

Die kleinste publizierbare Einheit

Oder: was ist messbare wissenschaftliche Qualität?



Rankings, Ratings, Hitlisten, Impactfaktor, Hirsch-Index, klassische Publikationsliste: das sind nur einige der Schlagworte, die in der gegenwärtigen Debatte um die „Evaluierung“ von Hochschulen, Fakultäten oder einzelnen Wissenschaftlern kursieren. Immer stärker geraten die Hochschulen unter Legitimationsdruck. Die „Leistungsorientierte Mittelvergabe“ soll Anreize schaffen, und es soll gerecht zugehen bei der vergleichenden Beurteilung. Der Chemiker Thomas Koop, der Literaturwissenschaftler Martin von Koppenfels, der Psychologe Thomas Mussweiler und der als externer Experte eingeladene Medizynpsychologe Elmar Brähler diskutieren: Wie vermessbar ist die Wissenschaft?

Warum sind Evaluationen von Hochschulen, Instituten oder Fachbereichen und das Messen wissenschaftlicher Leistungen „en vogue“?

Koop: In Zeiten von Ressourcenknappheit gelten Evaluationen als geeignetes Steuerinstrument, um Geld zielgerecht einzusetzen. Geld soll für Qualität ausgegeben werden.

von Koppenfels: Diese Debatte hat in Deutschland im Vergleich zu anderen Ländern spät begonnen, und zwar mit der Evaluierung der ostdeutschen Hochschulen nach der Wende. Außerdem zeigt sich ein Trend zu wachsender Diversifizierung und Autonomisierung der Hochschulen. Der Ruf nach verlässlichen Standards wird gleichzeitig immer lauter. Diese Tendenz wird sich noch verstärken, sollte der Hochschulbereich wirklich vollständig in die Hand der Länder übergehen.

Welchen Einfluss haben Evaluationen auf das wissenschaftliche Arbeiten?

von Koppenfels: Von Fach zu Fach existieren enorme Unterschiede: In einigen Fächern ist es einfacher, Qualität zu quantifizieren, als in anderen. Evaluationen erzeugen Anpassung, im schlimmsten Fall eine Verlagerung von Forschungsaktivitäten in Bereiche, die leichter quantifizierbar sind. Ein eher amüsantes Beispiel einer solchen Anpassungsleistung ist das Phänomen der „kleinsten publizierbaren Ein-

heit“ – die Zerlegung eines Buchmanuskripts in möglichst viele kleine Aufsätze, deren separate Publikation mehr Punkte bringt. Das verändert den Charakter des wissenschaftlichen Produktes erheblich.

Brähler: Die Wissenschaftler richten sich in ihrer Arbeit nach dem angewandten System der Quantifizierung. Zum Beispiel werden in der Medizin die Klinikchefs bei Publikationen an letzter Stelle genannt. Wird aber nach Zitierungen bewertet, sind die zuerst genannten Autoren im Vorteil – also stehen die Chiefs jetzt wieder vorn. Bei hochkarätigen Aufsätzen ist es in der Medizin keine Seltenheit, dass 40 Autoren beteiligt sind. Volle Punktzahl machen aber nur der erst- und der letztgenannte Autor, die mittleren Autoren erhalten nur reduzierte Anteile.

Koop: In Australien wurde seit Mitte der 90er Jahre für jede wissenschaftliche Publikation, unabhängig von deren Qualität, eine bestimmte Summe ausgeschüttet. Seitdem erhöhte sich zwar der Anteil australischer Artikel im SCI, allerdings vornehmlich derjenigen in weniger renommierten Zeitschriften. Die Grundfrage lautet also bei jeder Bewertung: Was will man fördern?

Kann eine Qualitätsmessung jedem Fach gerecht werden?

von Koppenfels: Es scheint in der gesamten Evaluationsdebatte mittlerweile Konsens zu sein, dass man stärker differenzieren muss. Bestimmte Fächer kommen unter die Räder, wenn sie nach Standards bewertet werden, die für andere Bereiche gelten. Für die Geisteswissenschaften ist es zum Beispiel nur ein relativer Fortschritt, mit den Sozialwissenschaften über einen Kamm geschoren zu werden. Andererseits können Bewertungen nicht unendlich verfeinert werden, weil es auch darum geht, Standardisierungen zu finden.

