

Internationale Hochschulrankings – Eine kritische Bestandsaufnahme

Gero Federkeil

Vor zehn Jahren ist mit dem so genannten „Shanghai Ranking“ das erste globale Hochschulranking entstanden. Seitdem sind in jährlichem Abstand neue globale Rankings hinzugekommen, die die Hochschulwelt verändert haben. Der Artikel beleuchtet kritisch die Methodik der wichtigsten globalen Hochschulrankings mit Blick auf ihre methodischen Grundsätze und die verwendeten Indikatoren. Es wird ein traditionelles Modell von Hochschulrankings skizziert, das durch drei Grundelemente gekennzeichnet ist: Vergleiche auf der Ebene ganzer Hochschulen ohne Differenzierung nach einzelnen Fächern, die Berechnung von Gesamtwerten („composite indicators“) aus gewichteten Einzelindikatoren und die Zuordnung zu exakten Rangplätzen („League tables“). Abschließend wird die Wirkung internationaler Rankings unter dem Aspekt der Diversität der Hochschulsysteme diskutiert.

1 Einleitung

Hochschulrankings existieren in einigen Ländern bereits seit den 1980er Jahren; seitdem entstanden in immer mehr Ländern Rankings mit unterschiedlichen Zielsetzungen, Zielgruppen, Methoden und Indikatoren. Aktuelle Bestandsaufnahmen verzeichnen nationale Rankings in mehr als 30 Ländern; in einigen Ländern konkurrieren mehrere Rankings um die Gunst des Publikums (vgl. *van Vught/Westerheijden 2010*). Internationale Rankings gab es bis Anfang der 1990er Jahre hingegen nur wenige; diese waren zudem weitgehend akademisch geprägt, bezogen sich meist nur auf ein Fach und fanden über die Hochschulwelt hinaus kaum öffentliche Aufmerksamkeit, auch wenn sie in den betreffenden Fächern kontrovers diskutiert wurden. Das änderte sich erst vor zehn Jahren (1993) mit der ersten Publikation des „Academic Rankings of World Universities“ (ARWU),¹ des so genannten „Shanghai Rankings“, durch eine Gruppe von Wissenschaftlern der Shanghai Jiaotong University. Das Ranking wurde zunächst im Auftrag der chinesischen Regierung entwickelt, um den Standort der chinesischen Universitäten in der Forschung sowie in Naturwissenschaften und Technik im weltweiten Wettbewerb der Hochschulen zu identifizieren. Die Verfasser des Rankings, das zunächst gar nicht veröffentlicht werden sollte, waren von der weltweiten Aufmerksamkeit, die sie damit – ohne jede eigene mediale Öffentlichkeitsarbeit – errangen, selbst überrascht. Seitdem sind innerhalb von nur zehn Jahren in

¹ Die Links zu den Webseiten der hier genannten Rankings finden sich am Ende des Artikels.

beinahe jährlichem Turnus neue globale Rankings hinzugekommen. Neben dem Shanghai Ranking gehören insbesondere das Ranking der britischen Hochschulzeitung Times Higher Education und das Quacquarelli Symonds (QS) World Ranking² zu denjenigen, die international die höchste Aufmerksamkeit finden.³ Eine längere Historie haben z. B. spezialisierte, internationale Rankings von Business Schools bzw. MBA Programmen (z. B. von der Financial Times).

Daneben sind eine Reihe weiterer globaler Rankings entstanden, die spezielle Aspekte der Hochschulen beleuchten, wie z. B. das Webometrics-Ranking, das die Internetpräsenz der Hochschulen bewertet, oder ein von der französischen Ecole des Mines de Paris erstelltes Ranking, das die Hochschulen in einer Arbeitsmarktperspektive daran misst, wie viele Vorstandsmitglieder der größten Unternehmen der Welt sie hervorgebracht haben. Ausschließlich auf die Forschungsleistung der Hochschulen konzentrieren sich Rankings, die sich allein bibliometrischer Methoden und Indikatoren bedienen; Beispiele hierfür sind das Ranking der Higher Education Evaluation and Accreditation Agency of Taiwan (HEEACT) oder das „Leiden Ranking“ des auf bibliometrische Analysen spezialisierten Center for Science and Technology Studies (CWTS) der Universität Leiden.

Diese Auflistung zeigt, dass die globalen Rankings im Unterschied zu den nationalen Rankings, die zu einem Großteil von Medien bzw. Zeitschriften konzipiert und produziert werden, durch sehr unterschiedliche Typen von Organisationen erstellt werden. Dem Muster der nationalen Rankings folgt von den breit angelegten globalen Rankings lediglich das Times Higher Education Ranking; die spezialisierten internationalen Rankings von Business Schools bzw. MBA-Programmen werden hingegen in der Regel von Zeitungen/Zeitschriften gemacht (z. B. Financial Times, Business Week). Quacquarelli Symonds ist ein kommerzieller Anbieter, der jedoch nicht im traditionellen Medienbereich tätig ist. Die anderen globalen Rankings werden von Hochschulen bzw. Forschungseinrichtungen (Shanghai Ranking, Leiden Ranking, Ecole des Mines de Paris, Webometrics) oder staatlichen Evaluationsagenturen (HEEACT) erstellt. Das neue, von der Europäischen Union geförderte U-Multirank-Projekt wird von einem Konsortium aus Hochschul- und Forschungseinrichtungen und unabhängigen Organisationen entwickelt (vgl. den Beitrag von Ziegele & van Vught in diesem Band).

