealth of Nations” würden 32 % der Rückgabefragebogen ausmachen, während Asien nur durch wenige Länder repräsentiert werde und 22 % ausmache. QS-Rankings stützten sich zur Hälfte der Analyse auf Reputationsindikatoren, somit solle mehr Gewicht auf qualitative, quantitative und praxisrelevante Indikatoren gelegt werden.

3.1.7. Times Higher Education World University Rankings (THE)

Die Times Higher Education World University Rankings (THE)⁴⁶ versuchen, die besten Universitäten der Welt zu identifizieren, indem insgesamt fünf verschiedenen Kategorien abgebildet werden: Peer Review (40 % Gewicht), Zitate (20 %), Lehre (20 %), Internationalität (10 %) und major (mainly international) graduate recruiters (10 %).⁴⁷ Daher spielt die Internationalisierung der Hochschulbildung eine wichtige Rolle bei der Bewertung der Universitäten. In sechs verschiedenen Fachbereichen werden Ranglisten der 1.000 besten Universitäten der Welt erstellt. Tatsächlich ist das THE Ranking eines der bekanntesten Rankings weltweit. Aus den genannten Parameterfeldern wird ein Wert (z-Score) abgeleitet, der benutzt wird, um verschiedene Kriterien zu standardisieren. Diese Prozedur ist für das THE Ranking sehr spezifisch. Nichtsdestotrotz gibt es Kritik an dem Ranking. Es würden Experten aus ausgewählten 30 Universitäten befragt, allerdings sei der Auswahlprozess unklar.⁴⁸ Die Bewertung des THE, dass eine höhere Anzahl internationaler Studenten generell besser sei, sei zumindest zu hinterfragen.⁴⁹ Zudem stelle sich die Frage, wie internationale Studenten gezählt würden (Staatsangehörigkeit, Schulabschluss, Aufenthaltsdauer in einem Land...).⁵⁰ Einige Parameter wie die Bewertung von Forschungspreisen (z. B. Nobelpreis) ist Gegenstand von kritischen Stimmen.⁵¹ Das Gesamtmaß wird als Kombination von Daten der Reputation aus Lehre und Forschung berechnet.

46 Quelle: https://www.timeshighereducation.com/world-university-rankings/2018/world-ranking#!/page/0/length/25/sort_by/rank/sort_order/asc/cols/stats [zuletzt abgerufen am 24. September 2018].

47 Farzana Anowar, Mustakim A. Helal, Saida Afroj, Sumaiya Sultana, Farhana Sarker und Khondaker A. Mamun: A Critical Review on World University Ranking in Terms of Top Four Ranking Systems; New Trends in Networking, Computing, E-learning, Systems Sciences, and Engineering, Lecture Notes in Electrical Engineering 312, DOI 10.1007/978-3-319-06764-3_72 (2015); im Internet abrufbar unter: https://www.researchgate.net/publication/278656267_A_Critical_Review_on_World_University_Ranking_in_Terms_of_Top_Four_Ranking_Systems [zuletzt abgerufen am 24. September 2018].

48 Ebd.

49 Ebd.

50 Ebd.

51 Ebd.

3.1.8. U-Multirank

Auf der Grundlage der Ergebnisse einer Machbarkeitsstudie, die 2010/11 für 150 Universitäten durchgeführt wurde, beschloss die Europäische Kommission, die Umsetzung des Rankingsystems „U-Multirank“⁵² zu finanzieren. Das erste U-Multirank wurde 2014 publiziert und umfasste mehr als 850 Hochschulen aus mehr als 70 Ländern. U-Multirank wird von einem Konsortium unter der Leitung von Professor Frans van Vught vom Center for Higher Education Policy Studies (CHEPS), der Universität Twente in den Niederlanden und Professor Frank Ziegele vom Centre for Higher Education (CHE) in Deutschland sowie dem Centre for Science and Technology Studies (CWTS) der Universität Leiden durchgeführt. U-Multirank wird von der Europäischen Kommission, der Bertelsmann Stiftung und der Banco Santander finanziert. Das Konsortium arbeitet auch mit einer Reihe von nationalen Ranking-Partnern und Interessenvertretern zusammen. Neben internationalen Forschungsuniversitäten werden auch Fachhochschulen und Spezialhochschulen wie Business-Schools bewertet. Untersucht werden fünf „Dimensionen“: Forschungsleistungen, Lehrqualität, Wissenstransfer, internationale Orientierung und regionaler Bezug. Bei der Bewertung stützt sich U-Multirank sowohl auf Onlinebefragungen der Universitäten als auch auf bibliometrische und Patentdaten. Eine Besonderheit des Rankings ist, dass der Benutzer selbst entscheiden kann, auf welche Kriterien er besonderen Wert legt. Die Liste der Indikatoren ist im Internet abrufbar.⁵³ Die Hochschulen werden für jeden Indikator einzeln bewertet (A=sehr gut bis E=schwach). Anders als bei anderen Hochschulrankings wird im U-Multirank keine Gewichtung der einzelnen Indikatoren vorgenommen. Es wird keine Gesamtbewertung der Hochschulen abgeleitet, anhand derer eine Rangliste erstellt wird. Die Bewertung jedes einzelnen Indikators steht für sich, sodass Stärken und Schwächen jeder Hochschule schnell für den Nutzer sichtbar sind. Medizin ist als Fach in U-Multirank enthalten.⁵⁴ An diesem Beispiel lässt sich zeigen, was auch verschiedentlich als Kritik formuliert wurde: „Alle teilnehmenden Unis können sich ein bisschen als Sieger fühlen.“⁵⁵ Schränkt man den Vergleich lediglich auf deutsche Universitäten ein sowie die Abschlüsse Bachelor, Master, Promotion, ergibt sich in der Voreinstellung der U-Multirank-Seite das nachfolgend dargestellte Bild. Für eine Reihe Universitäten liegen keine Daten und somit keine Bewertungen vor.

52 Quelle: <https://www.umultirank.org> [zuletzt abgerufen am 24. September 2018].

53 Quelle: <https://www.umultirank.org/about/methodology/indicators/> [zuletzt abgerufen am 24. September 2018].

54 Quelle: <https://www.umultirank.org/compare?section=comparePrefs&mode=likewithlike&field=36&pref-4=1&pref-4=2&pref-4=3&name=null&trackType=compare> [zuletzt abgerufen am 24. September 2018].

55 Quelle: <https://www.tagesspiegel.de/wissen/neues-uni-ranking-u-multirank-alle-sind-ein-bisschen-sieger/11576942.html> [zuletzt abgerufen am 24. September 2018].



The screenshot shows a table from the multirank website comparing various German universities. The table includes columns for Student-staff ratio, Academic staff with doctorates, Overall learning experience, Research publications (absolute numbers), Citation rate, Income from private sources, Co-publications with industrial partners, International orientation of master programmes, International joint publications, and Regional joint publications. The data is presented as a grid where each cell contains a letter grade (A-E) or a dash (-). A legend at the bottom explains the grades: A (Very good), B (Good), C (Average), D (Below average), E (Weak), Data unavailable, and Not applicable.

