

Forum 2: Leistungsmessung in der Lehre

Leistungsindikatoren als Qualitätsindikatoren – Eine Wegbeschreibung

Peer Pasternack

Als der Alchimist Johann Friedrich Böttger 1703 von August dem Starken in ein Zwangspraktikum verpflichtet wurde, sollte er aus unedlen Metallen pures Gold gewinnen. Am Ende war er daran beteiligt, das europäische Porzellan zu entdecken, dessen fabrikmäßige Herstellung er dann in Meißen organisierte. Das Ziel seiner Festungshaft hatte er gleichwohl verfehlt, denn Gold zu gewinnen war ihm nicht gelungen. Wie würde heute ein wissenschaftliches Praktikum bewertet werden, das derart zielverfehlend ausging, aber gleichzeitig in anderer Hinsicht überraschend erfolgreich war? Es hinge von der Art der Bewertung ab.

Grundsätzlich heißt Qualität zu erzeugen an Hochschulen zweierlei: Zum einen sollen bestehende Standards gesichert, d.h. deren Unterschreitung verhindert werden. Zum anderen soll die Normüberschreitung bzw. Normabweichung ermöglicht werden. Hochschulen sollen in der Forschung das bisher noch nicht Entdeckte entdecken und das bisher noch nicht Gedachte denken. In der Lehre sind sie aufgefordert, keine geschlossenen Wissensbestände zu vermitteln. Stattdessen sollen sie dem Stand der Forschung entsprechendes Wissen lehren, d.h. ein Wissen, das in seiner Gewissheit zumindest zum Teil noch fragil und nicht kanonisiert ist. Hinzu tritt der Auftrag, die Fähigkeit zu vermitteln, dieses Wissen selbstständig zu bewerten, zu hinterfragen und die Folgen seiner Anwendung zu beurteilen. Ebenso wenig sollen die Hochschulen ihre Studierenden auf irgendein normiertes Persönlichkeitsbild hin zurichten. Hochschulen sind also ausdrücklich gehalten, Normen zu überschreiten, statt sich von ihnen fesseln zu lassen. Zugleich gilt auch an Hochschulen die Paradoxie von Stabilität und Entwicklung: Die Einhaltung methodischer Standards – also Normen – ist Voraussetzung für die Erzielung inhaltlicher

Normabweichungen – etwa von Erkenntnissen, die paradigmatische Umbrüche einleiten.

Wie kann nun die Bewertung von derart spezifischen Leistungen organisiert werden?

1. Grundlegende Unterscheidungen

Hilfreich ist es zunächst, zwei Unterscheidungen zu treffen. Zum einen ist die Elementardifferenzierung zwischen Quantitäten und Qualitäten vorzunehmen: Erstere sind messbar, letztere nicht. Zum anderen sind zwei Arten von Qualität voneinander abzusetzen.

Wenn von Hochschulqualität gesprochen wird, sind regelmäßig zwei völlig verschiedene Phänomene gemeint. Zum einen wird von „Qualitäten“ – im Plural – gesprochen. Damit werden einzelne Merkmale oder Eigenschaften bezeichnet, etwa die Fremdsprachenkompetenz von Studierenden („Qualität erster Ordnung“). Es werden damit Aspekte des Betrachtungsgegenstandes benannt, die sich im Übrigen mit ganz unterschiedlichen weiteren Qualitätsmerkmalen verbinden können. Zum anderen wird mit „Qualität“ – im Singular – etwas bezeichnet, das einen Prozess, eine Leistung oder ein Gut ganzheitlich durchformt. Darunter sind komplexe Eigenschaftsbündel zu verstehen, die den Betrachtungsgegenstand in seiner Gesamtheit prägen („Qualität zweiter Ordnung“). So kann etwa ein Studiengang internationalisiert sein, wovon aber nur dann mit Berechtigung zu sprechen ist, wenn sich dies in allen seinen Bestandteilen – inhaltlich, zeitlich, (fremd-)sprachlich, personell, strukturell – niederschlägt. Die Einzeleigenschaften sind zwar nicht metrisch, aber immerhin verbal, also beschreibend standardisierbar; Qualität als ganzheitlich durchformende Güte dagegen ist nicht zu standardisieren.