Die öffentliche Aufmerksamkeit, die die globalen Rankings finden, verweist auf eine spezifische Dualität von Rankings: Auf der einen Seite ist das Entstehen globaler

² Quacquarelli Symonds ist ein privater Anbieter von Informationsdienstleistungen über Hochschulen und Karrieremöglichkeiten von Hochschulabsolventen. Vgl. <http://www.qs.com/>

³ Ursprünglich veröffentlichten QS und Times Higher Education ein gemeinsames Ranking. Nach anhaltender Kritik trennte sich Times Higher Education vom alten Partner und entwickelte in Kooperation mit Thomson Reuters ein eigenes Ranking.

Rankings und die hohe Aufmerksamkeit, die sie finden, *Ausdruck* eines zunehmenden internationalen Wettbewerbs zwischen Hochschulen (um Reputation, Talente und Ressourcen). Gleichzeitig sind die Rankings aber auch *Medium* dieses Wettbewerbs; sie reproduzieren und verstärken ihn durch ihre eigenen Ergebnisse. Dies wird z. B. daran deutlich, dass sich Maßnahmen zur Förderung von Spitzenforschungsuniversitäten in vielen Ländern explizit auf das Abschneiden in den weltweiten Rankings beziehen.

Die folgende Diskussion der Methodik globaler Rankings konzentriert sich auf die drei großen, breiteren Rankings (ARWU, QS, THE), die für sich beanspruchen, die weltweit besten Universitäten, die vielzitierten „world class universities“⁴ zu ermitteln.

2. Die Methodik internationaler Rankings

2.1 Das traditionelle Ranking-Modell

Die globalen Rankings (wie auch die Mehrzahl der nationalen Rankings) folgen trotz zum Teil erheblicher Unterschiede in den Indikatoren und der Berechnung der Ranglisten einem gemeinsamen methodischen Grundansatz, der bereits dem gewissermaßen als „Mutter aller Hochschulrankings“ geltenden Ranking der amerikanischen Hochschulen durch die Zeitschrift U. S. News & World Report zugrunde liegt. Dieser traditionelle Ranking-Ansatz ist durch drei grundlegende Charakteristika geprägt:

Es wird ein Ranking *ganzer Hochschulen* über ihre einzelnen Fächer hinweg erstellt. Zwar haben sowohl ARWU als auch QS daneben auch Rankings für Fächergruppen und zum Teil auch einzelne Fächer entwickelt, doch finden überwiegend die Rankings ganzer Hochschulen die größte öffentliche Aufmerksamkeit („Die weltweit beste Universität ist...“). Zudem basiert auch die Auswahl der Hochschulen, die in die fächergruppenbezogenen Rankings einbezogen werden, zum Teil auf der Grundlage der Ergebnisse der ganzen Hochschulen.

Auf der Basis von Gewichtungen der Einzelindikatoren wird ein *Gesamtwert* („composite indicator“) errechnet, der mittels einer einzigen Kennziffer die Leistung ganzer Hochschulen bewerten will.⁵

⁴Vgl. Zum Konzept und der Diskussion um die world class universities *Salmi (2009)*

⁵Von den eingangs erwähnten globalen Rankings verzichtet einzig das Leiden Ranking auf diese Zusammenfassung der einzelnen Indikatoren.

Auf der Grundlage dieses Gesamtwertes wird eine *Rangliste* – ähnlich einer Sporttabelle – berechnet, die die Hochschulen linear vom ersten bis zum letzten Platz einzelnen Rangplätzen zuordnet.⁶

2.1.1 Rankings ganzer Hochschulen versus fachbezogene Rankings

Die globalen Rankings begannen alle mit dem Vergleich ganzer Hochschulen (institutionelles Ranking). Erst in den letzten Jahren haben das Shanghai Ranking und das QS Ranking begonnen, zusätzlich auch Ergebnisse für Fächergruppen oder einzelne Fächer zu veröffentlichen. Das Shanghai Ranking bietet Vergleiche für fünf Fächergruppen (sciences, engineering and technology, life sciences, clinical medicine/pharmacy und social sciences) sowie für fünf einzelne Fächer (mathematics, physics, chemistry, computer science und economics/business). Das QS Ranking bewertet neben dem institutionellen Ranking rund 20 Fächer, allerdings werden in einigen Fächern nur sehr wenige Indikatoren zugrunde gelegt. Das THE Ranking ist nach wie vor ausschließlich ein institutionelles Ranking ganzer Universitäten.

Ein Problem der institutionellen Rankings besteht darin, dass sie Unterschiede in der Leistungsfähigkeit der einzelnen Fächer und Fakultäten an einer Hochschule nicht darstellen können und nur einen Durchschnitt über alle Fächer hinweg bieten. In dieser Perspektive kann eine Hochschule mit sechs durchschnittlich leistungsfähigen Fakultäten gleich abschneiden wie eine Hochschule mit drei sehr guten und drei sehr schlechten Fächern. Der Informationsgehalt solcher Vergleiche für Nutzer von Rankings ist sehr eingeschränkt – er kann sogar irreführend sein, wenn sich z. B. ein Studieninteressent ausgerechnet für ein Fach an einer insgesamt recht gut bewerteten Hochschule interessiert, das zu den schlechtesten Fächern der Hochschule gehört.

