
Dossier
Qualitäts- und Leistungsmessung in
den Geisteswissenschaften



Qualitäts- und Leistungsmessung als Basis für eine gemeinsame Wissenschaftskultur

(bk) Vorbei sind die Zeiten, in welchen ein Mäzen Wissenschaftler um der Erkenntnisse willen forschen liess. Von Steuergeldern finanziert, muss auch die Wissenschaft immer mehr Rechenschaft ablegen. Transparenz wird heute in allen Bereichen gefordert. Allerdings eignen sich nicht alle Messinstrumente für alle Disziplinen gleich gut. In der Wissenschaft dominieren Instrumente aus den Naturwissenschaften, welche für die Geistes- und Sozialwissenschaften jedoch nicht geeignet sind. Das Projekt «Mesurer les performances de la recherche» der Rektorenkonferenz der Schweizer Universitäten (CRUS) trägt dem Rechnung und erstellt eine Übersicht über «die wichtigsten Initiativen auf schweizerischer und internationaler Ebene, welche die Forschungsleistung in den Geistes- und Sozialwissenschaften messen und dabei mögliche Verzerrungen aufgrund sprachlicher, kultureller und nationaler Gegebenheiten vermeiden». Ingesamt drei Initiativen und vier unterstützende Massnahmen wurden in diesem Rahmen gefördert, mit dem Auftrag «Unterschiede zwischen Disziplinen, kulturelle Ausprägungen, spezifische Wertschöpfung, welche über die wissenschaftliche Gemeinschaft hinausgeht (use value) und Existenzwert (non use value)» zu untersuchen.

In diesem Dossier legen die CRUS sowie Verantwortliche der Initiativen und Massnahmen dar, dass es auch in den Geistes- und Sozialwissenschaften durchaus möglich ist, nachvollziehbare und überprüfbare Angaben über Leistung und Qualität zu machen; auch wenn eine völlig objektive Sicht nie ganz möglich sein wird, wie Michèle Lamont ausführlich belegt. Wir schliessen uns der Ansicht von Wiljan van den Akker an, dass die geforderte Transparenz für die Geistes- und Sozialwissenschaften ein grosse Chance darstellt. Einerseits ergibt sich dadurch die Möglichkeit, sich innerhalb der Wissenschaftsgemeinschaft auf Qualitätsstandards zu einigen, und andererseits bietet sich die Chance für einen klaren Leistungsausweis gegen aussen. Diese

Inhaltsverzeichnis Dossier

- 35 Qualitäts- und Leistungsmessung als Basis für eine gemeinsame Wissenschaftskultur
- 36 Projet «Mesurer les performances» – Présentation, état des travaux, prochaines démarches
Raymond Werlen
- 38 A la recherche de la valeur d'usage de la science
Olivier Glassey et Jean-Philippe Leresche
- 40 Décrire et mesurer la «fécondité» des recherches en sciences humaines et sociales. Aperçu d'un projet
Jean-François Perret, Philippe Sormani, Alain Bovet, Alaric Kohler, Edo Poglia
- 42 Qualitätskriterien für die Forschung in den Geisteswissenschaften – Eine Explorationsstudie
Sven E. Hug und Michael Ochsner
- 44 Evaluieren in der Kommunikations- und Medienwissenschaft – Warum Aktivitäts-Profile die bessere Alternative sind
Diana Ingenhoff, Benedetto Lepori, Carole Probst
- 46 Wie «gut» ist juristische Forschung in der Schweiz?
Andreas Lienhard
- 47 Opening the black box of evaluation: How quality is recognized by peer review panels.
Michèle Lamont, Katri Huutoniemi
- 50 Qualitäts- und Leistungsmessung in den Geisteswissenschaften – Eine Pilotstudie des deutschen Wissenschaftsrats.
Christina Mair, Elke Lütke-meier, Veronika Khlavna
- 52 Yes we should; research assessment in the Humanities.
Wiljan van den Akker
- 54 European Educational Research Quality Indicators (EERQI).
Ingrid Gogolin

35

beiden Aspekte können zu einem gemeinsam geteilten Verständnis der eigenen Wissenschaftskultur beitragen.

Die Wissenschaftskultur steht 2011 bei der SAGW im Fokus. Wir lancierten diese Diskussion im Bulletin 3/2010, als wir die Geisteswissenschaften in Grossprojekten analysierten. Welche Wissenschaftskultur haben die Geistes- und Sozialwissenschaften und welche brauchen sie in Zukunft? Ein gemeinsames Verständnis von Qualität und gemeinsam gesetzte Standards sind sicher ein wichtiger Bestandteil.

