SSOAR

Open Access Repository
www.ssoar.info

Exzellenz und Evaluationsstandards im internationalen Vergleich

Hornbostel, Stefan

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:
Hornbostel, S. (2007). Exzellenz und Evaluationsstandards im internationalen Vergleich. *Gegenworte - Hefte für den Disput über Wissen*, 17, 72-75. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-23627

Nutzungsbedingungen:
Dieser Text wird unter einer CC BY-NC-ND Lizenz (Namensnennung-Nicht-kommerziell-Keine Bearbeitung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
https://creativecommons.org/licenses/by-nc-nd/4.0/deed.de

Terms of use:
This document is made available under a CC BY-NC-ND Licence (Attribution-Non Comercial-NoDerivatives). For more Information see:
https://creativecommons.org/licenses/by-nc-nd/4.0

gesis
Leibniz-Institut
für Sozialwissenschaften

Mitglied der
Leibniz-Gemeinschaft

Stefan Hornbostel

Exzellenz und Evaluationsstandards im internationalen Vergleich

Amerika, du hast es besser
Als unser Continent, das alte,
Hast keine verfallene Schlösser
Und keine Basalte.
Dich stört nicht im Innern,
Zu lebendiger Zeit,
Unnützes Erinnern
Und vergeblicher Streit

Was Goethe 1827 mit Sehnsucht, Ironie und einer gehörigen Portion Fehleinschätzung in den *Zahme[n] Xenien* zu Papier brachte, ließe sich ohne Weiteres auch als Epigramm auf die Evaluationsdebatte in der Wissenschaft lesen: Die Vermutung, dass es anderorts besser sei – weil unbelastet von ständischen Relikten jahrhundertelanger Mandarinenherrschaft und Aufständen gegen selbige –, ebenso wie die Täuschung über die tatsächlichen Verhältnisse andernorts und der gleichwohl sehnsüchtige Blick auf die Verhältnisse in der Fremde charakterisieren zumindest die deutsche Diskussion um Evaluation.

Evaluation ist nicht neu – sofern man darunter verfahrensförmig bereitgestellte Informationen zur Bewertung von Wissen und Erkenntnis und auch ihrer institutionellen und personalen Träger versteht. Seit knapp 500 Jahren begleiten Evaluationen die neuzeitliche Wissenskultur. Die frühen Formen, wie der *Index librorum prohibitorum*, stützten sich noch auf die Autorität der katholischen Kirche und den universalen Anspruch der Theologie, auch wissenschaftliches Wissen auf Kompatibilität mit religiöser Dogmatik (und damit vermeintlich auch Wahrheit) prüfen zu können. Dem folgten Verfahren organisierter, wissenschaftlicher ›Selbstzensur‹ (durchaus im Spannungsfeld von Staat und Kirche), wie sie erstmals in Gestalt des ›Peer-Review‹ im 17. Jahrhundert von der Royal Society eingeführt wurden. In der Wirtschaft entwickelten sich im 19. Jahrhundert fortan die ersten Formen systematischer und öffentlich verbreiteter Bewertungen von Marktakteuren – Bewertungen, die selbst im Markt gehandelt wurden und deutlich Tendenzen der Oligopolbildung, aber auch der Produktion von Selffulfilling Prophecys aufweisen.