Wenngleich einige Nutzer von Rankings, wie z. B. Hochschulleitungen oder auch die Politik, an Informationen über die Hochschulen als Ganzes interessiert sind, sind für die meisten Nutzer vergleichende Informationen über ein bestimmtes Fach relevanter. Studierende bzw. Studieninteressenten benötigen Informationen über das Fach, das sie studieren wollen, und nicht über die durchschnittliche Qualität einer Hochschule insgesamt. Wissenschaftler wollen sich mit den Kollegen an anderen Hochschulen in ihrem Fach vergleichen und nicht mit dem Durchschnitt aller Fächer.

2.1.2 Gesamtwerte versus multidimensionales Ranking

Alle globalen Rankings (mit Ausnahme des Leiden Rankings) sowie die überwiegende Mehrzahl der nationalen Rankings berechnen aus den gewichteten Einzelindikatoren

⁶ Im Shanghai Ranking gilt das nur für die Plätze 1 – 100; die Hochschulen danach werden jeweils nur Klassen von 50 Hochschulen (101-150, 251 – 200 usw.) zugeordnet.

einen Gesamtwert („composite indicator“; Tabelle 1 zeigt die Gewichtungen für Shanghai, QS und Times Higher Ranking). Damit misst letztlich eine einzige Zahl die Leistungsfähigkeit der komplexen Institution Hochschule (vgl. hierzu *Federkeil, van Vught, Westerheijden 2012: 43 ff.*). Die Berechnung solcher „composite indicators“ basiert notwendigerweise auf einer Gewichtung der Einzelindikatoren, die auch eine Festlegung der unterschiedlichen Relevanz der einzelnen Indikatoren durch die Produzenten der Rankings impliziert. Unterschiedliche Gewichtungsfaktoren der einzelnen Indikatoren führen dazu, dass auch Rankings, die ein ähnliches Indikatorenset verwenden, zu unterschiedlichen Ergebnissen führen können. Die Nutzung dieser „composite indicators“ ist einer der Hauptgründe für das „Unbehagen“ an Rankings (*Usher & Savino 2006: 3*)

Unterschiedliche Nutzer von Rankings können jedoch sehr unterschiedliche Präferenzen und Prioritäten bezüglich der Relevanz der Indikatoren haben. Für Studierende dürften in der Regel andere Indikatoren ausschlaggebend sein als für Wissenschaftler oder Rektoren. Aber auch Studierende sind keine homogene Gruppe: Während manche Studierende eine kleine, auf die Lehre konzentrierte Hochschule mit einem engen Kontakt zu den Lehrenden bevorzugen, sind andere an einer international ausgerichteten Universität mit hoher Forschungsorientierung interessiert. Während Rankings oft beanspruchen, entscheidungsrelevante Informationen zu liefern („to make an informed choice“), bevormunden Gesamtwerte mit vordefinierten Gewichten der einzelnen Indikatoren die Nutzer.

Zudem haben empirische Studien gezeigt, dass die Gewichtungen in den globalen Rankings alles andere als robust sind (vgl. *Saisana & d’Hombres 2008*). Kleine Veränderungen in den Gewichtungen der Indikatoren führen zu sehr unterschiedlichen Ergebnissen. Sensitivitätsanalysen (ebd.) haben gezeigt, dass der Rangplatz für 67 Prozent aller Hochschulen im THE Ranking und für 60 Prozent aller Hochschulen im Shanghai Ranking in erheblichem Maß von der Gewichtung der Einzelindikatoren abhängt. Die Unterschiede sind in den unteren Bereichen der Rankings größer, doch auch für Top-Einrichtungen wie das Massachusetts Institute of Technology (MIT) hängt das Ergebnis stark von den Gewichtungsfaktoren ab. Letztlich gibt es weder empirische noch theoretische Argumente, den einzelnen Indikatoren ein bestimmtes Gewicht – und damit eine vordefinierte Relevanz – zuzuschreiben. Dies hat bereits 1977 ein Gutachten des National Opinion Research Center zum Ranking von US News & World Report konstatiert: „*the weights used to combine the various measures into an overall rating lack any defensible or theoretical basis*“ (zitiert nach *van Vught & Westerheijden 2010: 19*). Während bei ARWU und QS die Zahl der Indikatoren, die in den Gesamtwert einfließen, mit sechs überschaubar bleibt, berücksichtigt das THE Ranking 13 Indikatoren, die fünf Dimensionen abdecken sollen und sich in ihren Gewichten deutlich unterscheiden – die Spannweite der Gewichte der Einzelindikatoren reicht von 2,25 Prozent

bis zu 30 Prozent. Dadurch lässt sich kaum noch aussagen, was der Gesamtwert letztlich misst.

Sowohl unter methodischen Gesichtspunkten als auch mit Blick auf die Nutzerorientierung sind multidimensionale Rankings eine sinnvollere Alternative: Sie stellen alle einbezogenen Indikatoren nebeneinander und überlassen die Entscheidung über die Relevanz der einzelnen Indikatoren (und in webbasierten interaktiven Rankings zum Teil auch über die quantitative Gewichtung) den Nutzern. Dadurch wird „auf effektive Weise die Definitionsmacht von „Qualität“, die eine der Schlüsselrollen ist, die die *Produzenten* von Rankings einnehmen, auf die *Nutzer* von Rankings übertragen“ (Usher and Savino 2006: 35). Dies ist auch eines der wesentlichen Grundprinzipien von U-Multirank (vgl. Ziegele & van Vught 2013 in diesem Heft).