Projet «Mesurer les performances» – Présentation, état des travaux, prochaines démarches

Raymond Werlen, secrétaire général adjoint de la Conférence des Recteurs des Universités Suisses (CRUS)

36

Des rankings toujours plus nombreux établissent régulièrement des «comparaisons» entre universités en se basant sur des analyses souvent peu appropriées et toujours très délicates à interpréter. Les instruments conventionnels permettent de visualiser l'impact de la recherche sur la communauté scientifique dans certains domaines des sciences exactes et naturelles, mais ils ne «voient» que très partiellement la recherche dans d'autres domaines, en particulier dans les sciences humaines et sociales. Le projet «Mesurer les performances de la recherche» cherche à doter les universités d'instruments adéquats pour leurs situations particulières et tenant compte de leurs stratégies propres.

Le projet «Mesurer les performances de la recherche» fait partie des projets de coopération et d'innovation que la Conférence universitaire suisse a lancés en 2008¹. Il bénéficie à ce titre d'un financement de la Confédération dans le cadre des contributions liées à des projets prévus par la Loi sur l'aide aux universités².

Il a été précédé par un état des lieux – réalisé par un expert mandaté – des principales initiatives suisses et internationales, visant à mesurer les performances de la recherche dans les sciences humaines en évitant les biais linguistiques, culturels et nationaux³, ainsi que par des analyses bibliométriques réalisées par les universités en collaboration avec le Center for Science and Technology studies (CWTS) de l'Université de Leyde et dont les résultats consolidés sont présentés dans le premier rapport du projet⁴.

1 <http://www.crus.ch/la-crus/coordonne-harmonise/projet-mesurer-les-performances-de-la-recherche.html>

2 <http://www.cus.ch/wFranzoesisch/beitraege/2008-2011/index.php>

3 http://www.sbf.admin.ch/hm/themen/uni/ufg_fr.html

4 <http://www.crus.ch/dms.php?id=10474>

5 <http://www.crus.ch/dms.php?id=10722>

Le projet finance trois types de mesures:

- la création de postes de spécialistes dans les universités,
- des initiatives visant à développer des instruments pour mesurer les aspects de la recherche qui échappent à la bibliométrie conventionnelle et
- des actions, dont le rôle est à la fois de favoriser la diffusion des expériences acquises dans le cadre des initiatives et d'explorer les pistes qui, faute de moyens ou de compétences avérées, n'ont pas été retenues sous la forme d'initiatives.

Les spécialistes des universités sont organisés en un réseau qui se réunit 3 à 4 fois par an. Ce sont eux qui mettent en œuvre les instruments au sein des universités. Ils ont notamment participé aux analyses bibliométriques susmentionnées et sont souvent impliqués dans les initiatives. Les trois initiatives financées dans le cadre de ce projet font l'objet d'autres articles dans ce numéro⁶. Il en va de même d'une des actions⁷.

Rankings et des instruments appropriés

L'idée du projet est issue de plusieurs constatations dont découle, pour les universités, la nécessité de développer des compétences en matière de mesure et de comparaison internationale des performances de leur recherche. Des rankings toujours plus nombreux établissent régulièrement des «comparaisons» entre universités en se

6 Jean-François Perret, Université de Neuchâtel: Mesurer la fécondité de la recherche en sciences humaines et sociales,

Sven Hug, Université de Zurich: Entwicklung von Qualitätskriterien für die Geisteswissenschaften – Eine Explorationsstudie in den Literaturwissenschaften und der Kunstgeschichte,

Diana Ingenhoff, Université de Fribourg: Erfahrungen und Erkenntnisse aus den Kommunikationswissenschaften

7 Andreas Lienhard, Universität Bern: Erfahrungen und Erkenntnisse aus den Rechtswissenschaften

basant sur des analyses souvent peu appropriées et toujours très délicates à interpréter. Cela dans un contexte de globalisation de l'enseignement et de la recherche qui oblige les universités à se positionner dans un espace compétitif international. L'autonomie grandissante dont jouissent les universités entraîne enfin de leur part, comme contrepartie, un devoir accru de reddition de compte pour répondre aux attentes légitimes du public et des autorités politiques qui les financent.

Les instruments conventionnels, essentiellement basés sur la bibliométrie classique, permettent de visualiser l'impact de la recherche sur la communauté scientifique dans certains domaines des sciences exactes et naturelles. Mais ils ne «voient» que très partiellement la recherche dans d'autres domaines, en particulier dans les sciences humaines et sociales. Ils ne permettent pas non plus de rendre compte d'autres aspects de la recherche comme son impact sur l'enseignement ou sur l'innovation économique et sociale.

Les instruments doivent dépendre des objectifs de la recherche considérée

La CRUS a exprimé en novembre 2008 sa conception de la qualité dans le système universitaire en 10 principes sous le titre «The Swiss Way to Quality in the University System»⁸. Un élément central de cette approche est la conception de la qualité comme dépendant de la mission, des objectifs et de la stratégie de l'université. Lorsqu'il s'agit de mesurer les performances de la recherche, cela signifie que les instruments doivent dépendre des objectifs de la recherche considérée.

