

Dieter Lenzen

Ranking, Rating – Steuerung und Motivation.
Erfahrungen und Befunde zum Forschungsrating aus Sicht einer
Universitätsleitung

Vortrag gehalten am 21.9.2012 bei der „Tagung zur Bedeutung des Forschungsratings als Instrument der strategischen Steuerung und Kommunikation von Hochschulen und Forschungseinrichtungen“ veranstaltet vom Wissenschaftsrat und dem Stifterverband für die Deutsche Wissenschaft

Es ist einiges geschehen seit dem ich gebeten wurde, Wünsche an künftige Leistungsdarstellungen von Hochschulen zu formulieren:

0. Die Lage

- Wir haben gelernt, dass Moody's in der Lage ist, die Welt an den Abgrund zu führen. Nachdem diese und andere Finanzratingagenturen in der früheren Vergangenheit außer Stande waren, irgendeine Wirklichkeit vorherzusagen, produzieren sie sie nun selbst durch die breite Kommunikation ihrer Up- und Downgradings ganzer Nationen.
- Bei der Londoner Olympiade durften wir wieder einmal wieder die primitivste Form des Rankings betrachten. Eine Gruppe muskulöser Männer oder Frauen rennt um die Wette, 100-Meter, Erster Platz: 9,63 Zweiter Platz: 9,75, Dritter Platz: 9,79. Wir sehen keine Unterschiede, wenn wir selbst hinschauen, aber das Messinstrument sieht, nein produziert sie. Und, noch schlimmer: Die Bronzemedailien-Gewinnerin im Fechten sei zwei Ränge schlechter als der Goldmedailien-Gewinner im Hochsprung, obwohl es sich um ganz unterschiedliche Sportarten handelt.
- Und: Was wir alle kennen, weil wir alle zur Schule gegangen sind: Schulnoten. Wir wissen, PISA sei Dank, dass eine 1 in Mathe im Norden unseres Landes so viel Wert sein kann wie eine 3 im Süden. Und wir wussten immer schon, dass seit einem berühmten Experiment in den 50er Jahren für ein und denselben Deutschaufratz alle Noten von 1-6 vergeben werden, wenn man den Aufsatz verschiedenen Lehrern ohne Kenntnis der Autoren zur Bewertung gibt. Und wir wissen, das manchmal Mädchen, brave zumal, um zwei Notenstufen für dieselbe Leistung besser bewertet werden als wilde Jungs.
Wer sich dieses Geschehen, etwa als Außerirdischer vom Sirius kommend, betrachtet, der muss sagen: Die spinnen, die Erdlinge.

Da Wissenschaftler und Wissenschaftlerinnen vielleicht nicht ganz so extensiv spinnen wie Finanzmakler, Oberschulräte und das IOK, prasselt im Wissenschaftsbereich die Kritik an Rankings und Ratings vernehmlicher. Nachdem zunächst die Historiker, sodann etliche Medizinische Fakultäten nun auch die Deutsche Gesellschaft für Soziologie und die Gesellschaft Deutscher Chemiker zum Boykott von Rankings und Ratings aufgerufen haben und etliche Universitäten (Leipzig, Köln, Lüneburg, Hamburg, Vechta, FernUni Hagen) Daten überhaupt nicht mehr liefern oder nur gegen Vollkostenrechnung, was auf das gleiche hinausläuft, verlieren Wissen-

schaftsrangings nicht nur an Reputation, sondern auch an Validität, wenn sie eine solche jemals besaßen. Das muss man jedenfalls bezweifeln, wenn man zur Kenntnis nehmen musste, dass das Times Higher Education Ranking vor ca. 2 Jahren offen zugab, dass ihre Daten falsch waren, wenn man zur Kenntnis nimmt, dass in den USA bereits gut verkaufte Bücher existieren mit Ratschlägen, wie man durch „Gaming“ seine Ranking- Ergebnisse gezielt, auch in Absprache, verfälschen kann, wenn man zur Kenntnis nimmt, dass ein und dieselbe Universität in 5 verschiedenen Rankings einmal auf einem dritten Platz und ein anderes Mal auf einem dreistelligen landet (vgl. Glaser, 2012, S. 210) und wenn man zur Kenntnis nimmt, dass Phil Baty, Herausgeber des Times Higher Education Rankings nonchalant im Interview mit der Süddeutschen sagte: „Ich würde nie eine Entscheidung auf Grund eines Rankings treffen.“ (Interview mit Maria Holzmüller, Süddeutsche Zeitung vom 06.06.11).
Warum machen Menschen dann so etwas?