2.1.3 Ranglisten („league tables“)

Das dritte Merkmal des traditionellen Ranking-Ansatzes (und auch fast aller nationalen Rankings) ist die Übersetzung der Gesamtwerte in exakte Rangplätze – ähnlich einer Bundesligatabelle. Daher wird der Begriff Ranking häufig synonym mit „league table“ verwendet. Dieser Ansatz suggeriert, dass jeder Rangunterschied einen Leistungsunterschied markiert („Platz 123 ist besser als Platz 127“). Eine Analyse der Rankings zeigt jedoch, dass häufig bereits kleine Unterschiede im Zahlenwert des Gesamtwerts mit großen Unterschieden in den Rangplätzen einhergehen. So unterschieden sich im aktuellen THE Ranking die Plätze 60 und 80 nur um 5,1 Punkte auf einer 100er Skala; die Plätze 80 und 120 auch um nur 5,9 Punkte. Ranglisten haben so die Tendenz, Unterschiede zwischen nahe beieinander liegenden Hochschulen zu übertreiben. Zudem können sie mit ihrer Suggestion einer Pseudogenauigkeit statistischen Unsicherheiten in den Daten nicht gerecht werden.

Diese Kritik haben einige der Rankings aufgegriffen, indem sie zumindest in den unteren Bereichen statt Rangplätzen breitere Ranggruppen gebildet haben. Im Shanghai Ranking werden beispielsweise ab Platz 100 jeweils 50er Gruppen gebildet (101–150, 151–200 etc.). Einige wenige Rankings (neben dem CHE Ranking zum Beispiel ein von der Akkreditierungsagentur Taiwans (HEEACT) publiziertes nationales Ranking, das sich primär an Studierende richtet) verzichten von vornherein auf die Berechnung von Rangplätzen und sortieren die Hochschulen nur in Gruppen.

2.2. Die Indikatoren

Während die drei einflussreichsten internationalen Rankings alle dem geschilderten traditionellen Ranking-Modell folgen, unterscheiden sie sich hinsichtlich ihrer Indikatoren. Eine Synopse der verwendeten Indikatoren und ihrer Gewichte (s. Tabelle 1)

macht deutlich, dass sich das Shanghai Ranking fast ausschließlich auf die Messung der Forschungsleistung beschränkt. Der einzige Indikator, der sich in der Zuordnung der Ranking-Produzenten auf die Qualität der Ausbildung bzw. Absolventen bezieht (Zuordnung der Nobelpreise zur Hochschule des ersten Abschlusses der Preisträger), ist ebenfalls stark forschungsorientiert. Dies entspricht der ursprünglichen Intention des Rankings, die Forschungsleistungen der chinesischen Hochschulen mit denen der führenden Hochschulen der Welt zu vergleichen, es muss jedoch bei der Interpretation der Ergebnisse berücksichtigt werden. Demgegenüber fließen in die Rankings und in die Berechnung des Gesamtwertes von QS und Times Higher Education Indikatoren ein, die unterschiedliche Aspekte und Dimensionen messen. Neben den Indikatoren zur Forschung (Publikationen, Zitationen) werden auch Indikatoren zur Lehre (z.B. Betreuungsrelation), zur internationalen Orientierung (z.B. der Anteil ausländischer Studierender) und vor allem auch zur Reputation der Hochschulen einbezogen. Damit ist im Ergebnis keine klare Aussage mehr möglich, was der Gesamtwert letztlich misst.

Tabelle1: Indikatoren und Gewichtungen

	ARWU	QS	THE
Forschung	80%	20%	42,0%
Publikationen	40%		6,0%
Zitationen	20%	20%	30,0%
Forschungsgelder			6,0%
Nobelpreise	20%		
Lehre	10%	20%	15,0%
Betreuungsrelation		20%	4,5%
Promotionsintensität			8,25%
Budget			2,25%
Alumni mit Nobelpreisen	10%		
Reputation		50%	33,0%
unter Akademikern		40%	
unter Akademikern – Lehre			15,0%
unter Arbeitgebern		10%	18,0%
Internationale Ausrichtung		10%	7,5%
Int. Studierende		5%	2,5%
Int. wiss. Personal		5%	2,5%
Int. Co-Publikationen			2,5%
Sonstiges	10%		2,5%
Größe	10%		
Industriegelder			2,5%

ARWU: Academic Ranking of World Universities; QS: Quacquarelli Symonds (QS) World Rankings;
THE: Times Higher Education World University Rankings

Dies gilt umso mehr, als im Gegensatz zum Shanghai Ranking, das ausschließlich Fakten berücksichtigt, sowohl im QS als auch im THE Ranking der Reputation der Hochschulen ein hohes Gewicht zukommt. Im QS-Ranking resultiert der Gesamtwert zur Hälfte aus der Reputation der Hochschulen unter akademischen Experten (40 Prozent) und Arbeitgebern (10 Prozent); in das THE Ranking fließt die Reputation (getrennt erfragt für Studium/Lehre und Forschung) mit einem Gewicht von insgesamt 33 Prozent ein. Das heißt, beide Rankings basieren zu einem großen Teil nicht auf Leistungsindikatoren, sondern sie messen das Renommee und den Ruf der Hochschulen. Was *Dill & Soo (2004: 8)* in ihrer Analyse nationaler Rankings beobachten, dass nämlich bei der Beantwortung von Fragen zur Reputation in Surveys das Ergebnis stärker von der vorhandenen Reputation von Hochschulen als von faktischer Kenntnis ihrer Qualität abhängt, gilt im internationalen Maßstab sicherlich noch stärker. Damit begeben sich beide Rankings in die Gefahr, eine self-fulfilling prophecy in Gang zu setzen: Sie beeinflussen die internationale Reputation der Hochschulen maßgeblich selbst durch die Messung eben dieser Reputation.

