

Der Teufel der Unvergleichbarkeit

Über das quantitative Messen und Bewerten von Forschung

| WERNER PLUMPE | Wie lässt sich eine geistige Leistung messen? Lässt sie sich überhaupt messen? Fragen, die je nach Interessenlage unterschiedlich beantwortet werden. Die Hinweise mehrten sich indes, dass die bisherigen Forschungsrankings und -ratings nicht das halten, was sie versprochen hatten. Um welche Schwächen handelt es sich genau und welche Konsequenzen sind daraus zu ziehen?

„Eine Empfindung, die mit Worten ausgedrückt wird, ist allzeit wie Musik, die ich mit Worten beschreibe; die Ausdrücke sind der Sache nicht homogen genug.“

Georg Christoph Lichtenberg

Seit einigen Jahren ist es auch in der Wissenschaft geradezu modisch geworden, quantitative Leistungsmessungen anzustellen und Leistungsvergleiche durchzuführen. Die Mode entstammt der Unternehmensberatung, die auf diese Weise in

»Im Zweifelsfall will man auch mit der Axt messen: Was nicht passt, wird passend gemacht.«

den 1990er Jahren glaubte, die Stärken und Schwächen von Unternehmen feststellen zu können. Findige Wissenschaftspolitiker, umtriebige Hochschulleitungen und eilfertige Berater glaubten dann, diese Verfahren auch bei der Bewertung und Verbesserung universitärer Forschungs- und Lehrleistungen nützen zu können. Nur: Was in der Un-

ternehmensberatung Sinn haben mag, um die Leistungsfähigkeit von Unternehmen gezielt zu verbessern, ist im Bereich der Wissenschaft mehr als fragwürdig. Denn während die Aufwendungen und Erträge eines Unternehmens sich zu Marktpreisen bewerten und damit quantitativ abbilden lassen, hat Wissenschaft keinen Marktpreis, der ansonsten Unvergleichliches miteinander vergleichbar macht. Gäbe es einen Markt für die Wissenschaften in mehr als einem metaphorischen Sinne, wäre das anders. Aber die geistige Leistung eines Wissenschaftlers lässt sich kaum über deren Kosten erfassen. Und selbst wenn es ginge, wie ja einige Verfechter des evaluierten Zulagenwesens im Rahmen der W-Besoldung zu glauben scheinen, wäre ja überhaupt nicht klar, ob

die Kosten für die geistige Anstrengung eines Biologen mit denen eines Ägyptologen sinnvoll vergleichbar sind. Zum Glück ist auch bisher niemand auf die Idee gekommen, Buchpreise und Umsätze für einen Ausweis an Qualität zu halten, auch wenn das CHE mit der Bewertung von Büchern nach ihrem Umfang diesem Aberwitz immerhin nahegekommen ist. Da es mithin nutzbare und unbestrittene quantitative Parameter nicht gibt, ja bei einer derart heterogenen Sache wie Forschung und Lehre gar nicht geben kann, mussten, da man Ratings und Rankings unbedingt wollte, Surrogate her: Impact-



Faktoren, Peer-Reviews, Zitationsindizes, Drittmittelquoten, Notendurchschnitte, Examensfrequenzen, Publikationsintensitäten, öffentliche Sichtbarkeit etc. pp. Die Vielfalt der angebotenen Surrogate und die Vielfalt der Institutionen, die damit Leistungsmessung betreiben wollten und wollen,



AUTOR

Professor Werner Plumpe ist Vorsitzender des Deutschen Historikerverbandes.

machten die Sache zwar nicht einfacher, erlaubten aber ein hübsches Ablenkungsspiel: Um über den Sinn derartiger Surrogate nichts sagen zu müssen, redete man die Konkurrenz schlecht und bot sich als die bessere Alternative an, sodass die Diskussion seit einiger Zeit nicht mehr vorrangig um den Sinn derartiger Leistungsmessungen geführt wird, sondern darum, welches Instrument und welches Verfahren anderen vorzuziehen sei. Die Grundsatzfrage, ob das alles denn überhaupt sinnvoll ist, wurde und wird kurzerhand mit der Bemerkung abgefertigt, derartige Wissenschaftsbewertungen