Brähler: Etwas Bewertungs-Gerechtigkeit kann man erreichen, wenn ausschließlich innerhalb eines Faches verglichen wird. In Nord-

Immer häufiger werden bei der Besetzung von Professuren, bei leistungsbezogener Mittelzuweisung oder bei der Zulassung zur Habilitation **bibliometrische Methoden** angewendet. Im vom **Institute for Scientific Information (ISI)** herausgegebenen **Science Citation Index (SCI)** wird aufgelistet, welcher Autor welchen anderen Autor zitiert.

Der Quotient aus den Zitaten, die eine Zeitschrift innerhalb eines Jahres auf publizierte Aufsätze der letzten zwei Jahre erhalten hat, und der Zahl der Aufsätze, die in dieser Zeitschrift im gleichen Zeitraum veröffentlicht wurden, ist der **Impact Faktor (IF)**. Der IF ist ein Indikator für die Reputation einer Zeitschrift, dient aber immer häufiger in bestimmten Fächern zur Begutachtung von Wissenschaftlern.

rhein-Westfalen werden beispielsweise einzelne Fachgebiete in der Medizin bundesweit betrachtet.

Mussweiler: Dieser Vergleich funktioniert nur, wenn innerhalb eines Faches die Standards klar sind, und das ist bei weitem nicht immer so.

Koop: Wichtig ist, dass niemand aus der Verwaltung ohne Fachkenntnisse eine Beurteilung nach Zahlen und Quotienten vornehmen kann. Außerdem ist das Zählen von Publikationen und Zitaten extrem fehleranfällig. Es gibt einen Wissenschaftler, der denselben Nachnamen und Initialen wie ich hat, deshalb können mir auf den ersten Blick über 100 Publikationen zugeschrieben werden.

Der Impactfaktor einer Zeitschrift eignet sich ebenfalls überhaupt nicht dazu, die Leistung eines Artikels oder gar des Autors zu messen. 15 Prozent der Beiträge machen häufig 50 Prozent der Zitate einer Zeitschrift aus, das zeigt schon den unmöglichen Rückschluss auf die Qualität einzelner Beiträge. Mir gefällt der Hirsch-Index daher viel besser. Dort wird gefragt: Was wurde publiziert und wie oft zitiert? Wie viel Wirkung hatten die publizierten Beiträge? Allerdings benachteiligt der Hirsch-Index junge Wissenschaftler.

Was bedeuten quantifizierende Messverfahren für solche Fächer, die nicht zu den so genannten Zukunftsfächern gehören?

von Koppenfels: Die Quantifizierung soll ein Ausweg aus den traditionellen Peer-Review-Methoden sein. Ich frage mich allerdings, ob man sich nicht auf einer mathematisch-formalisierten Ebene die gleichen Probleme eingehandelt hat.

Mussweiler: Der Vorteil der quantifizierenden Methoden liegt doch darin, dass nicht mehr zwei oder drei Gutachter über das Wohl und Wehe einer ganzen Disziplin entscheiden.

Alle, die wissenschaftlich in diesem Feld arbeiten, können mit abstimmen. Das schafft eine größere soziale Kontrolle und gewisse Spielregeln.

von Koppenfels: Der Impactfaktor ist auch ein Wichtigkeitsindex. Dabei wird genau wie früher institutionelle Macht ausgespielt: Wen muss ich zitieren?, das bleibt nach wie vor eine wichtige Frage.

Wie müssen Evaluationen generell beschaffen sein, um Reformprozesse anzustoßen?

von Koppenfels: Die Geisteswissenschaften haben das große Dilemma, dass ihre Arbeit wenig kostet. Bei der Mittelzuweisung werden jedoch die Drittmittelausgaben berücksichtigt. Eine solche Politik bedeutet eine Umverteilung von den „preiswerten“ Wissenschaften zu den „teuren“.

Brähler: Auch in der Medizin gibt es so etwas, beispielsweise in der Geschichte der Medizin oder anderen Randfächern wie medizinischer Soziologie. Dort werden an einigen medizinischen Fakultäten hervorragend Drittmittel eingeworben, und die Wissenschaftler publizieren sehr viel. Dennoch wird ihnen auch dann oft wenig Geld zugestanden, weil die Arbeit „eigentlich nichts kostet“. Dann wird das gesamte System ad absurdum geführt: Eine Disziplin strengt sich an, alle Indikatoren zu erfüllen, und erhält trotzdem nicht die Mittel.