Zweifel bestehen auch an der Reliabilität der Reputationsmessungen. Die Ergebnisse von Reputationsmessungen hängen in hohem Maße von der Struktur der befragten Stichprobe ab – mit Blick auf Region, Fächer und Gruppe der Befragten. Es existiert keine Reputation an sich, sondern immer nur Reputation innerhalb einer bestimmten Gruppe. Empirische Befunde (vgl. *Federkeil 2009*) zeigen, dass insbesondere eine Messung der Reputation ganzer Hochschulen in hohem Maße von der Struktur der Stichprobe – nach Fachzugehörigkeit und Herkunft der Befragten – abhängt. Während man im QS Ranking kaum Informationen über die Zusammensetzung der Stichproben der beiden Reputationsbefragungen (unter Hochschulwissenschaftlern und Arbeitgebern) findet, die zudem sehr geringe Rücklaufquoten aufweisen, wird die Stichprobenzusammensetzung nach Regionen und Fächergruppen im THE Ranking kurz beschrieben: *“There is a balanced spread across disciplines: about 20 per cent of respondents hail from the physical sciences, a figure matched by engineering and technology, with 19 per cent from the social sciences, 17 per cent from clinical subjects, 16 per cent from the life sciences and 7 per cent from the arts and humanities.”*⁷ Die Frage, was “balanced” bzw. eine repräsentative Stichprobe darstellt, kann im globalen Maßstab jedoch kaum beantwortet werden. Soll die Gesamtzahl des akademischen Personals an allen Hochschulen weltweit als Maßstab genommen werden oder sollte angesichts der Zielsetzung, die weltweit führenden Universitäten zu identifizieren, die Stichprobe auf Wissenschaftler dieser Hochschulen begrenzt bleiben? Sollen die Ergebnisse großer Fächer proportional stärker in das Ergebnis einfließen oder sollen die Fächergruppen gleichgewichtig behandelt werden? Andere als willkürliche

⁷ <http://www.timeshighereducation.co.uk/world-university-rankings/2012/reputation-ranking/methodology>

Begründungen der Stichprobenstruktur und der Definition der Grundgesamtheit sind in diesem Zusammenhang kaum möglich.

Die Ergebnisse sind daher sehr von Zufällen und möglichen Verzerrungen in der Stichprobe geprägt. Dies erklärt auch, weshalb sich (insbesondere im QS Ranking, in dem die Hochschulen zudem selbst Vorschläge zur Auswahl der Befragten machen können) in der Reputation vieler Hochschulen von einem Jahr zum anderen teils drastische Sprünge ergeben, obwohl Reputation ein über die Zeit sehr stabiles Merkmal ist. In solchen Fällen ist eher von methodischen Artefakten als von realen Veränderungen in der Reputationshierarchie auszugehen. Die mangelnde Transparenz im QS Ranking über die Stichprobenstruktur ist insbesondere bei den fachbezogenen Rankings in den Geisteswissenschaften problematisch, da das Gesamtergebnis dort mangels anderer Indikatoren nahezu vollständig auf der Reputation basiert.

Das Shanghai Ranking beruht demgegenüber ausschließlich auf faktenbasierten Indikatoren, die die Forschungsleistung der Universitäten abbilden sollen. Im Wesentlichen (zu 60 Prozent) fließen Publikationen und Zitationen, d.h. bibliometrische Kennziffern in den Gesamtwert ein. Mit einem Gewicht von insgesamt 30 Prozent werden Nobelpreise und Fields-Medaillen (für die Mathematik) einbezogen, und zwar zum einen für Hochschulen, an denen die Preisträger zum Zeitpunkt der Preisvergabe tätig waren (mit 20 Prozent) und zum anderen für die Hochschulen, an denen die Preisträger ihren ersten akademischen Abschluss erworben haben (mit 20 Prozent). Das Ranking berücksichtigt – wenn auch mit einer geringeren Gewichtung länger zurückliegender Nobelpreise – alle Preisträger seit 1911. Das bedeutet, dass der Indikator eher historischer Art ist und nicht auf die Messung der gegenwärtigen „Exzellenz“ der Hochschulen zielt. Dies gilt umso mehr, als in der Regel viele Jahre zwischen den der Preisvergabe zugrundeliegenden Arbeiten und der Preisverleihung liegen und die Preisträger daher vielfach bei der Preisverleihung an einer anderen Hochschule tätig sind. So hat eine Auswertung der Nobel-Preise für Medizin/Physiologie aus den Jahren 1997 bis 2006 gezeigt, dass nur sieben von 22 Preisträgern ihre der Preisverleihung zugrundeliegenden Arbeiten an der Universität gemacht haben, an der sie zum Zeitpunkt der Preisverleihung tätig waren (*Ioannidis et. al. 2007*).