Ein Beispiel für die beiden getroffenen Unterscheidungen: Das Lehrkräfte-Studierenden-Verhältnis an einem Fachbereich ist ein zu messender, also quantitativer Sachverhalt. Dieser ist in qualitativer Hinsicht für sich genommen nur von sehr eingeschränkter Aussagekraft. Die Lehrkräfte-Studierenden-Interaktion ist ein verbal standardisierbarer Sachverhalt, der schon deutlichere Qualitätsaussagen erlaubt. Dabei handelt es sich um eine qualitative Einzeleigenschaft. Die Gesamtgüte eines Fachbereichs

hingegen kommt erst in den Blick, wenn man sich bemüht, die Lehrkräfte-Studierenden-Interaktionswirkungen zu ermitteln. Hierbei sind Standardisierungen wenig hilfreich, weil sie Fachbereichsspezifika nicht zu erfassen vermögen.

Diese Unterscheidungen haben Folgen für die Wahl des Interventionsinstrumentariums, mit dem Qualität an Hochschulen gesichert und entwickelt werden soll. Zu unterscheiden ist zwischen Single-issue-Ansätzen und Systemveränderungsansätzen. Erstere können geeignet sein, um zielgenaue Einzeleigenschaften zu sichern und zu entwickeln. Letztere sind von Nöten, sobald eine bestimmte ganzheitlich durchformende Güte erzeugt werden soll.

2. Qualität bewerten?

Mancher entzieht sich den immanenten Schwierigkeiten der Qualitätsbewertung durch dialektische Pfiffigkeit. So hatte bspw., nicht ohne taktische Raffinesse, die Mitgliedergruppe Musikhochschulen in der Hochschulrektorenkonferenz 1999 ein Papier zum Thema verfasst. Dort hieß es unter der Überschrift „Einführung von Kennzahlensystemen“ zunächst: „Die Musikhochschulen halten die Einführung von Kennzahlensystemen auch in ihrem Bereich für realisierbar.“ Darauf folgte eine Auflistung möglicher Kennzahlen (Anzahl der Studienbewerbungen, der Studienanfänger und Studierenden insgesamt, Anteil der Studentinnen, der ausländischen Studierenden, Studierenden in der Regelstudienzeit, Anzahl der Absolventinnen und Absolventen, der Musikhochschulwechsler, Studienabbrecher, der Lehrkräfte insgesamt und Anteil der weiblichen Lehrkräfte, Anzahl der Hochschulkonzerte/-veranstaltungen sowie Anzahl der Hochschulveröffentlichungen). Abschließend folgte der Satz: „Diese beispielhafte Aufstellung quantifizierbarer Kriterien macht gleichzeitig deutlich, wie problematisch es ist, die Leistungen von Musikhochschulen tatsächlich adäquat darzustellen, da sie von Quantifizierungen nur zu einem Teil erfasst werden.“¹ Die Botschaft also war: Wir haben doch nichts gegen Kennziffern, weisen aber auf ihre Unsinnigkeit hin.

¹ Mitgliedergruppe Musikhochschulen in der Hochschulrektorenkonferenz: Musikhochschulen an der Schwelle des 21. Jahrhunderts. Thesenpapier. Zur Kenntnis genommen vom 189. Plenum der HRK am 08.11.1999, Bonn 1999, S. 11

Auch außerhalb der künstlerischen Ausbildung bilden quantitative Kennziffern wie etwa Absolventen- oder Dropout-Quoten, die Auslastungsquote von Hochschulräumlichkeiten oder kurzfristige Verwertungseffekte den hochschulischen Leistungscharakter nicht adäquat ab. Das Studienziel ‚kritisches Denken‘ entzieht sich ebenso einer quantifizierenden Erfolgs-/Misserfolgsbewertung, wie die Messung einer Innovationsrate von Grundlagenforschung beträchtliche Schwierigkeiten bereitet. Der Wissenschaftshistoriker Klaus Fischer meint, nach „heutigen Evaluationskriterien wären viele Universitäten dafür bestraft worden, dass sie Pioniere der Wissenschaften beherbergten – Neuerer, die gegen den Strom der Orthodoxie forschten, nur eine bescheidene Zahl von Publikationen unterbringen konnten und deren Anerkennung in Form von Ehrungen und Zitationen spät oder zu spät kam“.²