		Student-staff ratio	Academic staff with doctorates	Overall learning experience	Research publications (absolute numbers)	Citation rate	Income from private sources	Co-publications with industrial partners	International orientation of master programmes	International joint publications	Regional joint publications
A-Z	Top scores										
Heidelberg University	DE	-	-	-	A	B	-	A	-	B	C
Free University of Berlin	DE	-	-	-	B	B	-	A	-	B	B
LMU Munich	DE	-	-	-	B	B	-	A	-	A	B
University of Münster	DE	-	-	B	B	B	-	A	-	B	D
University of Bonn	DE	B	-	C	B	B	-	A	-	A	C
University of Duisburg-Essen	DE	-	D	C	B	A	-	A	B	B	C
Ulm University	DE	-	-	C	B	B	-	A	A	B	D
Johannes Gutenberg-Universität Mainz	DE	-	-	D	B	B	-	A	B	B	C
University of Göttingen	DE	-	-	B	B	B	-	B	-	B	C
RWTH Aachen University	DE	-	-	B	B	C	-	A	-	B	C
Lübeck University	DE	D	-	B	B	C	-	A	-	B	D
Universität Witten/Herdecke	DE	C	-	B	C	C	-	B	-	C	B
Asklepios Campus Hamburg	DE	-	-	B	-	-	X	-	B	-	-

PDF export Latest update: 2018

● A (Very good) ● B (Good) ● C (Average) ● D (Below average) ● E (Weak) — Data unavailable × Not applicable

Blendet man nun weitere Kriterien der „Research Performance“ ein: external research income, doctorate productivity, absolute numbers of research publications, citation rate, top cited publications, interdisciplinary publications, research orientation teaching, kann sich das Ranking-Bild deutlich ändern. Während bei der Zitationsrate die Universität Duisburg-Essen den Spitzenplatz einnimmt, kommen unter den top cited publications die Universität Bonn, Universität Heidelberg, LMU München und die Johannes Gutenberg-Universität Mainz hinzu. Die RWTH Aachen hingegen punktet in „external research income“.



The screenshot shows a table from the multirank website comparing German universities across various research metrics. The columns include: External research income, Doctorate productivity, Research publications (absolute numbers), Citation rate, Top cited publications, Interdisciplinary publications, and Research orientation of teaching. The rows list universities such as University of Duisburg-Essen, University of Bonn, Heidelberg University, LMU Munich, Johannes Gutenberg-Universität Mainz, University of Münster, Free University of Berlin, University of Göttingen, Ulm University, RWTH Aachen University, Lübeck University, Universität Witten/Herdecke, and Asklepios Campus Hamburg. Each university has a row with its name, country (DE), and scores for each metric.

		Best University for Me	University Rankings	Press / Media	About	<input type="text"/> search
	Show symbols	External research income	Doctorate productivity	Research publications (absolute numbers)	Citation rate	Top cited publications
	Show the whole table	A-Z	Top scores			
		▼ ▲	▼ ▲	▼ ▲	▼ ▲	▼ ▲
University of Duisburg-Essen	DE	A	C	B	A	A
University of Bonn	DE	-	A	B	B	A
Heidelberg University	DE	-	-	A	B	A
LMU Munich	DE	-	-	B	B	A
Johannes Gutenberg-Universität Mainz	DE	A	-	B	B	A
University of Münster	DE	A	C	B	B	B
Free University of Berlin	DE	-	-	B	B	B
University of Göttingen	DE	-	-	B	B	B
Ulm University	DE	A	B	B	B	C
RWTH Aachen University	DE	A	B	B	C	C
Lübeck University	DE	-	A	B	C	C
Universität Witten/Herdecke	DE	A	A	C	C	B
Asklepios Campus Hamburg	DE	-	x	-	-	B

3.1.9. Best Global Universities

Im Jahr 2014 veröffentlichte der U.S. News & World Report sein erstes „Best Global University Ranking“. Die Ranking-Methodik basiert auf zehn verschiedenen Indikatoren, die die akademische Leistung und den Ruf der Universitäten messen. Das erste Best Global Universities Ranking des U.S. News & World Report wurde am 28. Oktober 2014 publiziert. Es basiert auf Daten und Kennzahlen von Thomson Reuters. Die Universitäten werden anhand von Faktoren wie der globalen Forschungsreputation, Publikationen und der Anzahl der hoch zitierten Arbeiten bewertet. U.S. News veröffentlicht auch regionalspezifische und fachspezifische globale Rankings, die auf dieser Methodik basieren. Am 24. Oktober 2017 wurde das aktuelle Ranking „2018 Best Global Universities Rankings“ veröffentlicht.⁵⁶ Die bekanntesten globalen Universitäts-Rankings: the Times Higher Education World University Rankings, Academic Ranking of World Universities und QS World University Rankings seien, so der Daten-Strategie der U.S. News Robert Morse,

56 Quelle: <https://www.usnews.com/info/blogs/press-room/articles/2017-10-24/us-news-announces-2018-best-global-universities-rankings> [zuletzt abgerufen am 24. September 2018].

britisch und chinesisch.⁵⁷ Mit diesem Ranking komme ein US-amerikanisches hinzu. Tatsächlich stellt die Washington Post stellte fest, dass einige US-Institutionen in der globalen Rangliste der U.S. News einen höheren Rang einnehmen als in ihrer nationalen Rangliste, insbesondere Princeton, die im Jahr 2014 zur besten Universität der USA gewählt wurde.⁵⁸

3.1.10. CWTS Leiden Ranking

Das CWTS Leiden (Centrum voor Wetenschap en Technologische Studies an der Universität leiden, Niederlande) Ranking ist ein jährlich erscheinendes globales Hochschulranking, das ausschließlich auf bibliometrischen Indikatoren basiert. Dabei werden die zugrundeliegenden Daten der bibliographischen Datenbank Web of Science von Thomson Reuters benutzt.⁵⁹ Tatsächlich ist das CWTS Leiden Ranking eng verwandt mit dem U-Multirank. Auf den Webseiten des CWTS Leiden Ranking wird hierzu bemerkt: „Universitätsrankings konzentrieren sich auf spezifische Dimensionen der Hochschulleistung, die in der Regel relativ leicht zu quantifizieren sind. Das Leidener Ranking beispielsweise hat einen recht engen Rahmen, der sich auf bestimmte Aspekte der wissenschaftlichen Leistungsfähigkeit von Universitäten konzentriert. Einige andere Hochschulrankings haben einen größeren Umfang, wobei U-Multirank wahrscheinlich das umfassendste Rankingsystem ist.“⁶⁰ Die technischen Unterschiede werden in einem Informationspapier des U-Multirank detaillierter erläutert.⁶¹ Nicht nur die Datengrundlage ist sehr ähnlich, auch die Wahl der bibliometrischen Indikatoren sind in beiden Ranking Systemen eng miteinander verbunden. Ein wichtiger Unterschied besteht darin, dass das Leidener Ranking auf 'Web of Sciences-indexierten Publikationen' in internationalen, peer-reviewten wissenschaftlichen Zeitschriften basiert. Veröffentlichungen anderer Web of Sciences basierter Quellen (nationale wissenschaftliche Zeitschriften, Fachzeitschriften und populäre Zeitschriften) sind nicht enthalten. Dies gilt insbesondere auch für Forschungspublikationen in anderen Sprachen als Englisch. Auch Veröffentlichungen, die nicht gut verlinkt sind (im Sinne von Zitierlinks zu anderen Zeitschriften), werden nicht berücksichtigt. Dies sind vor allem, aber nicht ausschließlich Zeitschriften aus den Bereichen Kunst und Geisteswissenschaften. Im Gegensatz dazu berücksichtigt U-Multirank alle Web of Sciences-indexierten Quellen und Publikationen.⁶² In einem Artikel aus

57 Quelle: <https://www.insidehighered.com/quicktakes/2014/10/10/us-news-issue-new-global-university-rankings> [zuletzt abgerufen am 24. September 2018].

58 Nick Anderson (28. Oktober 2014). "U.S. News rolls out global university rankings, with some surprises". The Washington Post.

59 Quelle: <http://www.leidenranking.com/> [zuletzt abgerufen am 24. September 2018].

60 Quelle: <http://www.leidenranking.com/information/responsibleuse> [zuletzt abgerufen am 24. September 2018]. (Übersetzung durch die Autorin)

61 Quelle: [https://www.umultirank.org/export/sites/default/press-media/documents/Bibliometric Analysis in U-Multirank 2018.pdf](https://www.umultirank.org/export/sites/default/press-media/documents/Bibliometric%20Analysis%20in%20U-Multirank%202018.pdf) [zuletzt abgerufen am 24. September 2018].