Dans ce sens, le projet «Mesurer les performances de la recherche» ne cherche pas à développer une méthode unique qui serait applicable à toutes les recherches dans toutes les situations. Il cherche au contraire à doter les universités d'instruments adéquats pour leurs situations particulières et tenant compte de leurs propres stratégies. Le but n'est donc pas de développer un instrument avec lequel un organe externe pourrait mesurer les performances de la recherche de manière systématique mais une boîte à outils que les universités pourront utiliser selon leurs besoins.

Les spécialistes des universités ont acquis de l'expérience dans la manière de mettre en place différents instruments et les initiatives et actions ont entrepris d'explorer et de développer de nouveaux instruments correspondant à des situations particulières.

Plan d'action

Un colloque aura lieu le 7 octobre 2011 à Lausanne, au cours duquel les partenaires du projet et les responsables de la politique de la recherche des universités auront l'occasion d'échanger sur les besoins des universités et les contributions des différents partenaires du projet pour y trouver des réponses appropriées. Ce sera aussi l'occasion de s'interroger sur les priorités qui devront prévaloir lors d'une seconde phase du projet. La CRUS a en effet jugé, dans sa planification stratégique⁹, qu'il serait nécessaire de poursuivre les efforts pendant la prochaine période. Cela pourrait d'ailleurs faire l'objet d'un programme de la Conférence universitaire suisse (CUS) pour les années 2013-2016.

8 <http://www.crus.ch/dms.php?id=7480>

9 <http://www.crus.ch/dms.php?id=14913>

A la recherche de la valeur d'usage de la science

Olivier Glassey et Jean-Philippe Leresche, Observatoire science, politique et société, Université de Lausanne

38

Un changement se fait sentir dans le domaine de l'évaluation des sciences. La perspective externe tendrait à primer sur celle interne de laquelle on se basait: le focus est mis sur l'impact sociétal des sciences et non plus sur les sciences elles-mêmes. Les indicateurs traditionnels ne suffisant plus à l'évaluer, il faut en trouver de nouveaux.

L'évaluation des résultats de la recherche et de la productivité scientifique constitue à l'évidence un domaine en fort développement dans un contexte contemporain marqué par l'importance attribuée à l'économie de la connaissance. Cette montée en puissance de la mesure de la production scientifique rend compte de besoins accrus d'outils de lecture de l'activité du champ scientifique tels que formulés par les différentes catégories d'acteurs des politiques scientifiques (pouvoirs publics, agence de moyens, direction des établissements universitaires, etc.). Ces acteurs ne sont pas de simples «usagers» des outils de l'évaluation mais contribuent à les façonner en les reformulant à partir de l'expression des buts spécifiques qu'ils poursuivent. Dans ce contexte, depuis quelques années, on observe le renforcement d'une demande pour des instruments d'évaluation «à spectre plus large» qui prennent notamment en compte l'«impact» sociétal de la production scientifique. Cette tendance croise également des débats sur les missions et les finalités des universités, en particulier sur leur «troisième mission» qui suppose des interactions entre l'université et son environnement régional et économique. Ce double processus contribue au phénomène de «débordement» de la mesure de l'activité scientifique, au-delà du champ scientifique stricto sensu.

Aller au-delà de la perspective interne

Les débats conduits autour de la question de la nature des productions et usages de la science dans différentes arènes (scientométrie, transfert de technologie et

relations sciences et société) contribuent à influencer les manières de conceptualiser l'évaluation scientifique (balance entre des perspectives «internalistes» et «externalistes» de la science). Cependant, la mesure des outputs scientifiques traditionnels (publications, etc.) n'épuise pas la question des impacts plus larges (outcomes) de la recherche scientifique, que ce soit dans les sciences sociales et humaines ou les autres sciences. D'une manière générale, la réflexion en termes d'usage de la science apparaît ainsi comme une extension du périmètre de la mesure et de l'évaluation des apports de la science. Autrement dit, elle s'inscrit dans un processus plus large qui est celui de l'intégration grandissante dans la réflexion sur l'évaluation de la recherche des modalités d'appropriation sociale et d'actualisation des connaissances scientifiques dans le cadre des pratiques quotidiennes des acteurs sociaux. Aussi, dans cette perspective, n'est-ce plus la façon dont les sciences pensent contribuer à la société qui prédomine mais bien les manières dont les acteurs sociaux sont capables de mobiliser des éléments (quels qu'ils soient) issus de la recherche scientifique qui devient le point de départ de la réflexion sur les indicateurs de la mesure de la science.