Diese Auswahl der Indikatoren und Datengrundlagen führt im Ranking ganzer Universitäten zu systematischen Verzerrungen. Zum einen resultiert z.B. aus der Beschränkung der Wissenschaftspreise auf Nobelpreise (sowie der Fields-Medaillen in der Mathematik), die nur in einigen wenigen Fächern und, mit Ausnahme des Nobelpreises für Wirtschaftswissenschaften, ausschließlich in naturwissenschaftlichen Fächern vergeben werden, eine systematische Verzerrung zu Ungunsten aller anderen Fächer(-gruppen) bzw. zu Ungunsten von Hochschulen mit einem breiten Spektrum anderer Fächer (vgl. dazu auch *Marginson & van der Wende 2007: 312*). Dieser bias

wird im Shanghai Ranking durch die bibliometrischen Indikatoren noch verstärkt. Aufgrund unterschiedlicher Publikations- und Zitationskulturen in den einzelnen Fächern und eines unterschiedlichen Grads der Abdeckung relevanter Publikationen in den zugrundeliegenden bibliometrischen Datenbanken wird eine Hochschule mit einer mittelmäßigen Medical School bei diesen Indikatoren immer besser abschneiden als eine sehr gute Hochschule ohne Medizin und mit einem Schwerpunkt in den Geistes- und Sozialwissenschaften. Diese Verzerrung wird im Shanghai Ranking dadurch noch weiter verstärkt, dass Publikationen in „Science“ und „Nature“ noch einmal separat, d.h. doppelt, berücksichtigt werden. Dies führt z.B. dazu, dass selbst eine Hochschule wie die London School of Economics, die in ihrem Fächerspektrum sicherlich zu den weltweit führenden Hochschulen zu rechnen ist, nur in der Ranggruppe zwischen Platz 101 und 150 liegt, nachdem sie lange Jahre sogar nur zwischen Platz 201 und 300 gerankt war.

Auf der Ebene des institutionellen Rankings ganzer Hochschulen laufen die bestehenden globalen Rankings daher Gefahr, eher Unterschiede in der Fächerstruktur der Hochschulen abzubilden als wirklich die Leistungsfähigkeit der Hochschule mit ihren einzelnen Fächern zu messen. Auf der Fächerebene mangelt es ihnen zudem an validen und international vergleichbaren Indikatoren, insbesondere in den Geistes- und Sozialwissenschaften, deren Publikationskultur (hohe Bedeutung von Büchern; Publikationen in der jeweiligen Mutter- und Kultursprache) bislang von den beiden großen bibliometrischen Datenbanken nur unzureichend abgedeckt werden, da beide trotz einer Verbreiterung ihrer Datenbasis in den letzten Jahren nach wie vor stark auf Zeitschriftenaufsätze beschränkt sind.

3 Wirkungen der Rankings und Diversität

Die bestehenden globalen Rankings ranken in der Regel 500 bis 600 Hochschulen. Damit bilden sie nur einen sehr kleinen Anteil aller Hochschulen weltweit ab. Schätzt man die Zahl der Hochschulen weltweit auf rund 17.000, so entspricht dies etwa 3,5 Prozent aller Hochschulen (*Rauhvargers 2011: 13*). Dies hat bezüglich der Aussage der Rankingergebnisse zwei Implikationen: Erstens gehören die unter den Top 500 gelisteten Universitäten, wenn man den Rankings zugesteht, Forschungsexzellenz zu messen, weltweit zu den besten 3,5 Prozent aller Hochschulen. In diesem Zusammenhang relativieren sich Aussagen, dass eine Hochschule „nur“ auf Platz 53 gerankt ist (wie die aktuell bestplatzierte deutsche Universität im Shanghai Ranking). Zweitens umfassen die globalen Rankings aufgrund ihrer Methodik und Indikatoren faktisch nur einen spezifischen „Typ“ von Hochschule, die international orientierte Forschungs(voll) universität. Hochschulen mit anderen Profilen und Zielsetzungen sind nicht Gegenstand der Rankings und haben keine Chance, sichtbar zu werden.

In dieser ausschließlichen Konzentration auf Forschungsexzellenz im internationalen Maßstab liegt hochschulpolitisch das größte Problem der globalen Rankings traditioneller Art. In der öffentlichen Rezeption – wie auch in den politischen Entscheidungen, die in einigen Ländern an Rankingergebnisse geknüpft sind – erscheinen die Hochschulen, die nicht unter den Top 500 auftauchen, undifferenziert schlechter als diese Top 500-Hochschulen. Dabei ist die überwiegende Zahl dieser Hochschulen schlichtweg anders: Sie haben andere Zielsetzungen und andere Profile. Die mit den globalen Rankings verbundene ausschließliche Fokussierung auf Forschungsexzellenz hat zu einer Obsession über die so genannten „world class universities“ (Salmi 2009) geführt⁸, wobei es den Rankings weitgehend gelungen ist, die Definitionsmacht des Konzepts „Weltklasse“ für sich zu gewinnen. So häufen sich Berichte, dass in einigen Ländern die Förderung von internationalen Hochschulkooperationen oder des Austausches von Studierenden auf Hochschulen beschränkt werden, die in einem der drei großen globalen Rankings unter den Top 200 gerankt sind.

In vielen Ländern haben die globalen Rankings das Bemühen verstärkt, hochgerankte Forschungsuniversitäten zu schaffen, die sowohl als Symbol der Leistungsfähigkeit der nationalen Hochschulsysteme als auch als Motor für wirtschaftliches Wachstum in der Wissensgesellschaft betrachtet werden (Marginson & van der Wende 2007). In einer Reihe von Ländern sind nationale Exzellenzprogramme mehr oder weniger explizit von den Ergebnissen der globalen Rankings, und damit häufig von dem wahrgenommenen schlechten Abschneiden der eigenen Hochschulen⁹ stimuliert worden. Vergleichbares gilt für Fusionen von Hochschulen (z. B. in Dänemark), da die Größe der Hochschule ebenfalls als ein Erfolgsfaktor in den globalen Rankings identifiziert wurde.