Für das Wissenschaftssystem stellte sich die Frage, auf welche Weise jene Leerstelle, welche die Erosion der theologischen ›Wahrheit‹ hinterlassen hatte, im Rahmen der Selbststeuerung einer autonom gewordenen Wissenschaft zu füllen sei. Der Rekurs auf einen säkularen Wahrheitsbegriff führte schnell in unauflösbare Zirkel. Die Einsicht, dass letztlich das als wahr angenommen wird, was Wissenschaftler einigermaßen konsensual für wahr halten, führte dazu, dass das soziale System der Wissenschaft mit seinen Reputationshierarchien als Ersatz für Wahrheitskriterien herhalten musste. Hier hat auch die Rede von der Exzellenz ihren Ursprung. In einer meritokratisch orientierten Gesellschaft steht eine mehr oder weniger ständische Wissenschaftsverfassung allerdings in einem legitimatorischen Dauerkonflikt. Er ist nur stillzustellen, wenn sich zeigen lässt, dass die soziale Verfassung auch der Erkenntnisproduktion dient. Dass ebendies der Fall ist, zumindest wenn alle Beteiligten den Ehrenkodex der Wissenschaft, das wissenschaftliche Ethos nicht verletzen, versuchte der amerikanische Soziologe Robert K. Merton zu verdeutlichen. Das Streben der Wissenschaftler nach Reputation führt bei funktionierenden Institutionen zu größtmöglichem Fortschritt. Allerdings blieb dies nicht ohne den Hinweis auf die Labilität der Konstruktion: Wissenschaft habe durchaus pathogene Züge. Das führt dazu, wie Niklas Luhmann trefflich paradoxierte, dass die Orientierung an Reputation nicht selbst in guten Ruf kommen darf. Die Evaluation der ostdeutschen Hochschulen nach der Wende kann als eindrückliches Beispiel gelten für die Vermutung einer schweren Erkrankung durch politische Unterwanderung der Selbststeuerung des Wissenschaftssystems.

Wissenschaft hat, lange bevor von Evaluation die Rede war, neben den fachlichen Diskursen um methodische Absicherung der Erkenntnisproduktion eine Reflexion auf ihre soziale Verfasstheit entwickelt. Diese Reflexion war nicht unbeeinflusst von den jeweiligen nationalen Institutionalisierungsformen und bediente sich durchaus unterschiedlich aus dem Steinbruch der historisch entwickelten Evaluationsverfahren. So beschäftigte sich die US-amerikanische Wissenschaftsforschung – in einer stark durch private Akteure und marktförmige Distributionsstrukturen geprägten Wissenschaftslandschaft – vor allen Dingen mit dem internen Belohnungssystem, den Stratifikationsmustern, den normativen Grundlagen und der Formation von Konsens bei der Bewertung wissenschaftlicher Leistungen. Sie entwickelte früh quantifizierende Techniken und Indikatoren zur Beschreibung des Wissenschaftssystems. Die britische Wissenschaftsforschung thematisierte hingegen eher die Veränderungen von Wahrheitskonzepten und normativen Vorstellungen über die Zeit und über die Unterschiede zwischen den Disziplinen. Statt Konsensbildung nahm sie stärker Konflikte und Verhandlungsprozesse in der Wissenschaft in den Blick.

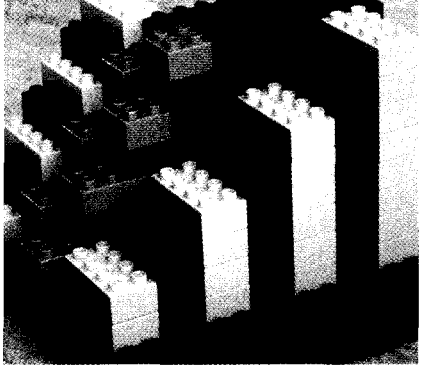
Einen besonders ausgeprägten Zweifel an der ›Passung‹ von sozialer Positionierung in der Wissenschaft und dem Beitrag zum Erkenntnisfortschritt steuerte der französische Soziologe Pierre Bourdieu bei. Er schlug vor, das alte Problem dadurch zu bearbeiten, dass man das soziale Kapital der Wissenschaftler nach rein wissenschaftlichem und institutionellem Kapital differenziert. Tendenziell ließen sich dann die eigentlichen Forscher und Wissenschaftler von denen unterscheiden, die ihre Karriere weitgehend dem Einsatz von sozialem oder politischem Kapital verdanken. Die damit transportierte Vorstellung, man könne über Publikations- und Zitationsindikatoren oder Wissenschaftspreise zu einer ›objektiven‹ Messung des wissenschaftlichen Kapitals kommen, hat nicht unerheblich zu jener Selbstmystifikation in der Wissenschaft beigetragen, die in den heftigen Debatten um die Exzellenz-Initiative fröhliche Urständ feiert.