1. Zur Geschichte der Leistungsskalierung

Für diese Frage gilt erstens der Satz: Das kann man nur historisch erklären. Sozialhistorisch betrachtet haben diejenigen, die Rankings einführten, immer spezifische Probleme zu lösen gehabt. Peter der Große ersetzte mit Hilfe von Dienstgraden, auch einem Ranking, den Erbadel und versuchte damit Leistungsgerechtigkeit und -adäquatheit herzustellen, die Olympiade diente eigentlich der Völkerverständigung, deswegen der Satz von Coubertin: „Nicht siegen, dabei sein ist wichtig.“ Moody's und ihre Artgenossen wollten Banken und Finanzministern dabei helfen, Kredite nicht an insolvente Kunden zu vergeben oder wenigstens dafür hohe Zinsen zu verlangen und auch die Schulnoten dienten dem Zweck, dem Prinzip Bildungsleistung gegenüber sozialer Herkunft Geltung zu verschaffen. Für Leistungsvergleiche im Wissenschaftsbereich gab es andere Motive: Das erste US-amerikanische Ranking in den 80er Jahren, das US News and World Report diente der Rechtfertigung von hohen Studiengebühren als Kundeninformationen über die Qualität der gebührenerhebenden Hochschulen, der erste Zitationsindex war dafür gedacht, Bibliothekaren Entscheidungshilfen bei der Bestellung von Zeitschriften zu liefern. Ob ihr Reputationsverfall damit ursächlich zusammen hängt, wird in Dissertationen zu klären sein. Anthropologisch betrachtet mag man Leistungsvergleiche und Wettbewerb als evolutionären Selektionskatalysator werten. Zumindest spricht das bei Hühnern beobachtete Ranking, die Hackordnung dafür. Dominante Hühner leben einfach länger.

2. Gelten die historischen Gründe noch?

Wir müssen uns nun zweitens fragen, ob die Entstehungsgründe eigentlich noch Gültigkeit haben, so dass Rankings auch im Wissenschaftsbereich weiterhin gerechtfertigt wären. Die Antwort ist einfach, sie heißt ja und nein. Natürlich sind die Entstehungsgründe zwischen Völkerverständigung und bibliothekarischer Hilfeleistung entfallen. Das Grundmotiv, das allen Vergleichen zu Grunde liegt, ist aber letztlich ein ökonomisches. Menschen möchten sich Vorteile verschaffen durch Vergleiche und als die jeweils besseren allererst dastehen. Indessen sind neue Gründe entstanden:

In der Literatur werden als Begründung für die Durchführung von Universitätsrankings bzw. Ratings in der Regel genannt

- Transparenz der Leistungsfähigkeit einer Hochschule für Studierende und deren Eltern
- Rekrutierungshilfe für akademisches Personal bei den Unternehmen (Heinemann 2012)
- Karriereplanung für das akademische Personal.

Letztere Absicht wird insbesondere durch das Times Higher Education Ranking in Anspruch genommen. Die Unternehmen, so die Einschätzung, werden am besten durch das CHE bedient, und mit Abstrichen noch durch das Ranking der Wirtschaftswoche. So werden dem CHE-Ranking valide Informationen über das Lernergebnis der Absolventen zugeschrieben, über die Qualität des Personals und der Ressourcen, des Lernumfeldes sowie über die Bedingungen, unter denen eine Hochschule Studierende überhaupt zulässt, über die Praxisorientierung der Curricula, die Internationalität der Hochschule und die Transferleistungen in der Forschung für Industrie und Gesellschaft. Transparenzbeschaffung für Studierende ist das erklärte Ziel des CHE selbst. Das gilt natürlich auch für Internationale Rankings, bei denen diffus angenommen wird, dass beispielsweise im Falle des Shanghai -Rankings der Aufenthalt eines längst verbliebenen Nobelpreisträgers von 1910 eine Art Genius Loci hervorgerufen hat, der die Studienbedingungen nachhaltig prägt und deshalb hohe Studiengebühren rechtfertigt.

In Wirklichkeit gilt für all diese Leistungsvergleiche aus privatwirtschaftlichen, zumindest in den USA, Händen: Es ist ein großes Geschäft. Das ist nicht zu beanstanden, hat aber Folgen für die Qualität, etwa wenn ein Großteil der erhobenen Daten überhaupt nicht veröffentlicht wird und so eine hohe Informationsselektivität existiert. Insoweit die Kunden solcher Rankings als Erwachsene für ihre Entscheidungen selbst verantwortlich sind, muss man das nicht kritisieren; umgekehrt muss aber ebenso klar sein, dass die öffentliche Hand sich auf keinen Fall an Daten orientieren kann, die nicht selbst durch öffentliche Einrichtungen beurkundet sind. Dementsprechend ist mit Besorgnis im Beschluss der KMK vom 12.05.2011 über die „Instrumente der Qualitätsfeststellung in der Hochschulforschung“ zu lesen, dass eben nicht nur das Förderranking des DFG und das Rating des Wissenschaftsrats zwei absolut seriöse Datenquellen, sondern auch privatwirtschaftlich erstellte Leistungsvergleiche nach Aussagen einer Reihe von Ministerien herangezogen werden zur Beurteilung ihrer Hochschulen und deren Position im Wettbewerb (Rheinland-Pfalz), als „wertvolle Elemente für die vergleichende Betrachtung“ (Baden-Württemberg), für die Außendarstellung der Hochschulen (Baden-Württemberg, Brandenburg, Bayern) und zur Qualitätsmotivation sowie für Wettbewerb und Profilbildung (Baden-Württemberg, Bayern, Brandenburg), wobei Bayern allerdings solche Leistungsvergleiche explizit für „wissenschaftspolitische und forschungspolitische Entscheidungen für überwiegend ungeeignet“ hält (KMK 2011 S. 14).