62 Quelle: [https://www.umultirank.org/export/sites/default/press-media/documents/Bibliometric Analysis in U-Multirank 2018.pdf](https://www.umultirank.org/export/sites/default/press-media/documents/Bibliometric%20Analysis%20in%20U-Multirank%202018.pdf) [zuletzt abgerufen am 24. September 2018].

dem Jahr 2010⁶³ kritisierte Loet Leydesdorff die vom CWTS Leiden Ranking angewandte Methode. Neben dem Fehlen von Fehlerbalken kritisieren die Autoren besonders die Normalisierungprozedur des CWTS Leiden Rankings. Allerdings wurde im Laufe der Zeit das Ranking mehrfach überarbeitet.

3.2. Beispiele für Publikationsleistungs-Indikatoren

Indikatoren fließen in Rankings ein und stellen letztendlich Rechenvorschriften dar, die aus der dem Ranking zugrundeliegenden Datenbank geeignete Größen ableiten, um eine Bewertung vornehmen zu können. Die Wahl der (bibliometrischen) Indikatoren ist entscheidend dafür, wie geeignet das Ranking ist, um das für die Bewertung gesetzte Ziel zu erreichen. Indikatoren werden immer wieder angepasst und neue vorgestellt.

Im Folgenden werden Indikatoren vorgestellt, die sich mit der Bewertung von Zitierleistungen und Publikationsleistungen als eine wesentliche Komponente der Forschungsleistungsbewertung beschäftigen. Selbstverständlich sind aber auch andere Indikatoren, die Lehrleistung, Transferleistung, klinische Leistung etc. messen, entwickelt worden. Alle Indizes greifen auf Datenbanken zurück, auf deren Einträge basiert die Berechnung des jeweiligen Index erfolgt. Dies ist bei der Bewertung der jeweiligen Resultate, insbesondere bei jeglichen weiteren Vergleichen zu beachten. Bekannte Datenbanken sind Scopus, SciVal, Web of Science, Google Scholar My Citations, USC Research Bank. Auf die Unterschiede dieser Tools wird in dieser Arbeit nicht eingegangen. Allen publikationsbasierten Indikatoren ist gemein, dass sie stark vom jeweiligen Fachbereich abhängen, da das Publikationsverhalten sich zwischen den Fächern erheblich unterscheidet. Weitere Faktoren, auf die Indikatoren unterschiedlich stark reagieren, sind das Alter der zu bewertenden Personen, die Gruppengröße und die Behandlung von Selbztitierungen.

3.2.1. Field weighted citation impact (FWCI)

Der FWCI wird berechnet, indem die Anzahl der Zitate, die eine Veröffentlichung erhält, durch die durchschnittliche Anzahl der Zitate in einer Datenbank geteilt wird, die Veröffentlichungen in einem Zeitraum, vom gleichen Typus und innerhalb derselben Fachkategorie erfasst⁶⁴. Der FWCI ist auf Zitier- und Indexierungsinformationen angewiesen, die in Abstract- und Zitatdatenbanken wie Scopus und World of Science verfügbar sind. Zusätzlich zu den Zitatzählungen klassifizieren diese Quellen Publikationen nach Jahr, Art und Thema. Der FWCI verwendet einen Mittelwert, der von Ausreißern beeinflusst werden kann. Die Verteilung der Zitate über Publikationen hinweg ist oft verzerrt. Die meisten Veröffentlichungen in einer Stichprobe erhalten nur wenige Zitierungen, abgesehen von einer kleinen Zahl von Ausnahmen. Ausreißer haben einen stärkeren Einfluss auf kleinere Proben und führen zu einem instabileren FWCI. Darüber hinaus ist das FWCI auf Zitierdaten aus dem Jahr der Veröffentlichung plus drei Jahre beschränkt, was Inhalte mit längeren Zitierzeiten beeinträchtigen könnte. Der globale Durchschnitt von 1.0 macht den FCWI zu einem einfach zu lesenden intuitiven Bewertungsindex.

63 Es handelt sich um einen Report, der nicht in einer wissenschaftlichen Zeitschrift erschienen ist: <https://arxiv.org/abs/1003.3977> [zuletzt abgerufen am 24. September 2018].

64 Somit ein Maß dafür wieviel „mehr oder weniger“ zitiert wurde, als man im Schnitt erwarten würde.

3.2.2. Citation Count and Citation per Output

Citation Count summiert die bisher erhaltenen Zitate in der Datenquelle. Zitierzahlmetriken sind nützlich, um die Sichtbarkeit zu bewerten, wenn Forscher aus ähnlichen Bereichen oder Disziplinen und mit ähnlichen Laufbahnen kommen.

3.2.3. H-Index

Diese Kenngröße wurde 2005 von dem Physiker Jorge E. Hirsch vorgeschlagen und nach ihm benannt. Der H-Index ist die (größtmögliche) Anzahl der Publikationen eines gegebenen Wissenschaftlers, die mindestens h-mal zitiert wurden.

Beispiel: Angenommen ein Wissenschaftler hat 15 Arbeiten, die 10, 9, 8, 8, 7, 6, 5, 5, 4, 2, 2, 2, 2, 1-mal zitiert wurden. Dann hat der Wissenschaftler einen H-Index von 6, weil 6 von seinen 15 Publikationen mindestens 6-mal und (15-6=9) 9 Publikationen höchsten 6-mal zitiert wurden.

Es ist zu beachten, dass natürlicherweise ein bestehender H-Index-Wert für einen Wissenschaftler immer lediglich steigen kann, selbst wenn dieser nichts mehr produziert.

3.2.4. Impact-Faktor

Der Impact-Faktor ist ein Indikator, der einer wissenschaftlichen Zeitschrift zugeschrieben wird und ihren wissenschaftlichen Einfluss messen soll. Teilweise werden für Wissenschaftler (in Abhängigkeit davon, an welcher Autorenposition sie auf einer Publikation stehen) alle Impact Faktoren ihrer Publikationen zusammengezählt, so dass man von einem „Summen Impact-Faktor“ spricht. Dieser ist natürlicherweise tendenziell höher, je länger sich ein Wissenschaftler bereits in einer wissenschaftlichen Laufbahn befindet. Der Impact-Faktor gibt an, wie häufig im Durchschnitt ein in dieser Zeitschrift veröffentlichter Artikel von anderen wissenschaftlichen Artikeln pro Jahr zitiert wird. (Zahl der Zitate im Bezugsjahr auf alle Publikationen der vorangehenden zwei Jahre, dividiert durch die Zahl der Artikel in den vorangehenden zwei Jahren). Der Impact-Faktor ist hochgradig umstritten, wird aber aufgrund seiner sehr einfachen Anwendung (er wird für jede wissenschaftliche Zeitschrift publiziert) immer wieder gerne herangezogen. Es existieren ein Fülle von Publikationen, die die Art der Berechnung, die Manipulierbarkeit, Behandlung unterschiedlicher Publikationstypen kritisieren. Auf den Internetseiten des Instituts für Forschungsinformation und Qualitätssicherung (iFQ) wird in einem Artikel auf die Problematik der Benutzung von Impact-Faktoren eingegangen und die Schlussfolgerung gezogen, dass dieser „ein gutes Mittel“ darstelle, „um die kurzfristige wissenschaftliche Wirkung von Zeitschriften zu bewerten. Zeitschriften mit ähnlichem Profil innerhalb einer Fachgruppe lassen sich auf der Basis dieses Wertes (evtl. ergänzt durch weitere Werte wie Immediacy Index und Cited Half-Life) vergleichen. Für die Bewertung der Publikationsleistung von Wissenschaftlern sind die Journal Impact Factors allerdings nicht geeignet.“⁶⁵

65 Quelle: http://www.forschungsinfo.de/iq/iq_inhalt.asp?agora/Journal_Impact_Factor/journal_impact_factor_inc.htmlXXXJournal%20Impact%20Factor [zuletzt abgerufen am 24. September 2018].