Die vermeintlich ›objektiven‹ Messungen sind selbst zutiefst durch die Schichtungsstruktur des Wissenschaftssystems geprägt; sie geben kein objektives, sondern ein anderes Bild. Und natürlich werden Indikatoren in das strategische Kalkül von Wissenschaftlern und wissenschaftlichen Institutionen einbezogen. Will man in diesem Feld der wissenschaftlichen Selbstreflexion von Standardbildungen sprechen, dann unterscheiden sich diese nicht von entsprechenden Prozessen in anderen Forschungsfeldern. Standardbildung kann hier verstanden werden als die methodische Debatte darüber, wie sich Urteilskonsens, partikulare Einflüsse in Entscheidungsprozessen, Geschlechterbias, Reputation, Erkenntnisproduktion, Rezeption und Ähnliches adäquat messen lassen. Wie in anderen Forschungsfeldern auch entwickeln sich solche Standards koevolutionär mit der Forschung und machen natürlich nicht an nationalen Grenzen halt.

Die Vorstellung, man könne über Publikationsindikatoren oder Wissenschaftspreise zu einer ›objektiven‹ Messung des wissenschaftlichen Kapitals kommen, hat nicht unerheblich zu jener Selbstmystifikation in der Wissenschaft beigetragen, die in den Debatten um die Exzellenz-Initiative fröhliche Urständ feiert.

Etwas anders sieht das Bild aus, wenn man nach Standards in jenem Feld fragt, das sich in den letzten 20 Jahren unter dem Titel ›Evaluation‹ etabliert hat. Als charakteristisch für das neue Verständnis von Wissenschaftsevaluation erweist sich eine Bedeutungs- und Anwendungserweiterung: Es handelt sich nicht mehr um einen reinen Selbstreflexionsprozess, sondern um eine Mischung aus Reflexion, Rechenschaftslegung, Positionsbestimmung, Managementinformation, Politikberatung, Mittelverteilung, Personalpolitik, öffentlicher Information, Reputationsrankings oder Planungsinstrumentarien. Entsprechend sind die Produzenten und Adressaten in diesem Feld nicht mehr nur Wissenschaftler, sondern eben auch Medien, Politik, Verwaltung oder – etwas euphemistisch ausgedrückt: ›Kunden‹ (Studenten, Postdocs, Forscher).

Der Ausgangspunkt für derartige Wissenschaftsevaluation (sowohl in der Lehre wie in der Forschung) war meist ein sehr nationales Phänomen: Politik, Medien und zum Teil die Wissenschaft selbst äußerten Besorgnis über Rückständigkeit, Ineffizienz, mangelnde Qualität oder Relevanz. Eine verstärkt utilitaristische Haltung gegenüber der Wissenschaft artikulierte Befürchtungen, im



internationalen Wettbewerb auch ökonomisch zurückzufallen und als Wissenschaftsstandort an Attraktivität einzubüßen. Diese Klagen und Befürchtungen sind nicht deutschlandspezifisch, sondern international zu beobachten. Die in der Folge entwickelten Evaluationen thematisierten genau besehen daher *drei verschiedene*, aber nicht voneinander unabhängige *Formen der Standardbildung*: a) die wissenschaftsinternen methodischen Standards, b) Verfahrensstandards und c) Standards im Sinne von ›Benchmarks‹ oder Referenzobjekten.

Der *erste Typ von Standards* greift auf die Wissenschaftsforschung zurück und fragt danach, inwieweit für eine wissenschaftlich fundierte Evaluation vonseiten der Forscher methodisch belastbare Informationen für den Bewertungsprozess bereitgestellt werden können. Hier prallen regelmäßig die Wünsche der Praktiker nach einem zweifelsfreien Superindikator und die vorsichtige ›Ceteris-paribus-Kultur‹ der Wissenschaft aufeinander. Die als analytische Instrumente entwickelten Messtechniken und Indikatoren lassen sich keineswegs problemlos in leicht administrierbare Kennzahlen transferieren.