Fazit: Die historischen Gründe für Leistungsvergleiche im Wissenschaftsbereich gelten fast ausnahmslos nicht mehr, an ihre Stelle sind andere getreten, die

entweder Entscheidungen von Nutzern des Hochschulsystems optimieren sollen oder von den Betreibern, also der Politik und der Hochschulleitungen. Dabei wird nicht selten auch die Auffassung vertreten, dass Leistungsvergleiche eine Motivation für Leistungssteigerungen mit sich führen. Ich komme darauf zurück.

3. Bedingungen des Skalierens

Wenn Leistungsvergleiche Steuerungsfunktionen wahrnehmen sollen, dann müssen sie methodisch so gestaltet sein, dass sie diesen Zweck auch tatsächlich erfüllen, das heißt, dass sie die Wirklichkeit, die sie zu beschreiben behaupten, auch tatsächlich abbilden, damit gelten für sie die drei methodischen Grundgütekriterien der Objektivität, der Reliabilität und der Validität. Alle drei Kriterien sind sogenannte Unabhängigkeitsmaße. Das heißt, dass das Ergebnis einer Untersuchung unabhängig sein muss von denen, die die Untersuchung durchführen, unabhängig von dem Zeitpunkt und den Umständen der Untersuchungsdurchführung, das heißt sie müssen replizierbar sein, und sie müssen unabhängig von Fremdeinflüssen sein, also das auch messen, was sie zu messen vorgeben und nicht irgendetwas anderes.

Zur Objektivität: Untersuchungsdurchführende sind die jeweilige Agentur und die Einrichtung, die die Daten erhebt und zur Verfügung stellt, also in der Regel die Hochschule selber. In beiden Fällen ist eine Unabhängigkeit nicht gegeben. Hochschulleitungen kennen viele Tricks, um Daten zu fälschen und zu beschönigen, auf das Stichwort Gaming habe ich bereits hingewiesen. Gaming findet allerdings auch schon bei den Hochschullehrern selber statt, wenn sie etwa mit berühmten Kollegen gemeinsam publizieren, um in den Genuss besserer Zitationsraten zu gelangen. Die methodische Frage ist also nicht, wie man Gaming ausschließt, sondern ob öffentliche Mittel wirklich dafür verwendet werden sollen, dass Hochschulleitungen und Verwaltungen sich zeitaufwendige Tricks ausdenken, um Agenturen zu betrügen. Denn eins ist sicher: Die Wissenschaftler sind auf diesem Feld immer besser. Eine Unabhängigkeit von den Mess-Agenturen ist sodann auch nicht gegeben. So sind es beispielsweise Experten, die die Auswahl von Zeitschriften oder von Indikatoren vornehmen. Sie bestimmen die Indikatoren selbst, deshalb die unterschiedlichen Rangplätze, und sie entscheiden über die Publikationsformen. Sie entscheiden auch darüber, was für eine ausreichende Fallzahl gehalten wird und ob im Falle der Nichtantwort gemahnt wird, - die deutsche Gesellschaft für Soziologie hat darauf hingewiesen. Insoweit Urteile von Hochschullehrern über Institutionen eingeholt werden, entsteht methodisch sogar ein gezieltes Nichtobjektivitätsmaß. Hier werden Einstellungen und Vorurteile zu Messkriterien gemacht. Dieses ist sozialwissenschaftlich gesehen eigentlich nur in der Einstellungs- und Vorurteilsforschung selbst sinnvoll. Besonders hinfällig sind auch die Tatsachen, auf die Kieser aufmerksam gemacht hat, beispielsweise auf die schwache Korrelation zwischen Gutachterurteilen bei Aufsatzbewertungen für die Annahme in den sogenannten A-Journals und in späteren tatsächlichen Zitationen, weswegen der Supreme Court in den USA Rückschlüsse auf die Qualität von Wissenschaftlern verboten hat, soweit sie sich auf Zitationsindizes bezieht. Hübsch auch Experimente mit bereits veröffentlichten

Manuskripten, die von sogenannten Experten nicht wiedererkannt wurden oder, um noch eines draufzusetzen, im Bereich der Befragung von Personalverantwortlichen. Hier wurden die Datengeber nach der Qualität von Studiengängen an Hochschulorten gefragt, die gar nicht existierten und dennoch von den Personalverantwortlichen hoch gerankt wurden.