3.2.5. Output in Top Percentiles

„Outputs in Top Percentiles“ gibt an, inwieweit Publikationen in den am häufigsten zitierten Perzentilen der Datenquelle vorhanden sind. Die Zitierzahlen, die die Schwellenwerte der am häufigsten zitierten Arbeiten von beispielsweise 10 % (top 10 Percent) in einer Datenbank pro Erscheinungsjahr darstellen, werden berechnet.

3.2.6. i-10 Index

Der i-10-Index bezeichnet die Anzahl der Publikationen mit wenigstens 10 Zitaten. Der Faktor wurde von Google Scholar eingeführt und wird in den „Google My Citation“ Profilen wiedergegeben.

3.2.7. g-index

Dieser Index wurde 2006 von Leo Egghe eingeführt.⁶⁶ Der Index wird berechnet auf Basis der Verteilung der Zitate, die von den Publikationen eines gegebenen Wissenschaftlers erhalten wurden. Die Artikel werden - wie beim H-Index - in abnehmender Zitat-Reihenfolge sortiert. Der g-Index ist die größte Anzahl n von hoch zitierten Artikeln, bei denen die *durchschnittliche* Anzahl der Zitate mindestens n beträgt. Der g-Index ist eine Alternative zum älteren H-Index, der die Anzahl der Zitate nicht mittelt.⁶⁷

3.2.8. Age-weighted citation rate (AWCR, AWCRpA), AW-Index

Mit der AWCR⁶⁸ wird die Anzahl der Zitationen einer Gruppe von Arbeiten unter Adjustierung des Alters der einzelnen Arbeiten berechnet. Somit handelt es sich um eine altersgewichtete Zitierrate. Konkret wird die Anzahl der Zitationen einer bestimmten Arbeit durch das Alter dieser Arbeit dividiert. Der AW-Index hingegen ist die Quadratwurzel der Summe aller altersgewichteten Zitatwerte über alle Papiere.⁶⁹

3.3. Zur Korrelation einiger Metriken

Wenn Indikatoren zur Darstellung von Publikationsleistungen herangezogen werden, ist es erforderlich, genau zu beachten, welche Metrik verwandt wurde. Ideal wäre die parallele Verwendung

66 Quelle: <https://link.springer.com/content/pdf/10.1007%2Fs11192-006-0144-7.pdf> [zuletzt abgerufen am 24. September 2018].

67 Das heißt g bezeichnet die Anzahl, so dass die Ungleichung $g^2 = \text{Summe } (i \leq g)$ der c_i, wobei c_i die Zitanzahlen bezeichnet.

68 The AR-index: complementing the h-index, ISSI Newsletter, 2007, 3(1), p. 6.

69 Quelle: <https://sci2s.ugr.es/sites/default/files/files/TematicWebSites/hindex/Jin2007.pdf> [zuletzt abgerufen am 24. September 2018].

und Darstellung einer ganzen Palette von Indikatoren. Tatsächlich unterscheiden sich die Ergebnisse nämlich - wie die nachfolgenden Studien zeigen - erheblich.

Im Juli 2015 wurde eine britische Studie veröffentlicht, in der 15 Indikatoren der Forschungsleistung verglichen wurden.⁷⁰ Insgesamt kommen die Autoren zum Schluss, dass keine Metrik geeignet sei, das Peer-Review gestützte System zu ersetzen. Sorgfältig ausgewählte Indikatoren könnten die Entscheidungsfindung ergänzen, aber eine Mischung aus Expertenurteilen, quantitativen Indikatoren und qualitativen Maßnahmen, die die Forschungsvielfalt berücksichtigten, sei weiterhin erforderlich. Es bestehe die berechtigte Sorge, dass einige Indikatoren missbraucht oder „getuned“ werden können: Zeitschriftenauswirkungsfaktoren, Hochschulrankings und Zitierzahlen seien drei herausragende Beispiele. Die Dateninfrastruktur, die der Verwendung von Metriken und Informationen über die Forschung zugrunde liege, sei noch immer fragmentiert, mit unzureichender Interoperabilität zwischen den Systemen. Verschiedentlich wurde Kritik an der Studie geübt. Beispielsweise wurde bemerkt, dass zwischen Buch-Artikeln und Publikationen in Journals unzureichend unterschieden wird. Zudem wird von „Times Higher Education“ einigen Universitäten vorgeworfen, dass diese das REF-System ganz gezielt ausnutzten. Es würden Mitarbeiter mit etablierten Forschungsergebnissen unmittelbar vor dem REF von ihren Universitäten angeworben, um die volle Anerkennung für ihre Publikationen für sich zu deklarieren, ohne das Risiko eingegangen zu sein, den Forscher vorher zu unterstützen.⁷¹ Zudem sei das System zu teuer, unnötig komplex; es gebe angesichts der Digitalisierung wesentlich effektivere Möglichkeiten.⁷²

Die britische Studie wird auch in einer Publikation aus dem Jahr 2017 weiterverfolgt.⁷³ Hier untersuchen die Autoren Zusammenhänge zwischen den stark zitierten Arbeiten in Psychiatrie, Klinischer Psychologie und Neurologie und deren Metriken. Die Korrelationen fallen meist schwach bis moderat aus. Die stärkste Korrelation zu einem „gold Standard Peer Review“ bestand mit dem SCImago Journal Rank, einer Variante des Journal Impact Factors, während die Höhe der Twitter-Aktivität keine Korrelation zeigte. Die Autoren schlagen vor, eine Kombination von Metriken anzuwenden, unter Hinzuziehen von Gewichtungen hinsichtlich des Autorenbeitrages (d.h. beispielsweise je nach Positionierung des Autors in der Autorenliste wird der Beitrag gewichtet dem jeweiligen Autor zugeschrieben).

70 Die Studie ist im Internet abrufbar unter: http://www.hefce.ac.uk/media/HEFCE_2014/Content/Pubs/Independentresearch/2015/The.Metric.Tide/2015_metrictideS2.pdf [zuletzt abgerufen am 18. September 2018].

71 Quelle: <https://www.timeshighereducation.com/news/twenty-per-cent-contracts-rise-in-run-up-to-ref/2007670.article> [zuletzt abgerufen am 3. September 2018].

72 Quelle: <http://blogs.lse.ac.uk/impactofsocialsciences/2011/06/10/ref-alternative-harzing-google-scholar/> [zuletzt abgerufen am 3. September 2018].

73 Seena Fazel and Achim Wolf: What is the impact of a research publication?; Evid Based Ment Health. 2017 May; 20(2): 33–34. Publiziert am 6. April 2017. doi: 10.1136/eb-2017-102668; Im Internet abrufbar unter: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5502238/> [zuletzt abgerufen am 24. September 2018].

4. Zu Bewertungen in der Gesundheitsforschung/Medizin

4.1. Forschung und Innovation in der Hochschulmedizin: EFI⁷⁴

Im Jahr 2014 war ein Kernthema des Berichts der Expertenkommission Forschung und Innovation (EFI) die Hochschulmedizin im internationalen Vergleich. Dabei geht die Kommission auf eine Vielzahl Aspekte ein; in der vorliegenden Arbeit wird lediglich die Bewertung von Forschungs- und Innovationsleistung im Hochschulbereich dargestellt. Daneben wurden im EFI-Bericht 2014 auch international vergleichend jeweils für die Hochschulmedizin die Forschung und Entwicklungs-Ausgaben (FuE), die organisatorischen Rahmenbedingungen, die länderspezifische Förderpraxis, die Bedeutung klinischer Studien, der Technologietransfer, Industriezusammenarbeit, die Einbindung von Informations- und Kommunikationstechnologie sowie die Arbeitsbedingungen kritisch beleuchtet.

