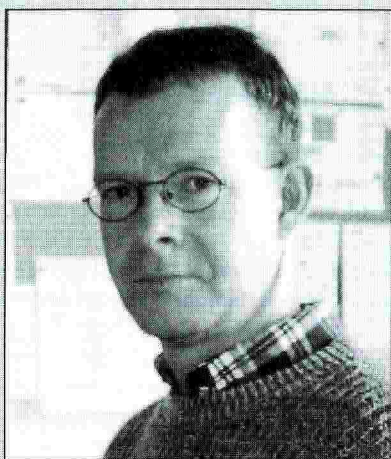




Wie auch immer das Besoldungssystem der Zukunft gestaltet wird, es wird leistungsorientiert sein. Eine leistungsorientierte Besoldung erfordert die Messbarkeit der Leistung. Der Autor zeigt die Schwierigkeiten, die Leistungsmessung zu operationalisieren, und kommt zu dem Schluss, dass eine Vielzahl von hochanspruchsvollen, komplizierten und aufwändigen Bewertungsprogrammen individuell entwickelt und ständig neu evaluiert werden muss.

Dr. phil. Peer Pasternack
Hochschulforscher am
HoF Wittenberg – Institut für
Hochschulforschung
(Universität Halle-Wittenberg)
Lehrbeauftragter für
Politikwissenschaft an der
Universität Leipzig
pasternack@hof.uni-halle.de

Eine Hand evaluiert die andere?

Was eine besoldungsrelevante Qualitätsbewertung zu beachten hätte

Leistungsorientierte Besoldung gehört – neben der Juniorprofessur, einer Relativierung der Habilitation und neuer Professorenbesoldungsgruppen – zu den wichtigsten Schlagworten, die sich mit der Hochschuldiensrechtsreform verbinden. Variable Vergütungsbestandteile sollen für individuelle Leistungen, aus Anlass von Berufungs- und Bleibeverhandlungen sowie für die Übernahme zusätzlicher Funktionen gezahlt werden. Mit anderen Worten: Belohnt werden sollen die Besseren und die Fleißigeren. Beides fällt nicht zwingend zusammen. Für die Fleißigeren werden Belastungs- und Funktionszulagen vorgeschlagen. Das erscheint weitgehend unproblematisch. Schwieriger ist es indessen mit den Leistungszulagen, denn hier kommt das Problem der Qualitätsbewertung hinein. Dazu hatte bereits die vom Bundesbildungsministerium eingesetzte Expertenkommission gelassen den Ratschlag ausgesprochen, dass „qualitativen Gesichtspunkten ausreichend Rechnung getragen werden“ müsse. Genauer sei in den Hochschulen und Fachbereichen festzulegen.

Früher oder später wird also an allen Hochschulen die Aufgabe stehen, Kriterien für solche besoldungsrelevanten Bewertungen unter Einbeziehung qualitativer Aspekte zu entwickeln. Wie aber lassen sich Qualitätsbewertungen operationalisieren?

Es mag in einer Hinsicht sinnvoll sein, dies tatsächlich von Hochschule zu Hochschule jeweils eigenständig festzulegen: Spezifische Einrichtungprofile können zu unterschiedlichen Gewichtungen führen, und diese lassen sich selbstredend nicht durch zentral vorgegebene Kriterienkataloge abbilden. Deshalb wird auch im Folgenden kein Katalog solcher Kriterien vorgeschlagen. Vielmehr sollen die Grundanforderungen und -unterscheidungen formuliert werden, die zu beachten und zu treffen sind, bevor man sich an einer beliebigen Hochschule daran setzt, die Besoldung leistungsorientiert zu gestalten.

Fünf methodische Anforderungen

Fünf Anforderungen hätte die methodische Gestaltung von Qualitätsbewertun-

gen mindestens zu erfüllen:

1. Als Faustregel sollte gelten, dass *ein Qualitätsbewertungsverfahren mindestens das methodische Niveau erreichen muss, das in dem zur Evaluierung anstehenden Bereich selbst üblich ist*. Um dies zu erreichen, sind alle quantitativen Kriterien qualitativ anzureichern. Das heißt, sobald Qualität abzubilden ist, müssen quantitative Kennziffern grundsätzlich in Indikatoren zur Feststellung von Qualitäten transformiert werden. Publikations- und Zitationsindizes etwa sind mit Gewichtungsfaktoren zu versehen, die den Charakter der betreffenden Veröffentlichung spezifizieren und die Bedeutung der jeweiligen Zeitschrift spiegeln. In Fachbereichs- oder Hochschulbewertungen sind Daten wie Lehrkräfte-Studierenden-Relation durch aussagekräftigere Indikatoren wie Lehrkräfte-Studierenden-Interaktion zu ersetzen.
2. Als allgemeiner methodischer Grundsatz sollte gelten: *Qualität als hochkomplexes Phänomen kann nicht unterkomplex abgebildet werden*, auch wenn Praktikabilitätsanforderungen Vereinfachungen erzwingen. Als Minimalvariante ist deshalb Triangulation anzuwenden: Jeder Bewertungsgegenstand wird durch mindestens zwei methodisch grundverschiedene Verfahren bewertet.
3. Ein anspruchsvoller Punkt ist die Einbeziehung von *Normabweichungen*. Die besoldungsrelevante Bewertung muss sowohl konventionelle Dauerleistungen erfassen wie auch Außergewöhnlichkeiten, d.h. langweilig sich einstellende Ergebnisse und Abweichungen von standardisierten Leistungsparametern. Denn Hochschulen existieren unter anderem, um Normabweichungen zu erzeugen: In der Forschung sollen sie das bisher noch nicht Entdeckte entdecken und das bisher noch nicht Gedachte denken. In der Lehre sind sie aufgefordert, dem Stand der Forschung entsprechendes, also in seiner Gewissheit fragiles Wissen zu vermitteln und zugleich die Fähigkeit, dieses Wissen selbstständig zu bewerten, zu hinterfragen und die Folgen seiner Anwendung zu beurteilen. Ebenso wenig sol-

len die Hochschulen ihre Studierenden auf irgendein normiertes Persönlichkeitsbild hin zurichten. Ein Weg für Leistungsbewertungen im Bereich der Forschung könnte z.B. in einer Zweistufigkeit des Verfahrens bestehen: Eine erste Stufe wird zweijährlich durchgeführt und dient der Bewertung offensichtlicher und kurzfristig wirksam werdender Qualitätsmerkmale. Eine zweite Stufe findet alle zehn Jahre statt und bemüht sich, langfristige Qualitätswirkungen der Arbeit des jeweils letzten Jahrzehnts (ggf. auch der davor liegenden Zeit) zu identifizieren. Wer bspw. infolge der zweijährlichen Qualitätsbewertungen regelmäßig auf Leistungszuschläge verzichten muss, hätte dann immer noch die Chance, eine Nachzahlung für die letzten zehn Jahre zu bekommen, sobald sich der bahnbrechende Charakter seiner Forschungstätigkeit mit Verzögerungseffekt offenbart. Ersatzweise könnte, wer mit solcher Verspätung als bedeutender Leistungsträger der Hochschule bewertet wurde, auch Anspruch auf Lehrdeputatsminderung oder zusätzliche Forschungsfreistellung haben – was die Sache wohl besoldungstechnisch vereinfacht. In diese Richtung könnten Versuche gehen, auch die Normabweichung als wesentliches Leistungsmerkmal von Hochschulen in Leistungsbewertungsprozesse einzubeziehen.

4. Die Akzeptanz des Bewertungsinstrumentariums ist zu sichern. Nur so kann Versuchen gegengesteuert werden, das Instrumentarium durch informelle Obstruktion zu unterlaufen. Daher müssen die Instrumente dem subjektiven Gerechtigkeitsempfinden entsprechen bzw. vom Grundsatz her diesem entsprechen können (ohne dass man es immer allen wird Recht machen können). Hilfreich hierfür ist der Einbau von automatisierten Rückkopplungen in die Qualitätsbewertungsprozesse. So könnte etwa die Möglichkeit geschaffen werden, anonymisierte Kommentare an einer benannten Stelle (z.B. in einem Chatroom der Hochschul- oder Fachbereichshomepage) zu hinterlegen, und ein Gremium beauftragt werden, auf der Grundlage dieser Kommentare in zuvor bestimmten Abständen eine transparente Überprüfung des Bewertungsinstrumentariums in Gang zu setzen. Bei all dem dürfte aber auch der gelegentliche Hinweis, dass es das perfekte Instrumentarium nicht geben kann, hilfreich sein: Er würde einerseits atmosphärisch entlastend wirken, andererseits plausibilisieren, warum das angewandte Instrumentarium einer regelmäßigen Überprüfung unterliegt.

5. Schließlich sind Lösungen für rationale Erfassungen der für Bewertungen benötigten Daten und für deren Auswertungen zu finden. Wo dies nicht gelingt, ist auf Dauer keine Akzeptanz für die Bewertungen zu erzeugen. Das heißt, dass jedes Nachdenken über Qualitätsbewertungskriterien sich immer zugleich auch den praktischen Fragen der Umsetzung widmen muss: Rechentechnische Erfassung?, Eingabemaske?, Eingabe durch die zu Bewertenden selbst? Wenn ja: wie viel Aufwand darf die Eingabe verursachen, ohne zu enervieren? Automatisierte Auswertung auf der Grundlage welchen Programms? Wie wird mit den Ergebnissen umgegangen, und was fordert dabei der Datenschutz?