Mit diesen Entwicklungen ist die Gefahr verbunden, dass staatliche Mittel zunehmend auf wenige Forschungsuniversitäten konzentriert werden. Dies führt zu einer Abwertung anderer Profile als der Forschungsexzellenz und ist letztlich eine Bedrohung der Vielfalt in den Hochschulsystemen. „World class“ Hochschulsysteme brauchen eine Vielfalt an Hochschulen mit unterschiedlichen Profilen, Schwerpunkten und Zielsetzungen. Diese Vielfalt darzustellen ist eine Herausforderung für multidimensionale und fachbezogene Rankings, die die beschriebenen Schwachpunkte der existierenden Rankings vermeiden. Ein spezifischer Ansatz für die Entwicklung eines solchen internationalen Rankings wird in U-Multirank verfolgt. Auf der Grundlage einer Machbarkeitsstudie (van Vught & Ziegele 2011, 2012) wird dieses Konzept zurzeit in einem von

⁸ Google notiert (Stand April 2013) 13,7 Millionen Treffer zu den exakten Suchbegriffen „world class university“ bzw. „world class universities“.

⁹ Nach der Publikation des ersten Shanghai Rankings erschien in der französischen Zeitung *Le Monde* beispielsweise ein viel diskutierter Artikel mit der Überschrift „The great misery of French universities.“ (24. Januar 2004)

der Europäischen Union geförderten Projekt implementiert (vgl. dazu den Beitrag von Ziegele & van Vught 2013 in diesem Heft).

4 **Schlußfolgerungen**

Internationale Rankings haben als Ausdruck eines wachsenden globalen Wettbewerbs um Ressourcen, Talente und Reputation in den letzten zehn Jahren enorm an Bedeutung gewonnen. Gleichzeitig verstärken sie diesen Wettbewerb durch ihre eigenen Ergebnisse. Dies hat zu einer Obsession über die so genannten „world class universities“ geführt: „Jeder will eine sein/haben, niemand weiß, was es ist und niemand weiß, wie man eine wird“ (Altbach 2004). Mittlerweile erfolgt die Definition vielfach über die Aufnahme in die Top 200 eines der drei maßgeblichen globalen Rankings.

Allerdings ist die Aussagekraft der bisherigen internationalen Rankings in mehrfacher Hinsicht beschränkt:

- Sie sind aufgrund der Auswahl ihrer Indikatoren (z. B. bibliometrischer Indikatoren, Nobelpreise) und der dahinter liegenden Datenquellen Rankings nur eines spezifischen Typs von Hochschule: der international orientierten Forschungs(voll)universität, der aufgrund des Erklärungsanspruchs der Rankings, die besten Universitäten zu ermitteln, zum Allgemeinmodell normiert wird.
- Auch wenn einige der internationalen Rankings versuchen, ansatzweise Indikatoren zu Studium und internationaler Orientierung aufzunehmen, sind sie weitgehend auf die Messung von Forschungsexzellenz ausgerichtet. Andere Dimensionen des Leistungsspektrums der Hochschulen wie Studium & Lehre oder Wissenstransfer bleiben weitgehend unberücksichtigt. Damit ist der Nutzen der Rankings für Studierende, die eine Entscheidung über ihre Hochschule treffen wollen, sehr eingeschränkt.
- Ebenfalls mit der Auswahl der Indikatoren und Datenquellen einher geht ein Bias zugunsten der Naturwissenschaften. Die Forschungsleistung in den Geistes- und Sozialwissenschaften, zum Teil auch der Ingenieurwissenschaften, kann auf der Basis der zugrundeliegenden bibliometrischen Datenbanken, die überwiegend Zeitschriftenaufsätze erfassen, nur unzureichend abgedeckt werden.
- Die Fokussierung auf einen medienwirksamen Gesamtwert verknüpft mit einer Rangliste schafft die Illusion einer eindeutigen Leistungshierarchie („Nummer vier ist besser als Nummer 7“), verstellt jedoch den Blick auf Profilunterschiede und die spezifischen Stärken und Schwächen der einzelnen Hochschulen. Damit bietet das traditionelle Ranking-Modell kaum Erkenntnisse, die für die Stärken-Schwächen-Analyse und die strategische Weiterentwicklung in den Hochschulen selbst genutzt werden könnten.

Rankings stehen in einem Dilemma: Einerseits wenden sie sich an eine breites (Laien-) Publikum außerhalb des Hochschulsystems selbst. Dies erfordert eine Reduzierung der Komplexität von Information, um für diese Zielgruppen – z. B. Abiturienten/Studieninteressenten – verständlich und anschlussfähig zu sein. Andererseits müssen sie sich an Qualitätsstandards empirischer Forschung orientieren, und die Expertennutzer von Rankings innerhalb des Hochschulsystems erwarten ausreichend differenzierte und laborierte Informationen. Eine Balance zwischen diesen beiden konfligierenden Anforderungen zu finden, ist eine Herausforderung für alle Rankings. Die bisherigen internationalen Rankings orientieren sich mit vordergründig klaren Gesamtwerten und eindeutigen Rangpositionen eher an der Simplifizierung.