Im Kapitel zur Bewertung der Innovationsleistung in der deutschen Hochschulmedizin stehen besonders zwei Indikatoren im Vordergrund: die Publikationsintensität und die Exzellenzrate. In die Analyse gingen folgende Hochschulmedizinstandorte ein: Berlin, Hannover, Heidelberg, München, Tübingen, Basel, Bern, Genf, Zürich, Amsterdam, Leiden, Rotterdam, Utrecht, Hamilton, Montreal, Toronto, Vancouver, Baltimore, Boston/Cambridge, Houston, San Francisco, Washington.

Die **Publikationsintensität** eines Standortes ist definiert als die Anzahl der Publikation, dividiert durch die Anzahl der Autoren am Standort.

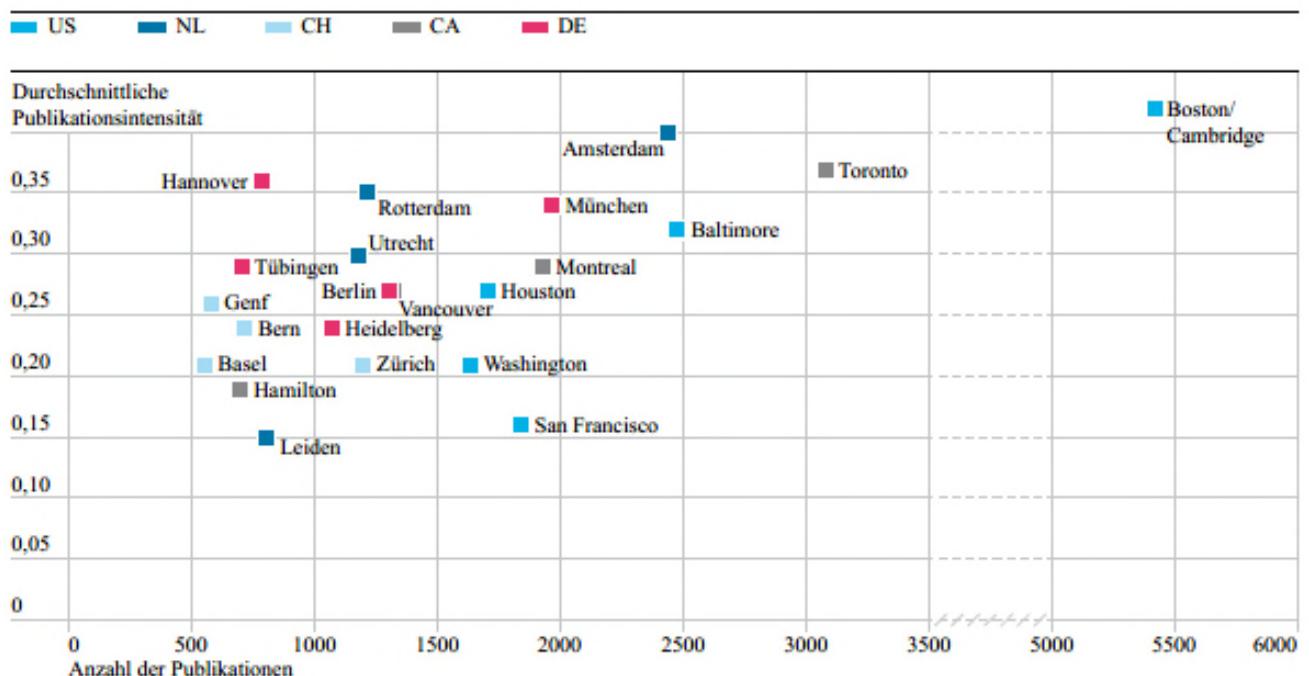
Publikationsintensität = # Publikationen / # Autoren

In der nachfolgenden Grafik wird die Anzahl der Publikationen an einem Standort (x-Achse) gegen die durchschnittliche Publikationsintensität am selben Standort (y-Achse) abgetragen. Deutsche Standorte sind pink-rot markiert.

Während sich sowohl Hannover wie München durch eine vergleichsweise hohe Publikationsintensität auszeichnen – Hannover sogar noch höher als München -, zeigt Hannover dabei eine recht niedrige Anzahl an Publikation. Obwohl am Standort München absolut gesehen mehr publiziert wird als am Standort Hannover, ist die Publikationsintensität des Standorts Hannover im Vergleich der ausgewählten deutschen Standorte am höchsten.

⁷⁴ Expertenkommission Forschung und Innovation: Jahresgutachten zu Forschung, Innovation und technologischer Leistungsfähigkeit Deutschlands 2014; <https://www.e-fi.de/gutachten-und-studien/gutachten/> [zuletzt abgerufen am 3. September 2018].

**Anzahl der Publikationen und Publikationsintensität
der Standorte 2012**



Publikationsintensität: Durchschnittliche Anzahl der Publikationen pro Autor.
Quelle: Web of Science, Scopus, Berechnungen des Fraunhofer ISI.

Quelle: EFI Gutachten 2014, Seite 58.

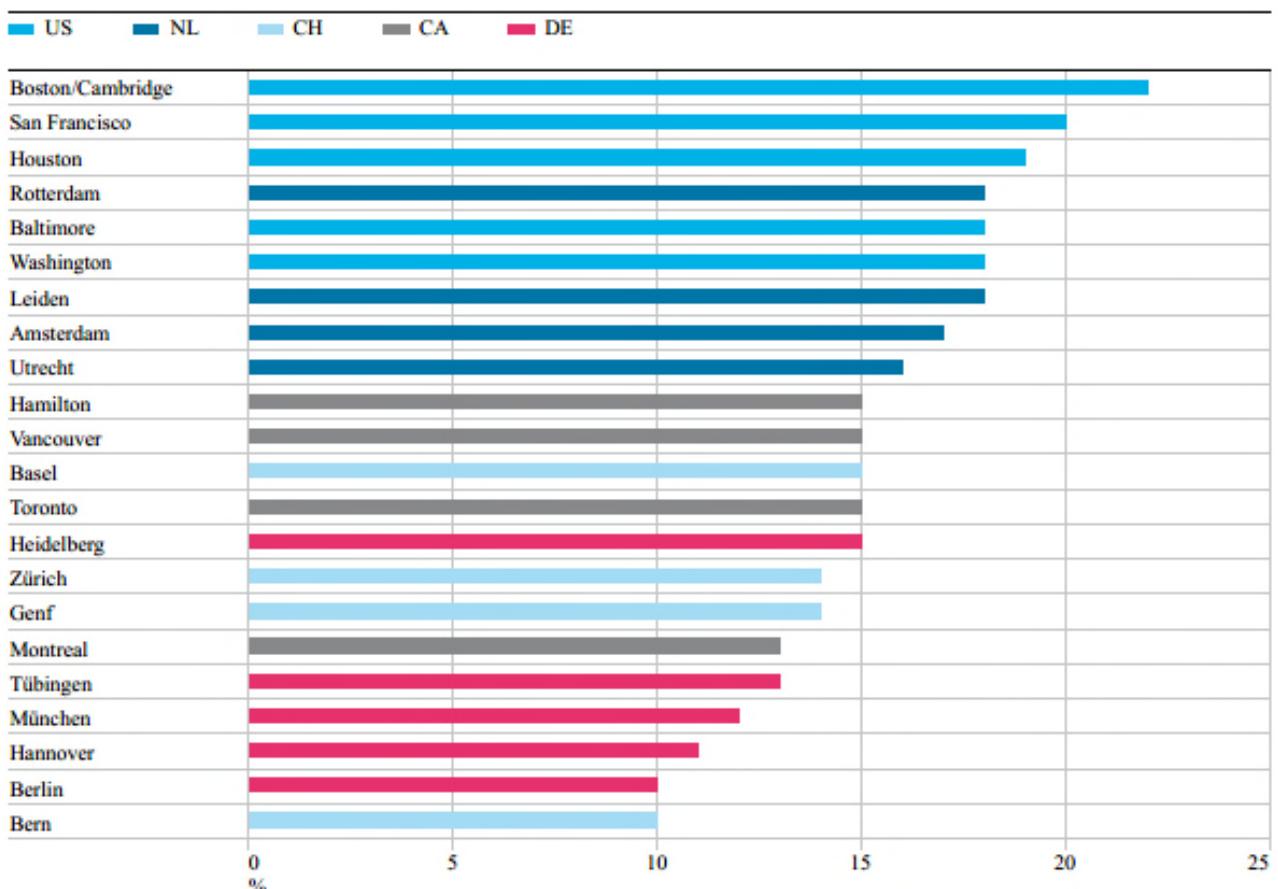
Zum anderen wird die **Exzellenzrate** definiert als die Anzahl der Publikationen, die zu den 10 % der weltweit am häufigsten zitierten Publikationen in der Medizin gehören, dividiert durch die Anzahl der medizinischen Publikationen an dem jeweiligen Standort.