Eine Unabhängigkeit von Zeitpunkt und Umständen der Untersuchung ist in der Regel gleichfalls nicht gegeben, anders sind die extrem hohen Schwankungen in den Rangplätzen einzelner Hochschulen kaum erklärbar. Denn keine Hochschule kann es schaffen, innerhalb eines Jahres von einem 80igsten auf einen 5. Listenplatz aufzusteigen, selbst dann nicht, wenn das ganze Personal auf Schlaf und Urlaub verzichtet.

Neben der notwendigen Unabhängigkeit von Untersuchern und Untersuchungs-umständen ist schließlich natürlich wesentlich, dass eine Leistungserhebung auch tatsächlich misst, was sie zu messen behauptet. Eine der diesbezüglichen entscheidenden Schwachstellen besteht darin, dass für Steuerungszwecke Leistungen von Fakultäten unterschiedlicher Fächer versucht werden vergleichbar zu machen, was grundsätzlich, auch durch die Beiführung von Koeffizienten nicht möglich ist. Ähnlich unsinnig wäre es, die Goldmedaille eines Biathlon Kämpfers mit der eines Dressurreiters zu vergleichen. Die ruhige Hand beim Abfeuern eines Schusses ist eben nicht dieselbe Qualifikation wie der Schenkeldruck für den fliegenden Wechsel.

Dieser Hinweis führt zu dem Problem der Skalierung als solcher. Rankings sind tatsächlich nämlich Formen von Skalierungen. Dazu muss man folgendes wissen: Es gibt Skalen wie die Temperaturskala oder die Schulnoten, die ein definiertes Anfang und ein definiertes Ende haben. Es gibt die heißest denkbare und die kältest denkbare Temperatur und es gibt eine 1 und eine 6. So wird bei Universitätsrankings nicht gemessen, sondern es wird so getan, als ob Leistungssteigerungen nach oben keine Grenze hätten. Wenn also die Anwesenheit eines Nobelpreisträgers in einem Lehrkörper ein hohes Qualitätsmerkmal ist, dann hätte jene Universität den höchstdenkbaren Platz erworben, deren gesamter Lehrkörper über einen Nobelpreis verfügt. Wir sehen sehr schnell, welcher Denkfehler hier gemacht wurde ist: Das sogenannte empirische Relativ einer Skala, das Eigenschaften vergleichbar macht, wird verwechselt mit einem numerischen Relativ, in dem die Eigenschaften arithmetisiert werden. Unter den grundsätzlich möglichen wählbaren Skalenniveaus, den Nominal-, Ordinal- und Kardinalskalen mit den Ausformungen Intervall- und Verhältnisskalen werden häufig Ordinalskalen gewählt, in dem sie auf die Reihenfolge der Elemente ungeachtet der Frage zielen, wie groß die Abstände zwischen den Leistungselementen sind. Das ist insbesondere bei den internationalen Rankings der Fall, die genauso unsinnig funktionieren wie Olympiaden, bei denen der Leistungsabstand zwischen den Besten ein Artefakt der Zeitmessung ist.

Wenn wir also zusammenfassend fragen, ob Wissenschaftsrankings ihre Zwecke erfüllen, dann ist dieses zu verneinen, weil sie nicht objektiv, nicht reliabel und

nicht valide sind, ihre Skalierungsformen sind fehlerhaft und irreführend. Hinzu tritt, dass internationale Rankings oftmals nur mit sogenannten „Proxies“ arbeiten, also mit Stellvertreterdaten für Qualität, dass sie sich im Wesentlichen auf ein Prozent sämtlicher Universitäten beschränken und dass im Übrigen ein Verzerrungseffekt dann eintritt, wenn englischsprachige Datenbanken als Quelle herangezogen werden und nicht auch nationalsprachliche andere. Umso erstaunlicher, dass selbst bei der Benutzung derselben Datenbanken, insbesondere von Thomson Reuters, durch verschiedene Rankings trotzdem völlig unterschiedliche Ranking-ergebnisse auch hier zustande kommen. Die Liste der Einwendungen ließe sich fortsetzen.