Künftige Rankings, die für verschiedene Nutzergruppen entscheidungsrelevante Informationen liefern müssen (z. B. für die Hochschulwahl für Studierende; für das strategische Management in Universitäten) sollten den Nutzern den Verzicht auf allzu vereinfachende Informationen zumuten und versuchen, ein differenzierteres Bild von komplexen Organisationen, wie Hochschulen es sind, zu vermitteln. Dazu gehören eine Differenzierung nach einzelnen Fächern/Fächergruppen, die einige der globalen Rankings angefangen haben, die aber in ihrer Positionierung noch deutlich hinter den Rankings der ganzen Universitäten zurückbleibt, sowie insbesondere ein multidimensionales Ranking, das die Tatsache berücksichtigt, dass es keine allgemein gültige Definition von Qualität und guter Performance an Hochschulen gibt (wie die composite indicators suggerieren). Unterschiedliche Nutzergruppen und Nutzer haben verschiedene Präferenzen und Prioritäten bezüglich der Relevanz von Leistungsdimensionen und Indikatoren. Dies sollten Hochschulrankings – nationale wie internationale – im Zeitalter interaktiver Darstellungsmöglichkeiten, die nicht mehr auf das Abdrucken vordefinierter, fester Listen angewiesen sind, in ihren Konzepten und Darstellungsformen berücksichtigen.

Übersicht internationaler Rankings:

Academic Ranking of World Universities (ARWU):

<http://www.shanghairanking.com/>

Ecole des Mines Paris:

<http://www.mines-paristech.eu/About-us/Rankings/professional-ranking/>

HEEACT: Performance Ranking for Scientific Rankings of World Universities

<http://ranking.heeact.edu.tw/en-us/2011/homepage/>

Leiden Ranking

<http://www.leidenranking.com/>

QS World University Rankings

<http://www.topuniversities.com/university-rankings>

Times Higher Education World University Rankings

<http://www.timeshighereducation.co.uk/world-university-rankings/>

Webometrics Ranking:

<http://www.webometrics.info/>

Literatur

Altbach, Philip (2004): The Costs and Benefits of World-Class Universities. *Academe* 90 (1, January-February).

Dill, David; Soo, Maarja (2005): Academic quality, league tables and public policy: A Cross-National Analysis of University Ranking Systems. *Higher Education* Vol. 49, 495–533.

Federkeil, Gero (2009): Reputation indicators in rankings of higher education institutions, in: Barbara M. Kehm und Bjorn Stens (Hrsg.): *University Rankings, Diversity, and the New Landscape of Higher Education*, Rotterdam, Taipei, 2009, S. 19–34

Federkeil, Gero; van Vught, Frans; Westerheijden, Don (2012): An Evaluation and Critique of Current Rankings. In: van Vught, Frans u. a. (Hrsg.): *Multidimensional Ranking. The Design and Development of U-Multirank*. Heidelberg/London/New York, S. 39–70

Ioannidis, John PA; Patsopoulos, Nikolaos A.; Kavvoura, Fotini K; etl a. (2007): International ranking systems for universities and institutions: a critical appraisal. *BMC Medicine* Vol. 5, 30-; download: <http://www.biomedcentral.com/1741-7015/5/30>

Marginson, Simon; van der Wende, Mareijk (2007): To Rank or to Be Ranked. The Impact of Global Rankings in Higher Education. *Journal of Studies in International Education* Vol. 11, S. 306–329.

Rauhvargers, Andrejs: Global University Rankings and their Impact. *EUA Report on Rankings 2011*. Brüssel, EUA.

Saisana, Michaela; d' Hombres Beatrice (2008): Higher Education Rankings. Robustness Issues and Critical Assessment. How Much Confidence Can We Have in Higher Education Rankings. Office for Official Publications of the European Communities, Luxemburg.

Salmi, Jamil (2009): The Challenge of Establishing World-Class Universities. Washington, D.C., The World Bank.

Usher, Alex; Savino, Massimo (2006): A World of Difference. A Global Survey of University League Tables. Toronto, Educational Policy Institute.

van Vught, Frans; Westerheijden, Don F. (2010): Multidimensional ranking: a new transparency tool for higher education and research. Higher Education Management and Policy, Vol. 22.

van Vught, Frans; Ziegele, Frank (Hrsg.) (2011): Design and Testing the Feasibility of a Multidimensional Global University Ranking. Final Report. Download: http://ec.europa.eu/education/higher-education/doc/multirank_en.pdf (Zugriff: 09. Januar 2013)

van Vught, Frans; Ziegele, Frank (Hrsg.) (2012): Multidimensional Ranking. The Design and Development of U-Multirank. Heidelberg/London/New York, Springer.

Ziegele, Frank; van Vught, Frans (2013): U-Multirank“ und „U-Map“ als Ansätze zur Schaffung von Transparenz im europäischen und globalen Hochschulsystem – Konzepte und Erfahrungen. In: Beiträge zur Hochschulforschung 2/2013, S. 50–74

Anschrift des Verfassers:

Gero Federkeil
Centrum für Hochschulentwicklung CHE
Verler Straße 6
33332 Gütersloh
E-Mail: Gero.Federkeil@che.de

Gero Federkeil ist Projektmanager am CHE Centrum für Hochschulentwicklung und Vice President von IREG Observatory on Academic Ranking and Excellence.