Aus jüngerer Zeit ist hier das niederländische Beispiel aufschlussreich: In den 1980er Jahren war dort ein neues System der Zuweisung von Forschungsmitteln eingeführt worden, das auf einer internen und externen Qualitätsevaluation der Forschung beruhte. Der Automatismus nicht leistungsbezogener Mittelzuweisungen sollte beseitigt werden. Beabsichtigt war statt dessen, durch den Einbau externer Bewertung von Projekten die ‚beste‘ Forschung zu identifizieren und dadurch besonders fördern zu können. Ende der 80er Jahre musste festgestellt werden, dass das neue System nur dazu beigetragen hatte, die Einhaltung von Mindeststandards zu sichern. Nicht aber war es gelungen, tatsächlich herausragende Forschung zu identifizieren und zu fördern: „Die Rahmenbedingungen begünstigten eher die Beantragung von risikolosen Forschungsvorhaben.“³ Das hätte nicht zwingend überraschen müssen, denn wie die Dinge in der Wissenschaft vonstatten gehen, war nicht gänzlich unbekannt: „Es ist eher die schöne Ausnahme als die Regel, dass entscheidende Durchbrüche in der Forschung in ihrem Anfangsstadium angekündigt

² Klaus Fischer: Evaluation der Evaluation Teil I, in: Wissenschaftsmanagement 5/1998, S. 16-21, hier S. 19

³ Roland Richter: Der niederländische Weg zur Modernisierung der Hochschulen. Ein Bericht über die Hochschulreform der 80er und 90er Jahre, hrsg. von der Gewerkschaft Erziehung und Wissenschaft, Frankfurt a.M. 1998, S. 75 und 77

werden als bedeutende Arbeit bedeutender Leute“, fasste de Solla Price schon vor längerem zusammen.⁴

Ebenso lässt sich die Qualität von Hochschulleistungen nicht an Merkmalen festmachen, wie sie z.B. bei industriellen Produkten Anwendung finden, etwa Haltbarkeit (Käse), Hitzebeständigkeit (Stahl) oder Ausstattungsgrad (Auto). Dem entspricht, dass Qualität an Hochschulen nicht nach industriellem Schema gefertigt werden kann, indem lediglich ein übersichtliches Handlungsprogramm in Gang zu setzen wäre, das die Ursachen erzeugte, als deren Wirkungen dann zwangsläufig Qualität entstünde. Vielmehr kann sie dadurch entstehen, dass die Bedingungen so gestaltet werden, dass Qualitätserzeugung nicht verhindert wird. Eine Entstehensgarantie ist das freilich nicht. Einer solchen Garantie steht eine Reihe von Abhängigkeiten entgegen: Abhängigkeiten von sozialen Konstellationen, Bewertungen durch die Community (die nicht alles, was Qualität ist, als solche goutiert), von kognitiven Situationen, Wettbewerbsbedingungen (eine Modellrechnung, die in der Volkswirtschaftslehre als wissenschaftlich großer Wurf gilt, erzeugt in der Mathematik mitunter ein müdes Lächeln statt ihrer Anerkennung als Qualitätsleistung) usw. usf.

Andererseits ist für die Bewertung von Qualität dreierlei typisch: Erstens ist, ob Qualität bewertet wird, nicht ins Belieben der Beteiligten gestellt. Denn solche Bewertungen finden permanent statt, ggf. auch ohne dass sich die Beteiligten dessen bewusst sind: in Berufungsverfahren, Abschlussprüfungen oder bei Entscheidungen über Mittelaufteilungen zwischen Fachbereichen. Zweitens finden sie oft nach nicht explizierten und damit intransparenten Kriterien statt. Insoweit ermöglicht die Bestimmung dessen, was Hochschulqualität sein soll, Kriterien explizit zu machen, die implizit ohnehin angewandt werden. Drittens finden Qualitätsbewertungen häufig auf der Grundlage von Kriterien statt, die gar keine Qualität bewerten (sondern meist Quantitäten), oder unter Verwendung von „Symptomen der Qualität“, nämlich der Reputation von Institutionen