L'utilité de la science pour la société, est-elle un instrument durable?

Cette piste de réflexion s'inscrit dans une analyse fine des rapports entre science et société. Elle considère les usages de la science non pas comme des produits dérivés d'une activité mais comme le point de rencontre (et parfois de tension/collision) entre, d'une part, la science, ses modes de fonctionnement, sa temporalité, ses buts et, d'autre part, des usagers de la science porteurs de leurs propres projets et valeurs (cf. l'Action 4 «Mesurer la valeur d'usage de la recherche» dans le programme (2008-2011) de la CRUS «Mesurer les performances de la recherche»). Une telle approche reconnaît que les règles de mise en circulation de valorisation, de validation et

de mise en œuvre ne sont plus de même nature une fois que les connaissances quittent le contexte spécifique de la communauté scientifique et de ses normes. Pour l'ensemble des disciplines, elle intègre le fait que la construction d'un référentiel visant à établir une évaluation des apports scientifiques à la société ne peut pas se limiter aux indicateurs traditionnels de l'activité scientifique.

Une telle approche en termes d'usage des sciences soulève cependant aussi un certain nombre de problèmes:

- l'accent mis sur la société n'est-il pas de nature à privilégier une vision essentiellement instrumentale de la science qui ne serait perçue qu'au travers de ses productions directement utilisables et utilisées par les acteurs sociaux?
- en l'absence d'un référentiel «stabilisé» des usages de la science dans la société, une mesure en termes de valeur d'usage qui s'établirait à partir de l'utilité d'une connaissance scientifique mobilisée en fonction d'un contexte et de besoins spécifiques ne serait-elle pas, par définition, de nature variable et volatile?

Une chance pour les sciences humaines et sociales?

Impliquant la «convocation» de nouveaux acteurs dans les processus d'évaluation sociale des sciences, cette perspective de la mesure par les usages s'inscrit aussi dans la définition d'un nouveau rapport de force entre évalués et évaluateurs et/ou entre l'Etat et les hautes écoles (lutte institutionnelle et politique pour le contrôle des instruments de «contrôle»). Elle pose également une question essentielle: une telle approche ouverte sur les usages sociaux offre-t-elle de nouvelles opportunités aux sciences humaines et sociales pour valoriser la diversité et la spécificité des formes de leurs contributions à la société?

A ce stade de la réflexion, on sait déjà que les sciences sociales figurent en première ligne dans l'analyse des pratiques et des modes d'appropriation des connaissances scientifiques. Dans une réflexion en termes d'usage, elles se trouvent donc particulièrement bien «outillées» pour étudier et révéler la capillarité des circulations de connaissances dans la société, notamment les flux qui ne passent pas par les canaux des publications scientifiques classiques mais qui débordent dans une pluralité d'espaces de vulgarisation et d'apprentissages formels et informels.

Décrire et mesurer la «fécondité» des recherches en sciences humaines et sociales. Aperçu d'un projet

Jean-François Perret, institut de psychologie et éducation, Université de Neuchâtel, avec la collaboration de Philippe Sormani, Alain Bovet et Alaric Kohler, Secteur Qualité de l'UniNe, et en partenariat avec Edo Pogliani de l'Université de la Suisse italienne

40

Le projet de coopération et d'innovation «Mesurer la performance de la recherche», piloté par la CRUS, comprend plusieurs volets. L'un d'eux vise le développement d'instruments qui permettent de mettre en évidence l'apport des recherches en SHS que la bibliométrie conventionnelle ne parvient pas à saisir. C'est dans cette perspective que s'inscrit notre «initiative» intitulée: «Décrire et mesurer la fécondité des recherches en sciences humaines et sociales, à partir d'études de cas». Nous présentons ici les principales options qui ont guidé la réalisation du projet.

Face à plusieurs options possibles, nous avons choisi de centrer notre attention sur la production scientifique des unités de recherche (instituts, laboratoires, centres ou autres groupes de recherche). Ces unités constituent le lieu par excellence où émergent de nouvelles problématiques, des projets de recherche et des sujets de thèse. Les apports multiples et divers de ces unités constitutives du paysage de la recherche sont parfois mal connus, ils méritent une attention particulière.

Regard sur la «fécondité» des activités de recherche

Nous avons retenu la notion de fécondité pour parler du déploiement des activités de recherche. L'intention est de mettre en lumière ce que les activités de recherche génèrent, au fil du temps, en termes de nouveaux travaux, de débats, d'interventions ou encore de collaborations locales, nationales et internationales.

Pour ce faire, nous préconisons d'analyser la production scientifique d'une unité de recherche sur une période minimum de 4 ou 5 ans, pour obtenir ainsi une image représentative du déploiement des activités d'un institut. Cette durée ne nous empêche cependant pas de prendre aussi en compte des contributions scientifiques plus anciennes, qui se sont révélées marquantes et ont inspiré des chercheurs quelquefois sur plusieurs décennies.