$$\text{Exzellenzrate} = \# \text{ Publikationen}_{10\% \text{ am häufigsten zitierten}} / \# \text{ Anzahl der medizinischen Publikationen}$$

Diese Maßzahl zielt in erster Linie auf die Messung von Exzellenz bzw. hoher Sichtbarkeit gemessen an Zitationen. „Die höchsten Exzellenzraten innerhalb der Hochschulmedizin werden von US-amerikanischen Standorten, allen voran Boston/Cambridge gefolgt von San Francisco und Houston, erzielt. [...] Die deutschen Standorte finden sich innerhalb dieser Vergleichsgruppe zumeist auf den unteren Rängen wieder. Die Ausnahme bildet hier der Standort Heidelberg, der einen Platz im Mittelfeld behaupten kann. Obwohl am Standort Heidelberg quantitativ – sowohl insgesamt gesehen als auch pro Kopf – weniger publiziert wird als an anderen hochschulmedizinischen Standorten in Deutschland, scheint es viele Publikationen zu geben, die eine im Vergleich zu den restlichen deutschen Standorten hohe Qualität besitzen. [...] Hier ist jedoch anzumerken, dass Publikationen in US-Zeitschriften generell am häufigsten zu den Top 10 Prozent zitierten Publikationen weltweit gehören, was zumindest in Teilen auch für die vorliegende

Rangfolge verantwortlich ist. In einer US-Zeitschrift zu publizieren ist für Autoren aus den USA vermutlich immer noch leichter als für Autoren aus anderen Ländern.“⁷⁵

Exzellenzrate der Publikationen der Standorte 2010



Exzellenzrate: Anzahl der Publikationen eines Standorts, die zu den 10 Prozent der weltweit am häufigsten zitierten Publikationen in der Medizin gehören, im Verhältnis zu allen medizinischen Publikation des jeweiligen Standorts.
Quelle: Web of Science, Berechnungen des Fraunhofer ISI.

Quelle: EFI 2014, Seite 59.

4.2. Arbeitsgemeinschaft wissenschaftlicher Medizinischer Gesellschaften: Evaluation der medizinischen Forschungsleistung⁷⁶

Die Arbeitsgemeinschaft wissenschaftlicher Medizinischer Gesellschaften analysiert in ihrem Evaluationspapier zur medizinischen Forschungsleistung die existierende Evaluationspraxis in

75 Ebd. Seite 57.

76 Christoph Herrmann-Lingen et al.: AWMF-Positionspapier zur Evaluation der medizinischen Forschungsleistung; GMS German Medical Science 2014, Vol. 12, ISSN 1612-3174.

Deutschland. Aus ihren Darstellungen leiten sie die nachfolgenden Forderungen ab. Evaluatoren sollten nur nach klarer Zieldefinition durchgeführt werden und an dieser Zielsetzung ausgerichtet werden. Die Autoren sprechen sich stark für die Verwendung von Peers aus; Indikatoren könnten unterstützend eingesetzt werden. Sie fokussieren in ihrem Papier drei Komplexe, die Gegenstand von Evaluationen sein sollten: **Forschungsleistung, Nachwuchsförderung und Drittmittel.**

Zur Beurteilung der Forschungsleistung halten die Autoren als geeignete Indikatoren des medizinisch-wissenschaftlichen Impacts „neben der Rezeption in der wissenschaftlichen Fachwelt auch die Nützlichkeit für die praktische Medizin (z.B. Leitlinien-Relevanz, Praxistransfer) bzw. die Gesellschaft insgesamt (z.B. Krankheitsprävention, ökonomischer Nutzen)“ für besonders relevant⁷⁷. Die Verwendung von Impact-Faktoren lehnen sie klar ab. „Geeignete Indikatoren für erfolgreiche Nachwuchsförderung sollten sowohl die Struktur- und Prozessqualität der akademischen Lehre als auch Maßnahmen zur Förderung des postgraduierten Nachwuchses und ihre jeweiligen Ergebnisse erfassen.“⁷⁸ Zudem sollten Drittmittel (Einwerbung und Ausgaben) beurteilt werden, wobei „antragsbasierte und unabhängig begutachtete Mittel öffentlicher Fördermittelgeber oder vergleichbarer Förderorganisationen höher bewertet werden als unbegutachtete Zuwendungen anderer Herkunft“.⁷⁹

4.3. Bibliometric Study in the disciplines Public Health and Epidemiology⁸⁰

In einer Publikation der Leopoldina, acatech und der Union der deutschen Akademien der Wissenschaften mit dem Titel „Public Health in Deutschland“ aus dem Jahr 2015 wird u.a. der Frage nach den Möglichkeiten der Verbesserung von Forschung und Wissenschaft zur Steigerung der Gesundheit der Bevölkerung nachgegangen. In diesem Zusammenhang wurde eine Studie vom Institut für Forschungsinformation und Qualitätssicherung durchgeführt: „Bibliometric study for German National Academy of Sciences Leopoldina in the disciplines Public Health and Epidemiology“.⁸¹ Basierend auf der Datenbank Scopus wurden Publikationen, Publikationsdynamiken und ihr Impact untersucht. In den internationalen Vergleich gingen die Länder USA, Großbritannien, Kanada, Australien, Brasilien, Niederlande, Schweden, Spanien und Frankreich ein. Es wurde zwischen „Public Health“ und „Epidemiologie“ unterschieden. Folgende Indikatoren wurden betrachtet:

77 Ebd. Seite 11.

78 Ebd.

79 Ebd.

80 Leopoldina, acatech, Union der deutschen Akademien der Wissenschaften (2015): Bibliometric study in the disciplines Public Health and Epidemiology; im Internet abrufbar unter: https://www.leopoldina.org/fileadmin/redaktion/Politikberatung/pdf/2015_Bibliometric_study_Public_Health_and_Epidemiology_web.pdf [zuletzt abgerufen am 3. September 2018].

81 Paul Donner, Pei-Shan Chi und Valeria Aman: Bibliometric study for German National Academy of Sciences Leopoldina in the disciplines Public Health and Epidemiology; Institut für Forschungsinformation und Qualitätssicherung (iFQ); Juni 2014, im Internet verfügbar unter: <https://www.leopoldina.org/politikberatung/arbeitsgruppen/abgeschlossene-arbeitsgruppen/public-health/> [zuletzt abgerufen am 3. September 2018].