renkonferenz und der Wissenschaftspolitik, nur mit derartigen Leistungsmessungen und -vergleichen zurecht kommen zu können, um eine angeblich kritische Öffentlichkeit über die Bedeutung der Wissenschaft aufklären und über die Verwendung der Steuermittel in diesem Bereich Rechenschaft ablegen zu können. Dabei weiß diese längst, dass derartige Versuche nichts taugen, sondern selber nur zeitintensiv und teuer sind. Auch die Vorstellung, nur auf der Basis derartiger Leistungsbemessungen und -vergleiche ließe sich das Wissenschaftssystem und die Universität sinnvoll steuern, vermag nicht zu überzeugen. Denn bisher ging es ohne Rankings und Ratings und die Kommunikation zwischen Wissenschaft und Öffentlichkeit ist produktiv. Die künstlich erzeugten „Tabellen“ und Übersichten nimmt folgerichtig niemand so richtig ernst, auch wenn sich Rektorate und Präsidien über gute Ranking-Plätze ihrer Einrichtungen freuen. Und dass auf der Basis etwa der CHE-Tabellen und Graphiken irgendwelche grundlegend neuen Entscheidungsmöglichkeiten gewonnen seien, die vorher gefehlt hätten, behauptet nicht einmal mehr das CHE selbst, das sich neuerdings darauf zurückzieht, seine Übersichten könnten doch immerhin sog. bildungsfernen Schichten bei der Universitätsauswahl helfen.

Eine klare Tendenz

Ein kurzer Blick in die Realität zeigt eine klare Tendenz. Die Hinweise häufen sich, dass die

bisherigen Rankings und Ratings nichts von dem halten, was sie versprochen hatten. Die neuen Macher des Times-Higher-Education-Rankings geben offen zu, dass die bisherigen Ranglisten das Papier nicht wert waren, auf dem sie gedruckt wurden. Dem CHE laufen die Wissenschaftler weg, und man be-

seien ohnehin nicht aufzuhalten; da komme es darauf an, deren Qualität zu verbessern. Und in der Tat scheint hier und dort der Wille vorhanden zu sein, im Zweifelsfall auch „mit der Axt“ zu messen: Was nicht passt, wird passend gemacht. Noch immer glauben Vertreter der Hochschulleitungen, der Rekto-



Foto: mauritius-images

KISSWIN.DE

HOW TO BECOME A PROFESSOR?

DAS HTBP-SEMINAR

22.-24.
Oktober
TEIL I

19.-21.
November
TEIL II

KARRIEREZIEL PROFESSUR

Wissenschaftliche Karrieren sind kein Zufall – sie sind strategisch planbar und individuell gestaltbar!

Das zweiteilige Kompakt-Seminar »HOW TO BECOME A PROFESSOR?« (HTBP-Seminar) bietet hochqualifizierten Nachwuchswissenschaftler/-innen, die mit ihrer Bewerbung überzeugt haben, eine intensive Karriereunterstützung auf dem Weg zur Professur.

TEIL I
STRATEGIE UND PLANUNG
22.-24. OKTOBER 2010

TEIL II
BEWERBUNG UND BERUFUNG
19.-21. NOVEMBER 2010

Pilot-Seminar
ERSTES MAL
PREIS



Anmeldung und Informationen unter KISSWIN.DE! Kontakt: RICK@KISSWIN.DE
Bewerbungsschluss: 31. August 2010

GEFÖRDERT VOM

hilft sich mit Ersatzdaten, die die Hochschulleitungen liefern. Das Rating des Wissenschaftsrates ist trotz des Bemühens um fachspezifische Kriterien bei sehr vielen Historikern auf Skepsis gestoßen; vor allem aber sind die Ratings in den Fächern Chemie und Soziologie weitgehend wirkungslos verpufft; kaum jemand spricht mehr davon, und auch auf Nachfrage gibt es keine klaren Aussagen zu Wirkungen und Folgen des Ratings und seiner Befunde. Jüngst hat der Mannheimer Betriebswirt Alfred Kieser die in den Wirtschaftswissenschaften üblichen und dort sehr verbreiteten Rankings, Zitationsindices und Peer-Reviews in seiner Abschiedsvorlesung einer ätzenden Kritik unterzogen. Die verwendeten Verfahren seien nicht nur sinnlos, sondern überdies gefährlich, weil sie jüngere Wissenschaftler in einer geradezu unakademischen Weise zu strategischem Publikationsverhalten zwingen und überdies innovationsfeindlich seien. In der Tat zeigen die bisher genutzten Verfahren zur Leistungsmessung derartig grundlegende Schwächen, dass es ausgeschlossen ist, diese Schwächen ließen sich durch Verfahrensverbesserungen beheben.

Zwei zentrale Probleme

Ein Blick auf zwei zentrale Probleme der Forschungsbewertung (Homogenität der Forschung, performativer Widerspruch zwischen Aktualität und Substantialität) macht diesen Einwand unmittelbar plausibel. Zunächst produ-

»Es ist an der Zeit, endlich auf diejenigen zu hören, die an den Universitäten die eigentliche Arbeit leisten.«

ziert – wie gesagt – Forschung keine Güter, die zu ihren Marktpreisen erfasst und über marktpreisbewertete input-output-Schemata kalkuliert werden können. Die Homogenität der Forschung als Grundlage ihrer Vergleichbarkeit muss erst künstlich (im Zweifelsfall eben „mit der Axt“) erzeugt werden, aber selbst daraus ergibt sich keine vergleichende Aussage etwa über Arbeiten zur frühmittelalterlichen Kosmologie und zur Rohstoffversorgung der chemischen Industrie im Zweiten Weltkrieg, um zwei Beispiele aus der historischen Forschung heranzuziehen.