4. Leistungsskalierung und die Folgen

Wir müssen aber noch eine andere Frage stellen, bevor wir uns damit beschäftigen, was man denn nun tun soll. Die Frage heißt, was sind die Folgen der Anfertigung und Veröffentlichung von Ratings? In diesem Zusammenhang haben Osterloh und Frey auf den sogenannten „lock-in-effect“ hingewiesen. Die von ihnen so bezeichneten perversen Lerneffekte aus der Paradoxie der Leistungsindikatoren bestehen darin, dass sich für einen Erfolg im Forschungsbereich die Vernachlässigung der Lehre als Mittel der Wahl ebenso anbietet wie die Auswahl von Forschungsthemen, für die leicht erreichbare Daten vorliegen, die Vernachlässigung von Buchpublikationen, die in Rankings nicht gezählt werden, die Unterlassung von Replikationsstudien, die sich schwer publizieren lassen (vgl. Osterloh/Frey 2012, S. 2). Gleichzeitig treten Zeitaufwendungen für den „wissenschaftlichen Diskurs“ und für einen „organisierten Skeptizismus“ zurück, ebenso wie Interdisziplinarität, eine Vielfalt von Forschungsansätzen oder der Dialog mit der Öffentlichkeit (ebd. S. 3). Diese Probleme gelten insbesondere für Forschungsrankings als Folgen für das Verhalten von Wissenschaftlern. Ergänzen lässt sich die Liste um die Neigung, vornehmlich in englischsprachigen Zeitschriften zu publizieren, also über Gegenstände, die überhaupt in der englischen Sprache darstellbar sind. Mit anderen Worten: Veröffentlichte Ratings, besonders wenn sie mit Belohnungsversprechen wie Reputationsgewinn oder zusätzliche Mittelausschüttungen - sei es im Rahmen von Ziel und Leistungsvereinbarungen für ein Individuum oder für die Arbeitsgruppe - verbunden sind, produzieren mittelfristig eine Scheinwirklichkeit und eine Fokussierung auf einen Typus von Mainstreamwissenschaft, bei dem die Sicherheit am größten ist, erfolgreich zu sein. Diese Scheinwirklichkeit korrespondiert mit einer Art Scheinmarkt, weil die Einsortierung von Individuen oder Einrichtungen, zu denen sie gehören, Marktvorteile auf dem Berufungs- und Besoldungsmarkt bringen können. Auf der Seite der Universitätsnutzer erzeugen sie Entscheidungseffekte, etwa wenn bei einer hoch „gerankten“ Universität die Zahl der Bewerber explodiert, obwohl ein Zusammenhang zwischen den angesetzten Forschungsparametern im Rahmen von Forschungsrankings und der Qualität des Unterrichts gar nicht angenommen werden kann. Auch hier entsteht ein Scheinmarkt. Bewerber und Bewerberinnen erhoffen sich ein späteres Vorteil bei der Arbeitsplatzsuche, weil sie an einer Universität tätig waren, in der

die Zahl der Professoren hoch ist, die große Mengen von Drittmitteln einwerben, egal zu welchem Forschungsgegenstand, auch in ganz fremden Fächern.

5. Zur Legitimation von Leistungsskalierung

Wenn man registriert, dass die bekannten Rankings ihre Zwecke nicht erfüllen oder sogar unerwünschte Nebenwirkungen haben, dann ist die Antwort auf die Frage, die mir gestellt worden ist, welche Anforderungen an Leistungsbewertungen zu richten sind, natürlich recht einfach: Leistungsbewertungen müssen das beschreiben, was tatsächlich der Fall ist und sie dürften keine unerwünschten Nebenfolgen haben. Es gibt eine Reihe von Bemühungen, minima moralia zu formulieren, die bei der Erstellung von Leistungsvergleichen erfüllt sein sollen. Dazu gehören die „Berlin Principals“, die 10 Thesen aus dem Zusammenhang der Friedrich Ebert Stiftung, sicher auch Überlegungen, die im Zusammenhang von Multirank angestellt worden sind und natürlich die Empfehlungen des Wissenschaftsrats. Sie sind so etwas wie die Genfer Konvention: Wenn schon Krieg geführt werden muss, dann dürfen Kriegsgefangene aber nicht zur Zwangsarbeit herangezogen werden. Wichtig sind in diesem Zusammenhang zweifellos die Empfehlungen von Osterloh und Frey, Rankings einzubetten in Rahmenbedingungen, so etwa diese, dass Rankings keine dominante Rolle spielen dürfen, das sie immer zu begleiten sind durch qualitative Expertenurteile und so weiter. Aber wer will das sicherstellen? Wer schützt uns vor der interessengeleiteten Benutzung von immer nur begrenzt objektiven, reliablen und validen Messergebnissen? Wer schützt uns vor dem falschen Schluss, dass es klug ist, einer hochgerankten Universität mehr Geld zu geben als einer anderen? Müsste es nicht umgekehrt sein und die schwächere gefördert werden, damit sie endlich leistungsfähig wird? Diese Überlegungen haben verschiedene Fächervertretungen dazu veranlasst, ihren Mitgliedseinrichtungen zu empfehlen, sich an der Datenlieferung zu Rankings nicht mehr zu beteiligen. Das ist eine pragmatische Entscheidung, sie löst aber das Problem nicht, weil es weiterhin Ratings geben wird, die dann nur noch ein bisschen schlechter sind als die jetzigen, weil die Datenbasis schwächer wird.