⁴ Derek John de Solla Price: Little Science, Big Science. Von der Studierstube zur Großforschung, Frankfurt a.M. 1974, S. 15

und ForscherInnen⁵, oder aber auf der Grundlage nicht oder unzulänglich offengelegter Maßstäbe.⁶

Die unzureichende Formulierung der Maßstäbe für Hochschulqualität hat problematische Folgen: Sie führt dazu, dass ein nicht oder nur vage rationalitätsgebundenes Instrument direkt oder indirekt verteilungsrelevant für Mittelallokation und individuelle Chancen wird. Qualitätsbewertung transparent zu gestalten, ist daher die Voraussetzung dafür, Qualität als hochschulpolitisches Steuerungsinstrument zu rationalisieren. Die Steuerung kann damit nicht nur funktional qualifiziert werden – sie würde zugleich auch demokratisiert: indem mit der entstehenden Transparenz eine wesentliche Voraussetzung für Aushandlungsprozesse geschaffen wird.

3. Die Gestaltung der Qualitätsbewertung

Qualitätsbewertungen lassen sich nicht nach Schemata organisieren, die dann für alle irgend denkbaren Fälle Anwendung finden können: Dafür sind die zu bewertenden Qualitäten zu verschiedenartig und zu komplex. Dennoch sollen Qualitätsbewertungen auch die Vergleichbarkeit verschiedener Fälle ermöglichen. Sie können also nicht willkürlich von Fall zu Fall festgelegt werden. Eine Lösung besteht darin, einige standardisierte Bewertungsmodule unterschiedlicher Reichweite und Gegenstandseignung zu entwickeln, die dann als Werkzeuge für differenzierte Anwendungen zur Verfügung stehen. Diese müssen methodisch zunächst immer zwei elementare Anforderungen erfüllen: Validität und Reliabilität. Valide sind solche Instrumente dann, wenn mit ihnen tatsächlich das erfasst wird, was erfasst werden soll. Reliabel sind sie, wenn wiederholte Überprüfungen am gleichen Objekt und unter gleichen Bedingungen zu den gleichen Ergebnissen kommen, d.h. wenn die Ergebnisse reproduzierbar sind.

Welche der standardisierten Bewertungsmodule dann jeweils eingesetzt werden, hängt von den Zielen der je konkreten Bewertung ab. Am An-

⁵ Peter Weingart: Prospektion und strategische Planung. Konzepte einer neuen gesellschaftsorientierten Wissenschaftspolitik, in: *Wirtschaft & Wissenschaft* 3/1995, S. 44-51, hier S. 48

⁶ Die diesbezüglich überzeugendsten Schilderungen und Interpretationen wurden in der Frauenhochschulforschung und anderen für Geschlechtsspezifika sensibilisierten Darstellungen vorgelegt.

fang von Qualitätsentwicklung stehen immer Zweckbestimmungen. Diese inspirieren Zieldefinitionen, und die wiederum sind elementare Voraussetzungen jeglicher Qualitätsorientierung. Wo anfangs Ziele definiert worden sind, werden nicht nur Zielverfehlungen und Zielerreichungen deutlich, sondern dort können auch zwar ungeplante, aber dennoch erreichte Ergebnisse – etwa Porzellan statt Gold – präziser identifiziert werden. Ebenso lassen explizierte Zieldefinitionen aber auch ggf. bestehende Zielkonflikte deutlich werden.

Praktisch muss jede Qualitätsbewertung zwei Bedingungen erfüllen: Zum einen hat sie abzubilden, was die Akteure interessiert. Zum anderen muss sie hinsichtlich der zur Verfügung stehenden Ressourcen praktikabel sein. Die Frage: welche personellen, finanziellen, sächlichen und zeitlichen Ressourcen stehen für die Durchführung der Qualitätsbewertung zur Verfügung?, ist zwar selbst nicht inhaltlich, doch prägt sie den Vorgang erheblich in auch inhaltlicher Hinsicht. Sie sollte daher möglichst frühzeitig geklärt werden. Darüber hinaus sollte sich jede Qualitätsbewertung folgenden Anforderungen regelhaft unterwerfen, um Akzeptanz für ihre Ergebnisse sicherzustellen:

- Benennung der Bewertungsmotivation, des Bewertungszwecks, der Bewertungsziele und -gegenstände;
- Benennung dessen, was nicht bewertet werden soll, d.h. Angabe der Grenzen des konkreten Bewertungsvorgangs;
- Referenzkonzept: in Bezug worauf soll die Qualität bewertet werden? (Effektivität, Effizienz, Demokratie, Output, Outcome, Bildungsbeteiligung, Impact o.a.)
- Angabe der Betrachtungsebene (Hochschule, Fakultät/Fachbereich, Institut, Fachdisziplin).

Methodisch sind sodann fünf grundsätzliche Entscheidungen zu treffen:

- 1) Entscheidung über geschlossene oder offene Bewertung: In der geschlossenen Bewertung wird die Gesamt-Qualitätsproduktion einer Organisation in einem Zeitraum gleich 100 Prozent gesetzt, und jeder Qualitätsindikator wird hinsichtlich seines prozentualen Anteils am Gesamtergebnis bestimmt, d.h. gegenüber den anderen Indikatoren ge-

wichtet. In der offenen Bewertung hingegen wird auf systemüberschreitende Maßstäbe Bezug genommen, etwa „internationales Niveau“.

- 2) Entscheidung darüber, ob eine vergleichende oder eine Einzelfallbetrachtung vorgenommen werden soll.
- 3) Entscheidung darüber, ob eine subjektive Zufriedenheitserhebung oder eine objektivierte Output- bzw. Outcomeerhebung stattfinden soll.
- 4) Entscheidung über den Einsatz von Peer Review oder Leistungsindikatoren oder einer Kombination aus beiden.
- 5) Entscheidung über die konkreten Bewertungsinstrumente und -verfahren: Was ist womit zu erheben, was in welcher Weise zu interpretieren?

Grundsätzlich stehen für die Qualitätsbewertung Instrumente auf drei Ebenen zur Verfügung:

- Bewertungsprogramme (Berichtssysteme, Monitoring, Evaluation, Akkreditierung, Qualitätsaudit, Zertifizierung, Benchmarking u.a.),
- Bewertungstechnologien (z.B. Mapping of Science, Wissensbilanz, Lehrberichte, Studentische Lehrbewertungen, Rankings, Absolventenbefragungen, Prüfungen, indikatorengestützte Bewertung),
- Leistungsindikatoren (Drittmittelquote, Impact-Faktor, Studienabbruchquote, Patentanzahl usw.).

Diese drei Ebenen stehen in einem hierarchischen Verhältnis zueinander: Ein Bewertungsprogramm wird aus einer Kombination mehrerer Bewertungstechnologien gebildet, und Bewertungstechnologien benötigen diverse konkrete Leistungsindikatoren (sofern sie nicht allein auf Peer Review bauen). Bewertungsprogramme müssen einem definierten Ziel folgen. Bewertungstechnologien dagegen sind vom Grundsatz her zielblind. Sie können im Rahmen an beliebige Ziele gebundener Bewertungsprogramme eingesetzt werden, soweit sie für den dortigen Bewertungszweck technisch geeignet sind. Je nach Zweck der angestrebten Qualitätsbewertung kann es genügen, nur einige Leistungsindikatoren heranzuziehen, oder aber eine bestimmte Bewertungstechnologie (innerhalb derer mehrere Leistungsindikatoren kombiniert werden) anzuwenden.

Wenn Qualitätsbewertungen abbilden sollen, was die Akteure interessiert, und sie zugleich praktikabel sein sollen, werden pragmatische Vereinfachungen benötigt. Diese ziehen aber leicht den Vorwurf auf sich, die Hochschulleistungen unterkomplex abzubilden. Gibt es Auswege aus diesem Dilemma?