Was sagt ein quantitativer Vergleich zwischen derartigen Forschungsrichtungen überhaupt aus? Der Wissenschaftsrat hat diese Schwierigkeiten gesehen und sich daher ergänzend zu quantitativen Verfahren für eine intensive Textlektüre durch eine Gruppe von Peers entschieden, die aber auch wie-

»Forschung produziert keine Güter, die über quantitative Schemata kalkuliert werden können.«

der in eine Art Notenschema einmünden. Hier sollen also qualitative Urteile über die Forschung einer Disziplin ebenfalls herangezogen werden, doch verschiebt dieser an sich begrüßenswerte Ansatz nur das Problem. Denn der „Teufel der Unvergleichbarkeit“ kehrt als Problem der Gutachterausswahl zurück, und selbst wenn das noch gelöst werden könnte, täte sich eine ganz neue Schwierigkeit auf. Rankings und Ratings leben angesichts des andauernden Kommen und Gehens an den Universitäten davon, dass sie in regelmäßigen, möglichst kurzen Abständen wiederholt werden. Denn schon nach einem Jahr sind die Daten u.U. veraltet. Damit gerät das anspruchsvolle Rating des Wissenschaftsrates aber in eine Art performativen Widerspruch von Aktualität und Substantialität. Je aktueller das Rating, um so weniger Zeit bleibt für die Lektüre und um so

substanzieller sind die Aussagen. Je substanzhaltiger hingegen das Rating ist, um so größer ist der Aufwand, der dafür betrieben werden muss, sodass es

zwangsläufig nicht wirklich aktuell sein kann. Beides zugleich, Aktualität und Substantialität, sind daher bestenfalls in ganz wenigen und nur sehr kurz anhaltenden Ausnahmesituationen zu haben, oder man muss dem gegenwärtigen System der permanenten, dezentralen Leistungsbewertung durch die Wissenschaftler an den Hochschulen vertrauen, die in Zeitschriften und Buchbesprechungen, auf Konferenzen und Kolloquien die Leistungen ihrer Kollegen laufend einer intensiven Bewertung unterziehen. Das aber wollen viele Vertreter in der Wissenschaftspolitik, der

Hochschulbürokratie und den Bewertungsagenturen nicht; statt dessen säen sie Misstrauen gegen die bestehenden Praktiken.

Unerklärliches Misstrauen

Dieses im Grunde unerklärliche Misstrauen gegenüber den Bewertungspraktiken an den Universitäten und in den akademischen Disziplinen, das sich bestenfalls auf anekdotische Evidenz berufen kann, ist um so weniger verständ-

lich, wenn man sieht, mit welcher Naivität Rankings und Ratings schöngeredet werden. Die Realität sieht ganz anders aus: Wo man auch hinschaut und hinhört, nur selten wird über positive Effekte von Ratings und Rankings berichtet, und dann zumeist von interessierter Seite. An den Universitäten werden unter den forschenden und lehrenden Kollegen, von denen ohnehin niemand auf die Idee gekommen wäre, derartigen Bewertungen das Wort zu reden, deren Schwächen und nachteilige Folgen fast durchweg beklagt. Warum, so fragt man sich, wird dennoch weiter daran festgehalten? Wissenschaftspolitik, Rektorenkonferenz und Hochschulleitungen haben sich, so scheint es, in eine Sackgasse manövriert, aus der sie endlich wieder herauskommen sollten. Es ist an der Zeit, endlich auf diejenigen zu hören, die an den Universitäten die eigentliche Arbeit leisten, das Misstrauen zu überwinden und nicht länger einer Mode hinterherzulaufen, die im übrigen nicht einmal in ihrem Kerngebiet, der Bonitätsbewertung von Schuldnern und der Unternehmensbewertung, unstrittig ist. Ein Verzicht hätte nur Vorteile, es könnte endlich wieder die permanent laufende, dezentrale fachliche Forschungsbewertung in den Vordergrund treten, die es zum Glück weiterhin gibt.

Gekürzte Fassung eines Vortrages, der bei einer Tagung der Konrad Adenauer-Stiftung am 19. Juni in Cadenabbia gehalten wurde. Eine Langfassung kann bei der Redaktion von Forschung & Lehre angefordert werden.