Les publications dans leurs contextes

Tout groupe de recherche se préoccupe de publier et de faire connaître ses travaux. La performance en termes de publications est cependant étroitement liée à d'autres activités interdépendantes qui exigent un engagement sur plusieurs plans: établir des contacts et des collaborations scientifiques; chercher des financements; former la relève; constituer et faire vivre un groupe de recherche, ou encore mener des activités en lien avec la cité. Chaque institut investit diversement ces champs d'action au gré des sollicitations, des subsides obtenus et autres événements qui façonnent leur agenda de recherche. Nous représentons ce système d'activités par le schéma suivant:

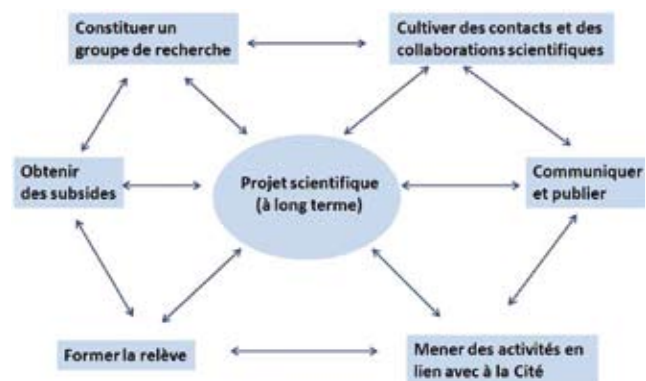


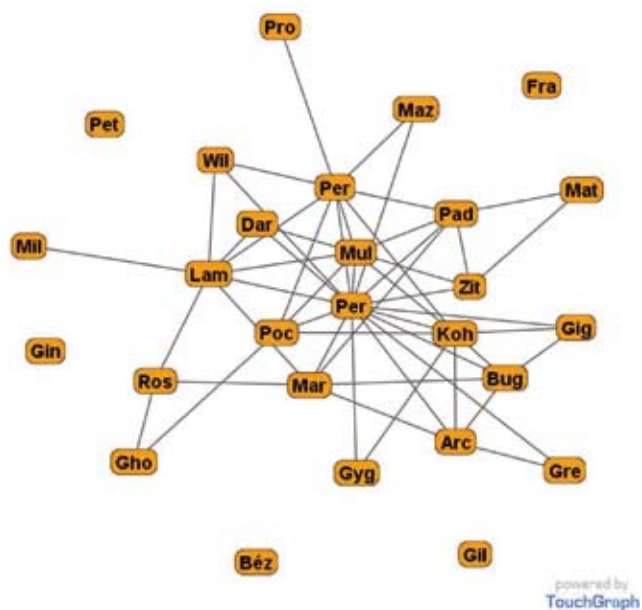
Schéma 1: Un système d'activités interdépendantes

Cartographier une production scientifique

Pour donner à voir les multiples aspects de la fécondité d'une unité de recherche, nous prenons en compte l'ensemble le plus complet possible de ses publications et communications. Tout texte produit est en effet susceptible de nous informer sur l'activité d'un institut. Dans ce but, nous avons développé une méthode d'indexation qui permet de caractériser tout écrit selon une quinzaine de descripteurs: auteurs et co-auteurs; type de publication et de communication, destinataires, langue, thématique, type d'apports, lien ou non avec

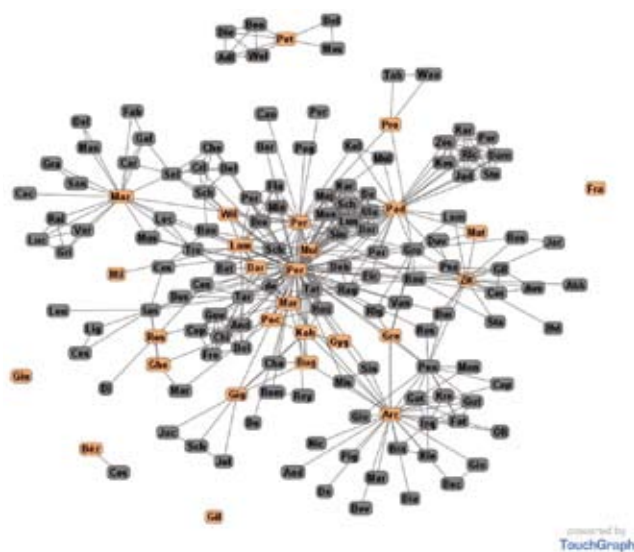
un projet subsidié, etc. Cette indexation est effectuée de manière à pouvoir cartographier la production scientifique d'un institut pour en visualiser les principales caractéristiques. L'outil de représentation graphique utilisé est le logiciel TouchGraph qui permet l'analyse de tout type de réseau. Le choix de cet outil s'est révélé judicieux pour mettre en relation, dans un même espace, les auteurs et leurs publications avec l'ensemble des descripteurs retenus. Le but n'est donc pas d'apprécier la valeur intrinsèque de chacune des publications mais plutôt de déceler ce qu'elles révèlent de l'orientation des activités d'un institut.