Publikationen:

- (1) absolute Publikationszahlen
- (2) Relativer Publikations-Anteil: Anteil des Landes an den Autorenadressen in der Veröffentlichung, d.h. als Lesebeispiel: sind drei Länder an einer Publikation beteiligt, zwei aus Deutschland, ein niederländisches Institut, so wird Deutschland 2/3 der Publikation zugerechnet, den Niederlanden 1/3.
- (3) Relativer deutscher Anteil: Anteil der Artikel in einem bestimmten Jahr und Fachgebiet, der von mindestens einem Autor aus Deutschland verfasst wurde.

Publikationsdynamik:

- (1) Durchschnittliches jährliches Wachstum der deutschen Publikationen: Verhältnis der Anzahl der Veröffentlichungen in einem Jahr bzgl. dem Niveau des Vorjahres
- (2) Sharpe Ratio (S): Diese Größe hängt ab von der Anzahl der Publikationen (Quantität, N) in einer bestimmten Disziplin (i), betrachtet in einem bestimmten Jahr. Es wird die Quantitätsveränderung in Abhängigkeit von der Zeit in einzelnen Disziplinen berechnet:

$$S(N_i) = \frac{\overline{D_i(t)}}{\sigma_{D_i(t)}}$$

with

$$D_i(t) = \frac{N_i(t) - N_i(t-1)}{N_i(t-1)} - \frac{N_{total}(t) - N_{total}(t-1)}{N_{total}(t-1)}$$

where $\overline{D_i(t)}$ is the average of $D_i(t)$, the differential of specific to general growth, and $\sigma_{D_i(t)}$ is its standard deviation.

Impact:

- (1) Absolute Zitationszahlen (3-Jahresfenster): Die absolute Zitierzählung ist die Anzahl der Zitate in den Publikationen,
- (2) die von einer bestimmten Untersuchungseinheit veröffentlicht werden.
- (3) Relativer Anteil an Zitierungen: Prozentsatz der Zitate der analysierten Länder zu allen Zitaten in einer bestimmten Disziplin und in einem Jahresfenster.
- (4) Zitate pro Papier eines Landes: Division der Zitate durch die Gesamtanzahl von Publikationen eines Landes.
- (5) **Field normalized citations per paper:** Zitate pro Papier (eines Landes) werden durch die entsprechende jährliche durchschnittliche Zitierrate des gesamten Feldes normiert, wodurch die normierte Zitierrate des Feldes ermittelt wird. Die Durchschnittszahl ist somit immer 1,0 für das gesamte Feld. Werte über 1,0 zeigen eine überdurchschnittliche Zitiertwirkung. Die Werte dieses Indikators können über verschiedene Disziplinen hinweg verglichen werden, im Gegensatz zu nicht normierten Indikatoren.
- (6) Uncited rate: Anteil der Papiere an, die drei Jahre nach der Veröffentlichung nicht zitiert werden.

Aus den zusammenfassenden Ergebnissen ergibt sich u.a.:

„Die 10 produktivsten Länder (nach Gesamt-Publikationszahl) sind die USA (42 958) > Großbritannien und Nordirland (13 663) > Kanada (6 723) > Australien (6 202) > Brasilien (5 343) > die Niederlande (3 926) > Schweden (2 516) > Spanien (1 709) > Deutschland (1 604) > Frankreich (1 508). [...] Deutschland weist außerdem im Vergleich zu den anderen Ländern eine relativ hohe Zitationsrate für Publikationen aus dem Bereich Public Health auf; die relative Zitationsrate hat sich auch hier in den letzten Jahren erhöht. An den nicht zitierten Publikationen hat Deutschland einen relativ geringen Anteil, das heißt, es bleiben nur wenige deutsche Forschungsergebnisse ungenutzt. [...] Die produktivste Einrichtung (Universität Heidelberg) stellte von 2000 bis 2012 einen Anteil von 9,6 Prozent der deutschen Forschungsartikel aus dem Bereich Public Health, gefolgt von der Universität Bielefeld (8,1 Prozent), der Charité Berlin (7,36 Prozent), der Universität Hamburg (7,04 Prozent) und der TU Dresden (6,61 Prozent). Die 10 am häufigsten zitierten deutschen Einrichtungen im Bereich Public Health in Deutschland sind in etwa identisch mit den 10 produktivsten deutschen Einrichtungen (siehe Abschnitt 10.2). Es gibt jedoch Variationen in der Reihenfolge der einzelnen Institute. Die Heinrich-Heine-Universität Düsseldorf ist die am häufigsten zitierte deutsche Einrichtung im Bereich Public Health und vereinte in 3-Jahres-Zitationsfenstern 429 Zitationen auf ihre Veröffentlichungen im Zeitraum 2000 bis 2010, während sie nach der Zahl der Veröffentlichungen auf Platz 7 landete.“⁸²

82 Seite 49f in: Leopoldina, acatech, Union der deutschen Akademien der Wissenschaften: Bibliometric study in the disciplines Public Health and Epidemiology (2015); im Internet abrufbar unter: https://www.leopoldina.org/fileadmin/redaktion/Politikberatung/pdf/2015_Bibliometric_study_Public_Health_and_Epidemiology_web.pdf [zuletzt abgerufen am 3. September 2018].

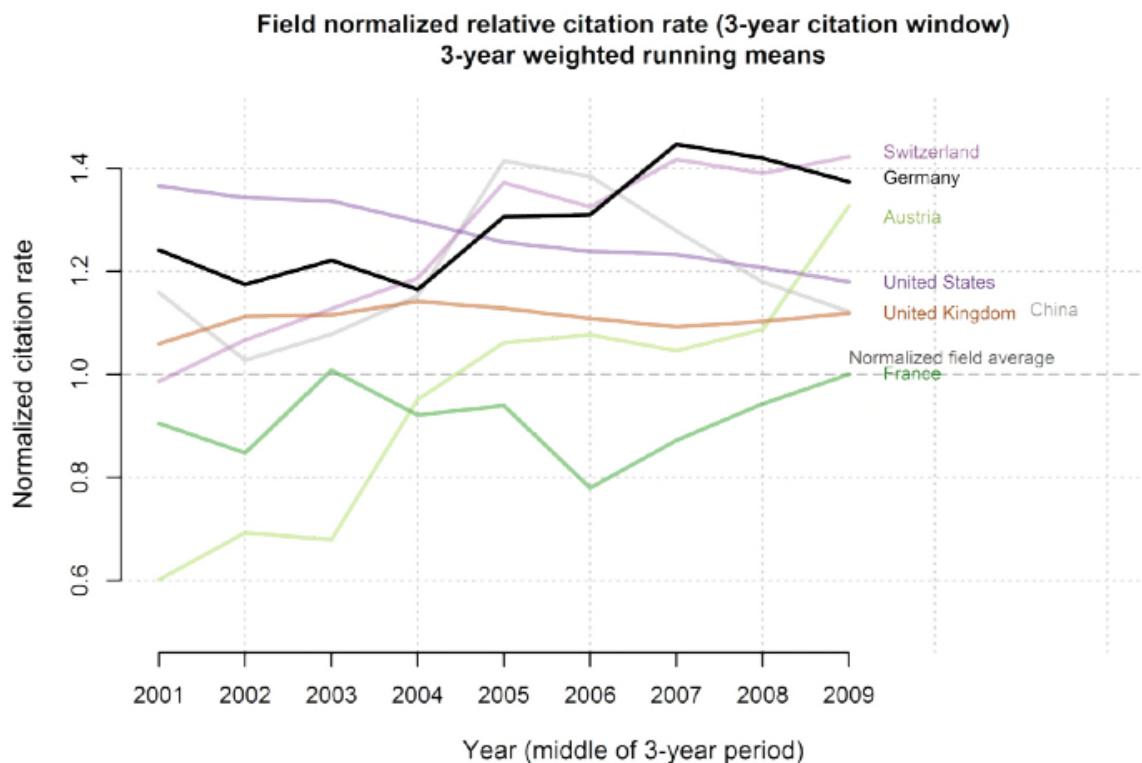


Figure 9. Average citation rates (raw and field normalized) of seven countries in public health (2000-2010).