Zentrale Unterscheidungen

Damit Operationalisierungswege gefunden werden, die den genannten methodischen Anforderungen entsprechen, sind am Beginn der Entwicklung von Bewertungsprogrammen einige wichtige Unterscheidungen zu treffen:

- Soll vorrangig die Bewertung von Leistungsgütern stattfinden oder die von Fleiß? In den Diskussionen um leistungsorientierte Besoldungsbestandteile entspricht dem die Unterscheidung zwischen Leistungszulagen einerseits sowie Belastungs- und Funktionszulagen andererseits.
- Soll die Beherrschung der Rollenkomplexität bewertet werden oder Leistungssubstitution zulässig sein? Dahinter steht, dass die Hochschullehrer-Rolle sich durch erhebliche Komplexität auszeichnet. Professoren und Professorinnen sollen vieles zugleich sein: Gut in der Forschung wie in der Lehre, begeistert in der Selbstverwaltung, erfolgreich im Netzwerkmanagement und bei der Drittmittelwerbung, hinreichend fittenreich gegenüber der Verwaltung, gelassen und kompetent in partnerschaftlicher Mitarbeiterführung, dazu souveräne Instrumentalisten auf allen neuen Medien, kognitive Innovateure wie auch unablässige Erzeuger öffentlicher Resonanz und nimmermüde Übersetzer wissenschaftlicher Fragestellungen auf gesellschaftliche Relevanzbedürfnisse hin. Zu unterscheiden ist nun: Soll die besoldungsrelevante Bewertung darauf zielen festzustellen, inwieweit die und der Einzelne dieser Rollenkomplexität gerecht wird – oder soll sie zulassen, dass Defizite im einen Bereich durch überdurchschnittliche

Leistungen in einem anderen Bereich ausgeglichen werden können?

- Soll eine Punktebewertung oder ein relationaler Outputvergleich stattfinden? Im ersten Falle müsste ein Punktesystem geschaffen werden. Mit dessen Hilfe könnten einzelne Leistungen bewertet und zu einer in absoluten Zahlen ausdrückbaren individuellen Gesamtleistung addiert werden. Die dabei bestehende Gefahr ist die Inflationierung: Jede Überarbeitung des Punktesystems bringt mit großer Wahrscheinlichkeit weitere bewertungsfähige Leistungselemente hinein. Im Falle des relationalen Outputvergleichs würde hingegen das Leistungssystem einer Organisationseinheit – etwa eines Fachbereichs – in einem bestimmten Zeitraum gleich 100 % gesetzt; anschließend wäre zu bestimmen, welche individuellen Anteile die einzelnen Kollegen und Kolleginnen an diesem Gesamtsystem haben.
- Soll mit einem einzigen Erfassungsintervall oder mit zwei (oder mehr) sich überlagernden Erfassungsintervallen gearbeitet werden? Bei einem einzigen Erfassungsintervall besteht die Schwierigkeit, Leistungsqualitäten, die in unterschiedlichen Zeiträumen zu Stande kommen, adäquat ins Verhältnis zu setzen. Zwei sich überlagernde Erfassungsintervalle könnten den Vorteil haben, dass kurz- und langwellig entstehende Ergebnisse nicht in ein Bewertungsraster gezwängt würden, sondern je für sich gegenstandsangemessener gewürdigt werden könnten.

Selbstredend kann jede dieser vier Unterscheidungen auch mit einem Sowohl-als-auch beantwortet werden. Dann aber ist immer noch zweierlei zu entscheiden: In welchem Gewichtungverhältnis soll das jeweils eine und andere zum Zuge kommen, und werden zur Bewertung des einen und des anderen jeweils spezifische Instrumente benötigt?

Diese Anforderungen an die Entwicklung angemessener Bewertungsprogramme mögen vielleicht als zu anspruchsvoll, kompliziert oder aufwändig erscheinen. Sie zu ignorieren oder nur teilweise zu berücksichtigen wird aber mit großer Wahrscheinlichkeit dazu führen, dass besoldungsrelevante Leistungsbewertungen entweder keine hinreichende Akzeptanz finden oder aber informell unterlaufen werden. In beiden Fällen wäre es ressourcenschonender, von vornherein auf Leistungsbewertungen zu verzichten bzw. etwaige gesetzliche Vorgaben dadurch zu bedienen, dass man sich intern darauf einigt, alle gleichermaßen profitieren zu lassen. □