Quatre exemples, parmi les multiples graphes qu'il est possible de produire, donneront un aperçu de notre démarche. Ils portent sur l'analyse de quelques 500 publications et communications d'un institut sur la période 2005–2010. Le premier exemple (graphe 1) visualise le réseau des co-auteurs, membres de cet institut.



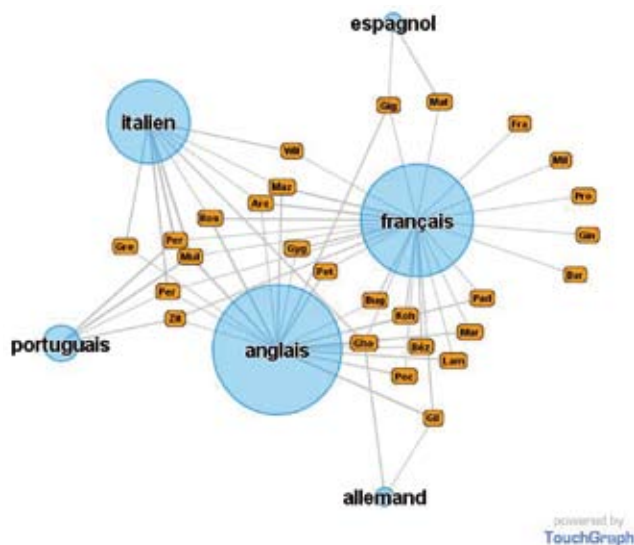
Grappe 1: Le réseau des co-auteurs, membres d'un institut

Ce graphe met en évidence que 22 des 27 chercheurs représentés ici ont cosigné un ou plusieurs textes avec leurs proches collègues. Bien entendu, ces chercheurs signent aussi des publications avec d'autres collègues suisses ou étrangers formant un plus large réseau de collaboration qui s'est construit au fil des ans (graphe 2).



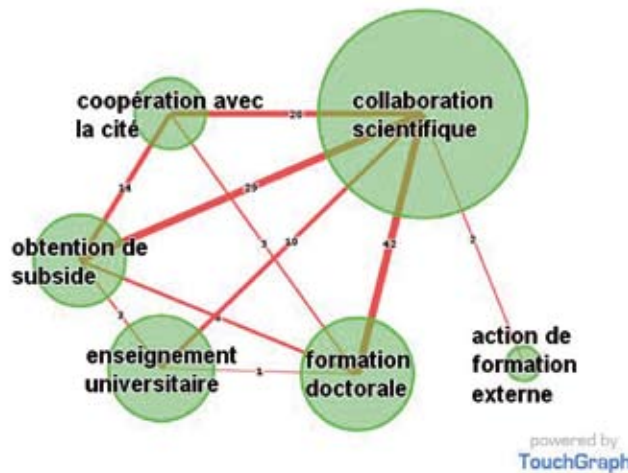
Grappe 2: Le réseau complet des co-auteurs internes et externes à l'institut

L'analyse peut encore être affinée pour révéler par exemple la carte institutionnelle de ces collaborations, ou encore la part d'entre elles associée à l'obtention de subsides. La démarche permet ainsi d'éclairer la réalité des collaborations scientifiques, en allant au-delà du score bibliométrique qu'il est possible en la matière de calculer. La question des langues dans lesquelles publient les membres d'un institut ouvre à diverses analyses possibles (graphe 3). Il serait intéressant de montrer aussi le lien entre les langues de publication, les réseaux de collaboration et les parcours des chercheurs.



Grappe 3: Les chercheurs d'un institut et les langues des publications

Un dernier exemple révèle les liens plus ou moins forts qu'entretiennent les publications avec les différents champs d'action d'un institut (graphe 4). Ce graphe permet d'identifier non seulement l'importance relative de chacune des dimensions considérées, mais aussi leur degré d'interdépendance que révèlent les co-occurrences pour une même publication (liens rouges sur le graphe).