Wir müssen also jenseits der Qualitätsfrage von Rankings und Ratings nicht nur nach der methodischen Gewissheit, sondern doch sehr viel mehr nach der sozialen Wahrheit fragen. Kant würde gefragt haben: Sind Leistungsvergleiche sittlich erlaubt?

Die Antwort ist eine konditionale: Leistungsvergleiche sind dann gerechtfertigt, wenn durch das, was sie auslösen und bewirken, die Allgemeinheit schneller und besser an die Lösung von Problemen herangeführt würde, als wenn es solche Leistungsvergleiche nicht gäbe. Wir müssen also wissen wollen, ob der Leistungsvergleich den dahinterliegenden Wettbewerb mit Gewinnern und Verlierern dadurch rechtfertigt, dass Leistungen so gesteigert werden, dass auch die Verlierer davon profitieren. Für diese Annahme gibt es keine empirischen Evidenzen. Im Gegenteil wird in der psychologischen Kleingruppenforschung davon ausgegangen, dass Kooperation leistungsfördernder als Wettbewerb ist. Gleichwohl sind Leistungsvergleiche nach Taijfel und Turner (1986) Elemente eines

Grundbedürfnisses von Individuen, nämlich eine positive Selbsteinschätzung zu erhalten. Sie kommt dann zustande, wenn unter anderem soziale Identität bei einem Individuum entsteht. Sie entsteht durch die Mitgliedschaft in Gruppen also zum Beispiel der Zugehörigkeit zu einem Forschungsinstitut und durch die Bewertung dieser Mitgliedschaft. Diese Bewertung findet optimaler Weise durch den Vergleich der einen Gruppe mit einer anderen, also sagen wir, einem anderen Forschungsinstitut statt. Wenn eine Forschungsgruppe und auch ein Individuum dabei positiv abschneiden, wächst die positive Selbsteinschätzung und auch die Leistungsbereitschaft zum Erhalt dieser Bewertung.

Da die Zahl der Verlierer aber größer ist als der Gewinner müssen die Konsequenzen einer negativen Bewertung im Leistungsvergleich betrachtet werden. Sie bestehen, psychologisch gesehen, darin,

- dass die Individuen versuchen, die Gruppe so schnell wie möglich zu verlassen. Sie lassen sich wegberufen.
- dass Verlierer die Vergleichsdimensionen wechseln, in dem sie Einfluss auf die Leistungsindikatoren nehmen,
- dass sie die Interpretation der Attribute wechseln, indem sie hochbewertete Leistungen anderer schlecht reden,
- in dem eine andere Vergleichsgruppe herangezogen wird, die noch schlechter ist oder
- indem eine direkte aggressive Konfrontation mit der Vergleichsgruppe gesucht wird. Diese besteht etwa in gezielten Zitationskartellen, Mobbing, Berufungskartellen und ähnlichem, wie wir es aus dem Alltag kennen. Diese aggressive Form verdankt sich einer empirischen Tatsache, die in dem sogenannten „minimal-group-paradigma“ festgehalten ist: Interessanterweise verschaffen Mitglieder einer eigenen Gruppe sich selbst nämlich nicht nur den größten Vorteil, sondern sie versuchen in erster Linie den Abstand zu allen anderen Gruppen mit Hilfe der benannten Instrumente zu vergrößern. Wir sehen sehr schnell, dass dieser psychologische Mechanismus der Reaktion auf den Verliererstatus für die Qualität der gesamten Wissenschaftseinrichtung desaströs ist, wie im übrigen auch für das gesamte Wissenschaftssystem, innerhalb dessen es gleichfalls nur eine kleine Zahl von Gewinnern geben kann.

Mit anderen Worten: Leistungsvergleiche bringen eben nicht die Allgemeinheit sehr schnell und zuverlässig in den Genuss guter Problemlösungen, da diese Problemlösungen nicht von einer kleinsten Zahl bester Wissenschaftler abhängen, sondern vom Wissenschaftssystem insgesamt. Genau dieses, dass das System im Wesentlichen aus Verlierern bestehen muss, das ist der Sinn von Rankings, es enthält aber zerstörerische Einflüsse.