Wie lässt sich verhindern, dass Evaluationen zum politischen Kontrollinstrument werden?

Koop: Eine Hochschule kann Schwerpunkte setzen und beispielsweise entscheiden, dass Drittmittel und Patente im Mittelpunkt stehen. Fällt die Entscheidung nur für Anwendungsforschung, mag das für die nächsten Jahre nützen, aber nicht mittel- und langfristig. Über Evaluationen kann man in solche Prozesse eingreifen.

Fotos: Cynthia Rühmekorf



Fortsetzung auf Seite 6



von Koppenfels: In der Praxis besteht dieser Spielraum nicht. Welche Hochschule setzt in der heutigen Zeit auf Geisteswissenschaften? Solange die Geldströme nach den erwähnten Kriterien gelenkt werden, existiert diese Freiheit der Schwerpunktsetzung nicht. Es scheint sich das Prinzip „Wer hat, dem wird gegeben“ durchzusetzen.

Welche Rolle spielt die Bewertung der Lehre?

Brähler: Die Evaluation der Lehre muss vielschichtig sein: Wie bewerten Studierende und Absolventen den Unterricht? Wie wurden sie auf die Berufswelt vorbereitet? Arbeitgeber sollen nach den Qualifizierungsmängeln der Bewerber befragt werden und Dozenten nach den Schwachstellen in der Studentenschaft, wie beispielsweise mangelnde Homogenität am Studienbeginn. Wünschenswert wäre eine auswärtige Kommission, die mit allen Protagonisten spricht.

Mussweiler: Zurzeit wird die Lehre meistens anhand von standardisierten Befragungen evaluiert, die nicht von Dozenten entworfen wurden und den Vergleich zwischen vielen Lehrenden ermöglichen.

Brähler: Wie soll das Geld über die Lehr-evaluation verteilt werden?

von Koppenfels: Wenn es um die Verteilung von Geld geht, ist die Wirklichkeit viel schematischer. In Berlin wird die Qualität der Lehre bei der Verteilung der Mittel mit 50 Prozent berücksichtigt, also höher als die Qualität der Forschung. Aber wie wird die Qualität der Lehre bestimmt? Zur Hälfte nach

der Erfolgsquote. Es geht um den so genannten „Durchsatz“: Wie viele Studierende schaffen in welcher Zeit den Abschluss?

Auch hier zeigt sich: Die Studiendauer differenziert von Fach zu Fach sehr stark, die Fächer haben nicht nur verschiedene Forschungsbegriffe, sondern auch unterschiedliche Bildungsbegriffe. Es mag Fächer geben, in denen die Studiengeschwindigkeit ein Index für Qualität ist, weil ein klar umrissener Wissenskanon existiert. In anderen Fächern gibt es eine vernetzte Wissenswelt, in die die Studierenden sich einarbeiten müssen. Dort sind die besten Studierenden die, die etwas länger brauchen. Hier ziehen ebenfalls die geisteswissenschaftlichen Fächer den Kürzeren.

Brähler: Auch die Erfolgsquote schafft Veränderungen: Soll ein großer Anteil der Studierenden den Abschluss schaffen, oder soll vorher ausgesiebt werden, um eine bestimmte Qualität zu halten? Denn ein schneller Durchsatz lässt die Noten sinken, und das hat Auswirkung auf Beurteilungen in Rankings. Je nach den Vorgaben der Politik – entweder Regelstudienzeit oder Notendurchschnitt – ändert sich das Lehrverhalten.

Sind Preise und Akademiemitgliedschaften Bewertungsparameter für einen Wissenschaftler?

Mussweiler: Im Bewerbungsverfahren ist alles, was hilft, in die engere Wahl zu kommen, sehr willkommen. Für uns Nachwuchswissenschaftler sind solche Unterscheidungsmerkmale genauso wie die Beurteilung unserer Lehre wichtige Bausteine für die wissenschaftliche Karriere.



Wie sollen künftig sinnvolle Evaluationen aussehen?