Graphe 4: L'importance relative des champs d'action dont témoignent les publications et communications

Un outil à disposition des instituts

L'instrument d'analyse que nous développons dans ce projet s'adresse en premier lieu aux instituts intéressés à mettre en valeur leurs contributions scientifiques. L'instrument proposera un socle commun de descripteurs et d'indicateurs permettant des comparaisons entre instituts et offrira aussi la possibilité d'élaborer des descripteurs ad hoc pour éclairer les orientations de recherche spécifiques à chaque institut. Dans chaque cas, les chercheurs seront appelés à vérifier la pertinence des descripteurs utilisés et à contribuer à l'interprétation des cartes obtenues. Cette démarche offrira ainsi aux unités de recherche la possibilité d'utiliser cet outil, de participer à son adaptation et à son développement.

Qualitätskriterien für die Forschung in den Geisteswissenschaften – Eine Explorationsstudie

Sven E. Hug und Michael Ochsner, Sozialpsychologie und Hochschulforschung, Universität Zürich

Die aus den Natur- und Lebenswissenschaften stammenden bibliometrischen Verfahren eignen sich für die Geisteswissenschaften nur sehr bedingt, um Forschungsleistungen sichtbar zu machen oder zu messen. Deshalb fördert die Rektorenkonferenz der Schweizer Universitäten (CRUS) die Initiative «Entwicklung und Erprobung von Qualitätskriterien für die Forschung in den Geisteswissenschaften am Beispiel der Literaturwissenschaften und der Kunstgeschichte».

Die Initiative verfolgt das Ziel, Qualitätskriterien und Indikatoren zu entwickeln, die die Forschungsleistungen in ausgewählten geisteswissenschaftlichen Disziplinen adäquat abbilden und die von den Forschenden akzeptiert werden. Die Initiative, die an den Universitäten Basel und Zürich durchgeführt wird, ist in zwei Phasen gegliedert. In der Explorationsphase wurden mit Hilfe von Interviews, einer Literaturanalyse sowie einer vertieften Studie von etablierten Evaluationspraktiken adäquate Qualitätskriterien für die Englische und Deutsche Literaturwissenschaft sowie für die Kunstgeschichte/-wissenschaft ermittelt. In der zurzeit laufenden Validierungsphase werden diese Qualitätskriterien den Forschenden zur Modifikation, Ergänzung und Bewertung vorgelegt, um von den drei Fächern akzeptierte Qualitätskriterien zu ermitteln. Dies erfolgt im Rahmen einer international-europäischen Befragung mit der Delphi-Methode.

Erster Katalog von Qualitätskriterien

Die in der Explorationsphase gewonnenen Erkenntnisse wurden in einen ersten Katalog von Qualitätskriterien übersetzt; die 17 Kriterien wurden wie folgt benannt:

- Anerkennung
- Anschlussfähigkeit
- Begeisterung, intrinsische Motivation
- Bildungspotenzial
- Gelehrsamkeit, Scholarship, Belesenheit
- Gesellschaftsbezug

- Innovation, Erweiterung des Erkenntnisraums
- Inspiration
- Kontinuität
- Offenheit, Integration
- Originalität
- Pflege des kulturellen Gedächtnisses
- Reflexion, Kritik
- Selbststeuerung, Unabhängigkeit
- Vielfalt
- wissenschaftlicher Austausch
- Wissenschaftlichkeit

International-europäische Befragung von Professorinnen und Professoren

Dieser vorläufige Kriterienkatalog stellt die Grundlage für eine international-europäische Befragung mit der Delphi-Methode dar. Als Delphi-Methode bezeichnet man eine mehrstufige, schriftliche Befragung von Experten mit anonymer Rückmeldung nach jeder Befragungsrunde. Im Projekt sind drei Befragungsrunden vorgesehen. Die erste Runde diente der Modifikation und Ergänzung des ersten Kriterienkatalogs und wurde im August 2010 abgeschlossen. Um die Qualitätskriterien international abzustützen, wurden nicht nur ProfessorInnen aus der Schweiz angeschrieben (je 30 pro Disziplin), sondern auch ProfessorInnen von Universitäten, die Mitglied der League of European Research Universities (LERU) sind (ebenfalls je 30 pro Disziplin). Von den insgesamt 181 verschickten Fragebogen wurden 50 retourniert (Rücklaufquote 28%). Für jedes Qualitätskriterium war im Fragebogen eine Seite reserviert.

Dabei setzt sich ein Qualitätskriterium aus drei Teilen zusammen, nämlich aus der Benennung, aus den Aspekten, die das Qualitätskriterium spezifizieren und illustrieren, und aus den Indikatoren, die angeben, woran sich die Aspekte eines Qualitätskriteriums konkret erkennen, festmachen oder messen lassen. Aus den Antworten der ersten Befragungsrunde wurde ein zweiter, modifizierter Kriterienkatalog zusammengestellt, der im ersten Quartal 2011 durch eine breite Fachöffentlichkeit (Schweiz: alle ProfessorInnen der drei Fächer, d.h. ca. 130; LERU: ca. 450 ProfessorInnen) bewertet werden soll. Es ist angedacht, eine dritte Runde durchzuführen, in welcher der Kriterienkatalog einer erneuten Bewertung durch die gleichen Teilnehmenden unterworfen wird, um die Stabilität der Meinungen über die Qualitätskriterien und den Grad des Konsenses festzustellen.