Ein Vorschlag: Ein Schichtenmodell könnte die Lösung sein, die heuristisch zusammenführt, was getrennt zusammengehört: Ganzheitliche Qualität („Qualität zweiter Ordnung“) ist die Tiefenschicht hinter der Benutzeroberfläche, auf der die Akteure mit den Einzeleigenschaften („Qualitäten erster Ordnung“) operieren. Die Akteure entlasten sich davon, fortwährend das Ganze in seiner chaotischen, d.h. nichtlinearen Verfasstheit zu denken und zu prozessieren: Die Mehrdimensionalität der Qualität zweiter Ordnung wird in der Zweidimensionalität der Benutzeroberfläche abgebildet. Dort lassen sich in Abhängigkeit von den je zur Verfügung stehenden Bearbeitungskapazitäten bestimmte – einzelne oder mehrere – Qualitäten erster Ordnung fallweise aufrufen und wegklicken. Organisationen, kollektive und individuelle Akteure wären überfordert, unablässig das Ganze in all seinen prozessualen und Bedeutungsverzweigungen zu bearbeiten. Zumal unter Bedingungen unvollständiger Information – also dem Normalfall sozialen Handelns – bietet die Unterscheidung von Benutzeroberfläche und Tiefenschicht ein entlastendes Orientierungsmuster.

Diese Form der pragmatischen Vereinfachung würde jedoch problematisch, sofern die Oberflächenschicht in der Wahrnehmung der Akteure essentialisiert wird. Wenn also das Bewusstsein, mittels pragmatischer Vereinfachungen auf der Oberfläche einer darunter liegenden Tiefenschicht zu agieren, verloren geht oder auch erst gar nicht entsteht, dann verfehlt Qualitätsbewertung ihr Eigentliches. Denn hochschulische Leistungsqualität besteht nicht beispielsweise darin, viele Drittmittel einzuwerben. Sie besteht vielmehr darin, z.B. Wissen zu erzeugen und zu vermitteln, welches gegebene Zustände so in Frage stellt, dass Optionen eröffnet werden, diese gegebenen Zustände zu verändern, und zwar positiv und reflektiert, d.h. im Wissen um die Risiken jeder Veränderung und im Wissen um die Risiken jeder Nichtveränderung. Wenn das dann auch noch drittmittelträchtig ist – umso besser. Oder: Wird ein hoher

Impact-Faktor zum Fluchtpunkt aller wissenschaftlichen Bemühung, dann kann z.B. Forschungsinnovation zwar ein zufälliger Kollateraleffekt sein, kann aber ebenso Opfer fehlgeleiteter Energien der Forscher/innen und also unter Umständen vollständig verfehlt werden.

Akteure, welche die Hintergrundregeln, nach denen Hochschulen eigen-sinnig funktionieren, nicht zu decodieren vermögen, können diese natur-gemäß nicht berücksichtigen und werden also mit großer Wahr-scheinlichkeit scheitern bei der Hochschulqualitätsbewertung. Die Komplettzer-tifizierung einer Hochschule etwa erzeugt weder sowohl kenntnisreiche wie auch reflektionsfähige Absolventen noch nobelpreisfähige For-schungsergebnisse, sondern Verdruss. Denn wo auch das zertifiziert wird, was besser un-zertifiziert bliebe – z.B. Wahrheitssuche, oder sagen wir: die Suche nach plausiblen und erklärungskräftigen Konstruktionen –, dort wird lediglich das Personal durch das Ausfüllen der Zertifizierungs-bögen enerviert.

Wenn also forsche Hochschulqualitätsmanager auf der Benutzeroberflä-che beliebig ihre Instrumente platzieren und durch unbekümmertes An-klicken zu aktivieren suchen, ohne die Funktionslogik der Hochschule in Rechnung zu stellen, dann werden sie fortwährend eines produzieren: erzwungene Deaktivierungen, vergleichbar den ‚Schutzverletzungen‘ älterer Windows-Versionen. Systemabsturz und notwendiger Neustart sind die kurzfristigen Folgen, Programmblockaden die langfristigen. Danken werden es ihnen die hochschulangemessen agierenden Quali-tätsentwickler an anderen Standorten, deren Hochschulen an den alsbald zermanagten Problemfällen vorbeiziehen.

Autorenangabe

Peer Pasternack, Dr. phil., Staatssekretär a.D., Forschungsdirektor am HoF Wittenberg
– Institut für Hochschulforschung an der Universität Halle-Wittenberg.
peer.pasternack@hof.uni-halle.de; <http://www.peer-pasternack.de>