4.4. Bewertungen laut UK Research Base 2016⁸³

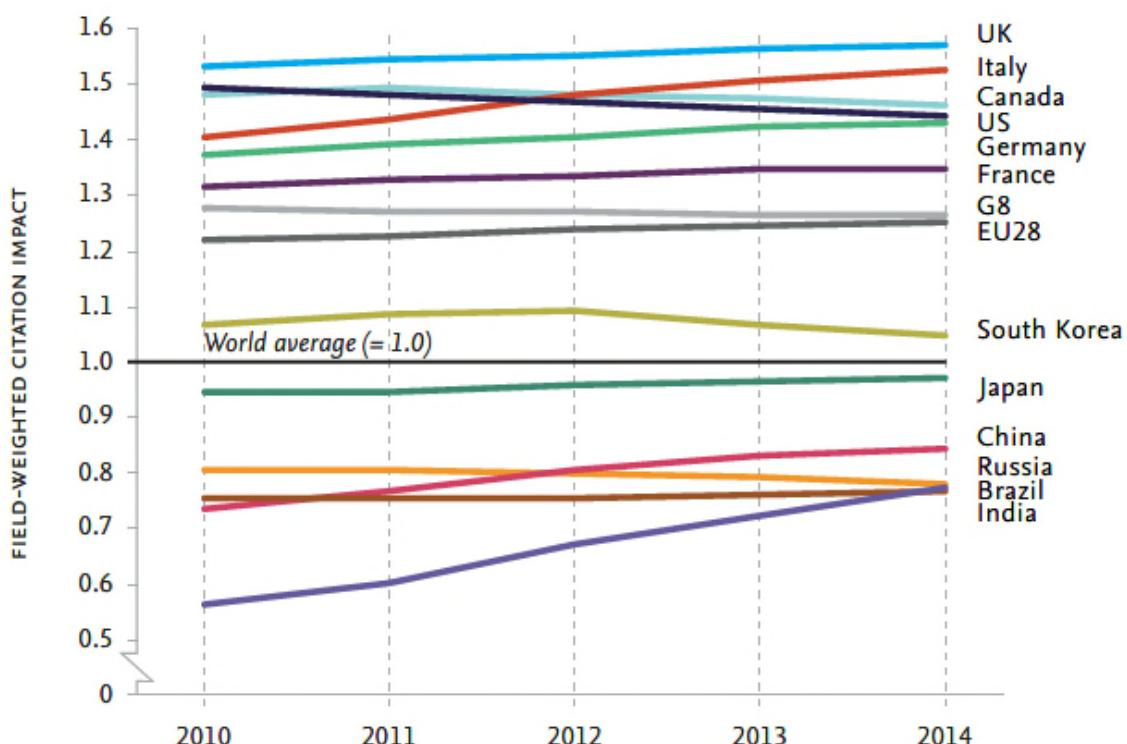
Seit Oktober 2017 liegt der im Auftrag des britischen Ministeriums für Wirtschaft, Energie und Energie erstellte Bericht „International Comparative Performance of the UK Research Base 2016“ vor. Hierin wird die Leistung des Vereinigten Königreichs (UK) im Vergleich zu sieben anderen forschungsintensiven Ländern (Kanada, China, Frankreich, Deutschland, Italien, Japan und den USA) bewertet. Es werden auch Vergleiche mit vier anderen schnell wachsenden Nationen (Brasilien, Indien, Russland und Südkorea) sowie internationale Benchmarks vorgenommen.

Auch hier wird ein gewichteter Zitationsindex herangezogen: der field weighted citation impact. Der FWCI ist ein Indikator für die Zitatwirkung und vergleicht die tatsächliche Anzahl der Zitate mit derjenigen, die von einem Artikel zu erwarten wäre (d.h. es wird durch die Anzahl der Zitate für einen Artikel desselben Typus und Referenzzeitraumes geteilt). Entspricht ein Artikel der zu erwartenden Zitationsanzahl, ergibt sich der Wert 1.0. Die nachfolgende Grafik zeigt die Ergebnisse der Jahre 2010 bis 2014. Für Italien, China und Russland ist im Referenzzeitrahmen ein

⁸³ UK's Department for Business, Energy & Industrial Strategy (BEIS): INTERNATIONAL COMPARATIVE PERFORMANCE OF THE UK RESEARCH BASE 2016; <https://www.gov.uk/government/publications/performance-of-the-uk-research-base-international-comparison-2016> [zuletzt abgerufen am 3. September 2018].

deutlicher Zuwachs zu verzeichnen. Italien erreicht 2014 nach dem Vereinigten Königreich Rang zwei. Deutschland landet bei nur geringer Steigerungsrate auf Platz 5 (hinter UK, Italien, Kanada und den USA).

Figure 4.6 — Field-weighted citation impact for the UK and comparators, 2010-2014. UK ranking in the world is amongst 92 countries with at least 1,000 articles in 2014. Source: Scopus.



Im Vergleich der beiden statistischen Grafiken der UK-Studie⁸⁴ und der Leopoldina Studie⁸⁵ ist eindeutig erkennbar, dass Deutschland in der Rangreihenfolge im europäischen Vergleich sehr unterschiedlich abschneidet. Es wurden sehr ähnliche Kenngrößen angewandt. Unmittelbar offensichtlich ist, dass in den beiden Analysen verschiedene Fächer (einmal nur Public Health, im

⁸⁴ UK's Department for Business, Energy & Industrial Strategy (BEIS): INTERNATIONAL COMPARATIVE PERFORMANCE OF THE UK RESEARCH BASE 2016; <https://www.gov.uk/government/publications/performance-of-the-uk-research-base-international-comparison-2016> [zuletzt abgerufen am 3. September 2018].

⁸⁵ Leopoldina, acatech, Union der deutschen Akademien der Wissenschaften: Bibliometric study in the disciplines Public Health and Epidemiology (2015); im Internet abrufbar unter: https://www.leopoldina.org/fileadmin/redaktion/Politikberatung/pdf/2015_Bibliometric_study_Public_Health_and_Epidemiology_web.pdf [zuletzt abgerufen am 3. September 2018].

anderen Fall verschiedene Fachbereiche, unten denen sich auch Public Health findet) betrachtet⁸⁶, unterschiedliche Länder ausgewählt, ein unterschiedlicher Zeitreferenzrahmen und Normalisierungsfaktoren angewandt wurden. Nichtsdestotrotz ist es wichtig, die grundsätzlichen Unterschiede und Hintergründe zweier daraus resultierender und in der Öffentlichkeit kommunizierter Rankings genauer zu verstehen:

Resultierendes Ranking UK-Studie:

Gemessen am field weighted citation impact der betrachteten Fächer steht Deutschland hinter UK, Italien, Kanada und den USA im Jahr 2014 auf Platz fünf der untersuchten 12 Länder und belegt damit einen Mittelfeldplatz. Das die Studie beauftragende Land, UK, schneidet sehr gut ab (erster Platz).

Resultierendes Ranking Leopoldina:

Gemessen an der field normalized relative citation rate im Bereich von „Public Health“ belegt Deutschland (das die Studie beauftragende Land) im Jahr 2009 hinter der Schweiz den zweiten Platz, im Jahr 2008 sogar den ersten Platz von insgesamt sieben ausgewählten Ländern.

86 Allerdings schneidet Gesundheitsforschung (ein Fachbereich unter vielen der UK-Studie) nicht erheblich besser ab als die übrigen in den Indikator einfließenden Fächer.