Demzufolge kann unsere Aufgabe, ethisch betrachtet, nicht darin bestehen, Leistungsvergleiche methodisch zu verbessern, sondern wir müssen zur Kenntnis nehmen, dass sie, egal wie gut, aus sozialpsychologischen Gründen zu den beschriebenen Effekten führen und damit dysfunktional sind. Die zerstörerischen

Wirkungen bestehen im Alltag aus Missgunst, Misstrauen, Frustration und anderen Formen der Aggression.

Insofern käme es eigentlich darauf an, Operationen, die sich auf das Leistungsgeschehen beziehen, so zu gestalten, dass sie Verlierer verhindern.

Im Bereich des dem New Public Management ja zu Grunde liegenden wirtschaftlichen Denkens ist dieses der Tendenz nach mit der Einführung der sozialen Marktwirtschaft ja versucht worden, auch wenn sie unter dem erheblichen Globalisierungsdruck heute in vormoderne Formen zurück zu gleiten droht. Im Bereich des wissenschaftlichen Geschehens gibt es allerdings kein funktionales Äquivalent, also keinen sozialen Wissenschaftsmarkt. Hier herrscht, gleichfalls unter Globalisierungsdruck, inzwischen ein nicht selten gnadenloser Wettbewerb, dessen Schattenseiten wir gerade erst ahnen, sei es der Kauf von Titeln, das Plagiiere, das Fälschen von Daten, die Ehrabschneidung und vieles andere.

Solche Aggressionsformen können wohl nur durch zwei Instrumente abgewehrt werden, zum einen durch, ich sage es ungern, verstärkten Staatsinterventionismus in dem Sinne, dass der inzwischen entstandenen Ellenbogenmentalität im Wissenschaftssystem durch den Staat selbst Grenzen gesetzt werden müssen und diese Haltung nicht noch verstärkt werden darf. Der Staat hat also nunmehr die Pflicht dafür zu sorgen, dass die Leistungsschwächeren über die Arbeitsbedingungen verfügen, die erforderlich sind, um Erkenntniseffekte für das Wohl der Allgemeinheit zu erzielen, auch wenn es nicht gleich die Higgs-Teilchen sind. Die zweite Notwendigkeit betrifft nicht den Staat, sondern sie besteht in der Wiedergewinnung von Grundregeln des akademischen Anstands, der darin besteht, Fairness im Umgang mit den Wissenschaftskolleginnen und -kollegen walten zu lassen und sich selbst in seinem Expansionsdrang zu begrenzen. Wenn Sie mich, so gesehen, nach Erfahrungen aus der Sicht einer Universitätsleitung fragen, dann würde ich sagen, dass mein Problem nicht darin besteht, die Leistungen von Forscherinnen und Forschern nicht einschätzen zu können, sondern die Leistungsträger daran zu hindern, die 95 Prozent übrigen Wissenschaftlerinnen und Wissenschaftler um ihre Arbeitsbedingungen und ihren eigenen Erfolg zu bringen. Das Wettbewerbsgeschehen hat nicht selten eine Maßlosigkeit der Gewinner hervorgebracht, die auch nicht davor zurückschreckt, den Hochschulleitungen Ahnungslosigkeit im Umgang mit den künftigen Nobelpreisträgern zu bescheinigen. Das ist die Exemplifizierung eines weisen Satzes von Ludwig Marcuse, der hieß: „Das Leid der Unterlegenen kommt aus dem Vergleich.“ Da wir uns an dem Leiden von Unterlegenen nicht beteiligen möchten, hat das Präsidium der Universität Hamburg beschlossen, Rankings bis auf weiteres nicht zu beliefern.

6. Die bessere Alternative: Auditierungen und Ratings

Für Organisationen wie den Wissenschaftsrat mag das eine Auskunft sein, welche seine künftige Orientierung nicht weitertreibt. Deswegen sei zu seinen erfolgreichen Bemühungen um ein Rating gesagt: Es ist vorstellbar, dass die Deutschen Hochschulleitungen sich auf folgende Entwicklung einlassen können:

- 6.1 Zu begrüßen sind valide Informationen und Einschätzungen zu den Leistungen der eigenen Institution, die geeignet sind, eine sachgerechte Weiterentwicklung einer Hochschule voranzutreiben, nicht aber, um negative Selektionsentscheidungen gegenüber Teilinstitutionen, Fächern oder gar Personen zu treffen. Solche sind insbesondere vor dem Hintergrund der Tatsache unangemessen, dass der Staat grundsätzlich die Pflicht hat, Hochschulen von hoher Qualität vorzuhalten, weil z.B. das Angebot eines Studiums in minderwertigen Institutionen kaum legitimiert werden kann. Für solche Bewertungen eignen sich die Dokumentationen der Deutschen Forschungsgemeinschaft (z. B. der Förderatlas), des DAAD, der Alexander von Humboldt-Stiftung ebenso wie Anlass bezogene Evaluierungen durch Peers aus dem DFG-Zusammenhang und des Wissenschaftsrates. Diese Instrumente sind international beispielhaft.
- 6.2 Für staatliches Handeln, zu dem die Steuerungsentscheidungen von Hochschulleitungen gehören, sind staatlich oder wenigstens "staatsnah" beurkundete Dokumentationen und Bewertungen die erforderliche Legitimation. Dieses ist bei den unter 6.1 genannten Einrichtungen der Fall.
- 6.3 Da Leistungsvergleiche, insbesondere in veröffentlichter Form, nicht geeignet sind, die Leistungsfähigkeit schwächerer Einrichtungen zu verbessern, sondern aus sozialpsychologischen Gründen im Gegenteil zu einer Vertiefung der Leistungsunterschiede führen, auch durch einseitige Finanzallokationsentscheidungen, werden Rankings nicht unterstützt, insbesondere dann nicht, wenn Parameter und Erhebungs- sowie Auswertungsmethoden nicht von den Einrichtungen selbst nicht definiert werden.
- 6.4 Die Ergebnisse von Ratings stehen ausschließlich Entscheidungsträgern, also insbesondere den jeweiligen Hochschulleitungen für Steuerungszwecke zur Verfügung, so dass der Leistungsstand dieser Hochschulen mit dem anonymisierten durchschnittlichen Leistungsstand der anderen Hochschulen vergleichbar wird und eine Einordnung der eigenen Hochschule im Leistungsspektrum möglich ist. Dieses Verfahren wurde erfolgreich im Rahmen des Profildatenprojektes zur Internationalisierung von HRK, DAAD und Alexander von Humboldt-Stiftung entwickelt und angewendet.
- 6.5 Dem Wissenschaftsrat ist zu empfehlen, das Konzept des Ratings unter den Bedingungen von 6.1 bis 6.4. weiter zu entwickeln und einen Auditierungszyklus anzubieten, an dem Hochschulen sich nach eigener Entscheidung beteiligen können. Die Bewertungen sind mit einrichtungsspezifischen Steuerungsempfehlungen zu versehen, wie sie in Auditierungsverfahren üblich sind. Diese können ggf. mit den von der HRK angestrebten Auditierungen als künftigen Ersatz für Akkreditierungen verbunden werden.

Literaturverzeichnis

- Borgwardt, A.: Rankings im Wissenschaftssystem. Zwischen Wunsch und Wirklichkeit, Friedrich-Ebert-Stiftung, Berlin 2011
- Deutsche Gesellschaft für Soziologie (DGS): Wissenschaftliche Evaluation ja – CHE-Ranking nein. Methodische Probleme und politische Implikationen des CHE Hochschulrankings, Stellungnahme der GDS, Juni 2012
- Glaser, H. A.: Neue Zahlenspiele. Internationale Rankings von Universitäten. In: Forschung und Lehre 19 (3), 2012, S. 210-213
- Heinemann, K.: Hochschulrankings aus Unternehmenssicht. Wie Ranking-Tabellen Unternehmen bei der Beurteilung von Hochschulen unterstützen. In: vhw Mitteilungen 39 (1), 2012, S. 15-19
- Holzmüller, M.: Bewertung von Hochschulen. „Uni-Rankings sind nie objektiv“, Interview mit Phil Baty, In: Süddeutsche Zeitung, 06.06.2011
- Kieser, A.: Akademische Rankings. Die Tonnenideologie der Forschung, In: Frankfurter Allgemeine Zeitung, 11.06.2010
- Kultusministerkonferenz (KMK): Instrumente der Qualitätsfeststellung in der Hochschulforschung. Erfahrungen der Länder, Beschluss der KMK vom 12.05.2011
- Leonard, W. P.: Gaming in the American university ranking system. In: University World News, 25.03.2012
- Lessenich, S.: Sind Rankings sinnvoll? – Nein! In: Die Zeit (29), 12.07.2012
- Osterloh, M./Frey, B. S.: Forschungsrankings und das Paradox der Leistungsmessung, Ms. 14.06.2012 o. O.
- Tajfel, H./Turner, J. C.: The social identity theory of inter-group behavior. In: S. Worchel/L. W. Austin (Hrsg.), Psychology of Intergroup Relations, Chicago 1986
- Wissenschaftsrat: Empfehlungen zu Rankings im Wissenschaftssystem (Drs. 6285-04), Hamburg, November 2004
- Wissenschaftsrat: Empfehlungen zum Forschungsrating (Drs. 8485-08), Rostock, Mai 2008
- Wissenschaftsrat: Empfehlungen zur vergleichenden Forschungsbewertung in den Geisteswissenschaften (Drs. 10039-10), Köln, Juni 2010