Beim *zweiten Typus* geht es darum, mit welchen Verfahren sich Evaluationen bewerkstelligen lassen. Hier wurde sehr früh in das europäische Ausland geschaut und Praxen wie etwa zweistufige Evaluationen, bestehend aus Selbstbericht und externer Begutachtung, etabliert. Teilweise hat man sich (mit begrenztem Erfolg) auch an Verfahrensstandardisierungen des Qualitätsmanagements aus der Industrie angelehnt (ISO 9000, Total Quality Management und Ähnliches). Schließlich sind Standards durch Übung in dem Maße entstanden, wie Evaluationsagenturen ihre Evaluationen iterierten. Und last not least finden sich inzwischen Übereinkünfte (zum Beispiel für die Akkreditierung) auf europäischer Ebene. Dort empfehlen *Standards and Guidelines for Quality Assurance in the European Higher Education Area* bestimmte Verfahrensabläufe ebenso wie die Komposition von Gutachtergremien (etwa die Hinzuziehung ausländischer Experten). Im Unterschied zu den erstgenannten Standards handelt es sich bei den Verfahrensstandards um mehr oder weniger reflektierte Empfehlungen, die aus Praxiserfahrungen entstanden sind. Sie sichern keineswegs automatisch ein gehaltvolles Evaluationsergebnis, wohl aber die Einhaltung eines Verfahrensweges. Dass auch bei Einhaltung eines solchen Weges skurrile Ergebnisse entstehen können, lässt sich aus der Arbeit der Inquisitionsbehörde durchaus lernen. Gleichwohl führt Mimesis zu einer gewissen Harmonisierung der Verfahren. Das gilt auch für die Forschungsevaluation, wenn etwa der Wissenschaftsrat sich mit seinem Forschungsrating an das britische Vorbild des *Research Assessment Exercise* anlehnt. Ob allerdings deutsche Gutachter die gleichen Urteilsskalen und -kriterien verwenden, steht auf einem anderen Blatt.

Der *dritte Typus von Standards* oder besser Benchmarks hat zurzeit Hochkonjunktur. Internationales Niveau, internationaler Standard, internationale Entwicklung – so lauten die Schlagworte, mit denen meist recht nebulös auf ein Mehr, Schneller, Höher, Weiter, Besser oder auch Schlimmer in der Welt verwiesen wird. Handfeste Hinweise sind dagegen recht schwer zu beschaffen. Am weitesten vorangeschritten ist die Messung derartiger Standards sicherlich im Bereich der Bibliometrie. Dort lässt sich zumindest für einzelne Disziplinen und Auditorien so etwas wie ein weltweites ›Normal‹ an Publikationen und Zitationen definieren, allerdings durchaus um den Preis, dass kultur- oder sprachaffine Forschungsbereiche (darunter auch die anwendungsbezogene Forschung) durch das analytische Raster fallen. Der Trend bei der Festlegung eines geeigneten Referenzrahmens geht daher auch deutlich in Richtung eines profilbezogenen Benchmarkings. Forschungsinstitutionen suchen gezielt nach vergleichbaren Einrichtungen, an denen sie sich messen wollen. Weitaus unklarer ist die Frage der internationalen Standardbildung beim Einsatz von Peer-Review-Verfahren. Auf der einen Seite beobachten wir eine schnell anwachsende Internationalisierung des Gutachterwesens. Auf der andern Seite gibt es bisher wenig Wissen über eine entsprechende Standardisierung der Gutachterkriterien und der Maßstäbe, der Rücksichtnahmen und der Wirkung nationaler Review-Kulturen – zumindest dann, wenn der enge fachwissenschaftliche Beurteilungsdiskurs überschritten wird.

Als Fazit bleibt die Erkenntnis: Die Bildung von Standards wird wohl auch in Zukunft eine Mischung aus eingeübter und durchgesetzter Praxis, aus strategischen Kalkülen und einer kritischen Reflexion sein. Das ist in der Wissenschaft nicht anders als in der Ökonomie oder der Politik.

Literatur
P. Bourdieu: Vom Gebrauch der Wissenschaft. Für eine klinische Soziologie des wissenschaftlichen Feldes. Konstanz 1998
N. Luhmann: Die Wissenschaft der Gesellschaft. Frankfurt a. M. 1990
R. K. Merton: The Sociology of Science: Theoretical and Empirical Investigations. Chicago 1973