Mussweiler: Nachwuchswissenschaftler werden häufig Opfer von undurchschaubaren Parametern, beispielsweise in Berufungsverfahren. Quantifizierbare Faktoren scheinen mir ein Ausweg zu sein. Diese Verfahren sind zwar auch fehlerbelastet, aber trotzdem fairer. Die Karriere wird dadurch berechenbarer. Jeder Einzelne kann verfolgen, wie viel er publiziert hat, wie viele Mittel er eingeworben hat und wie oft er zitiert wird. Dadurch fällt die Entscheidung für oder gegen eine wissenschaftliche Karriere leichter.

Brähler: Für Mittelzuweisungen könnten noch andere Kriterien herangezogen werden, die ein Wissenschaftler erfüllen muss: Die Veranstaltung von Tagungen, Gutachtertätigkeit für Zeitschriften, Kollegiatenzeit bei der Deutschen Forschungsgemeinschaft. Ein Preis ist wichtig für die persönliche Karriere. Aber sollte er in die Leistungsbewertung einfließen?

Wie lässt sich die Forschungsarbeit eines einzelnen Forschers bewerten? Welche Rolle spielt Teamarbeit?

von Koppenfels: Die drittmittelbasierte Vergabe von Mitteln zwingt Geisteswissenschaftler mitunter dazu, reine Zweckbündnisse zu schließen, bei denen Schein-Teams entstehen. Die Chance, Drittmittel zu erhalten, ist bei Gruppenanträgen am größten, und diese Gelder gehen wiederum in die Leistungsbilanz ein. Das zeigt deutlich, wie Bewertungsstandards die Forschungsrealität verändern. Geisteswissenschaftliches Arbeiten funktioniert aber nicht immer nach einem bestimmten Arbeitsplan, der systematisch abgearbeitet werden kann. Und geisteswissenschaftliches Publizieren ist immer noch individuelles Publizieren.

Koop: Einige naturwissenschaftliche Zeitschriften sind bestrebt, am Ende der Artikel den Arbeitsanteil eines jeden Beteiligten zu nennen. Natürlich gibt es immer noch die Möglichkeit, dass ein Vorgesetzter behaupten kann, die eigentliche Grundidee stamme von ihm, aber grundsätzlich schafft dieser Weg trotzdem mehr Transparenz.

Brähler: Ich bin trotz aller Kritik ein großer Anhänger von Quantifizierung, weil sie berechenbar ist und die Spielregeln klar sind. Das gilt immer nur innerhalb eines Fachgebietes; skeptisch bin ich bei der Bewertung einer gesamten Hochschule. Die Hoffnung, dass diese Faktoren zu mehr Gerechtigkeit bei Berufungen führen, teile ich nicht, weil die Stellenausschreibungen nicht allgemein unter dem Motto „Wir suchen den Besten“ verfasst werden, sondern sich nach lokalen Schwerpunkten richten. Bei Berufungen spielt nicht ein hoher Impactfaktor die entscheidende Rolle, sondern die Passfähigkeit des Kandidaten.

Koop: Die Spielregeln und Ziele müssen noch vor der Evaluierung klar sein. Für quantifizierbare Evaluationen sind Bewerber mit Spezialkenntnissen des jeweiligen Faches notwendig. Ein einziger Index kann kein umfassendes Bild geben, vielmehr müssen verschiedene Methoden gemischt werden.

von Koppenfels: Für die Literaturwissenschaft ist der bisherige Umgang mit quantifizierbaren Methoden oft desaströs. Er ersetzt das ältere System der Rezensionen und Gutachten nicht und schafft zugleich neue Zwänge, ohne Transparenz zu bringen. Quantifizierung ist jedoch im Hochschulkontext mittlerweile unerlässlich, deshalb kann es für uns nur darum gehen, die rein quantitative Erfassung von Publikationen, wenn sie denn sein muss, möglichst fein differenziert durchzuführen.

2005 schlug der amerikanische Physiker Jorge Hirsch einen neuen Index für die Beurteilung der wissenschaftlichen Leistung vor. Der **Hirsch-Index** gibt die Zahl der Arbeiten eines Autors an, die mindestens „h-mal“ zitiert wurden. Ein Hirsch-Index von 30 bedeutet demnach, dass der Wissenschaftler 30 Publikationen nachweisen kann, die jeweils mindestens 30 Mal zitiert wurden.

□ Das Gespräch führten:
Isabell Lisberg-Haag und Uschi Heidel