Grundlage für Dialog über Qualitätskriterien

Der in der Englischen Literaturwissenschaft, der Deutschen Literaturwissenschaft und der Kunstgeschichte/wissenschaft entwickelte Kriterienkatalog kann auch für andere geisteswissenschaftliche Disziplinen die Grundlage für einen Dialog über Kriterien der Forschungsqualität bilden.

Weitere Informationen

Link

<http://www.psh.ethz.ch/crus>

Evaluieren in der Kommunikations- und Medienwissenschaft – Warum Aktivitäts-Profile die bessere Alternative sind

Das CRUS-Projekt «Measuring Research Output in Communication Sciences between international benchmarks, cultural differences and social relevance»

*Diana Ingenhoff, Organisationskommunikation und Kommunikationsmanagement, Universität Fribourg
Benedetto Lepori, Carole Probst, Centre for Organisational Research – Management and Performance of Research and Higher Education Institutions, Universität Lugano*

44

Das Thema Evaluation und Messung von Leistungoutput hat in nahezu allen wissenschaftlichen Disziplinen Einzug gehalten. Während es in einigen Bereichen wie in den Naturwissenschaften fasst ausnahmslos um die Wahrung der englischsprachigen «Journalpublikationen» geht, die in internationalen Publikations- und Zitationsdatenbanken gerankt sind, und die Exzellenz anhand dieses eindimensionalen Indikators gemessen wird, ist dies für andere Disziplinen, insbesondere die Geistes- und Sozialwissenschaften, sehr umstritten. Dies liegt insbesondere darin begründet, dass die Gepflogenheiten in Bezug auf wissenschaftliche Publikationen hier stark differieren, sowohl hinsichtlich der Sprache als auch des Formats. Neben Artikeln in Fachzeitschriften haben z.B. auch häufig Buchkapitel, Herausgeberschaften oder Monographien einen hohen Stellenwert.

Die Kommunikations- und Medienwissenschaft ist ein Fach, welches geprägt ist durch eine hohe Diversität in Bezug auf theoretische Zugänge und historische Entwicklungen, die je nach Kultur- und Sprachregion unterschiedliche Ausprägungen aufweist. So lassen sich sozialwissenschaftlich, kulturwissenschaftlich, geisteswissenschaftlich und wirtschaftswissenschaftlich orientierte Programme in der Schweiz finden, deren Forschende sich häufig innerhalb der jeweils eigenen Sprachregion orientieren und die sich je nach Ausrichtung interdisziplinär oder transdisziplinär verankern. Als bei den Studierenden äusserst beliebtes Fach, welches in der Schweiz meist ohne Zulassungsbeschränkung studiert werden kann, ist es gekennzeichnet durch begrenzte Ressourcen, einen hohen Betreuungsaufwand und eine grosse Vielfalt im Hinblick auf Methoden, inhaltliche Schwerpunkte und Sprache.

Vor diesem Hintergrund ist es eine grosse Herausforderung, einheitliche Standards zur Messung von akademischer Exzellenz zu definieren. Neben dem Output in Form von wissenschaftlichen Publikationen muss auch der Output in der Aus- und Weiterbildung

sowie Betreuung der zahlreichen Studierenden berücksichtigt werden. Zu einem ganzheitlichen Bild gehören zudem ebenfalls z.B. die Relevanz der Zusammenarbeit mit universitären und nicht universitären Partnern innerhalb von Forschungs- und Beratungsprojekten und der damit verbundene Transfer in die Gesellschaft.

Aktivitäts-Profile gewinnen an Bedeutung

Um diese vielfältigen kontextuellen und fachlichen Unterschiede zu berücksichtigen, gewinnen Aktivitäts-Profile immer mehr an Bedeutung, die für einzelne Forschungseinheiten (bestehend aus dem wissenschaftlichen Forschungsteam) erstellt werden. Hierbei findet einerseits der Gedanke stärker Berücksichtigung, dass sowohl Forschung als auch Lehre eine teamorientierte Arbeit darstellen und andererseits der Output in verschiedenen Kontexten unter Berücksichtigung der Rahmenbedingungen integriert werden sollte. In einem Profil können nun die verschiedenen Dimensionen dargestellt und vergleichbar gemacht werden, so zum Beispiel von den Semesterwochenstunden und betreuten Qualifikationsarbeiten in der Lehre (BA, MA, Weiterbildung), der Ausbildung von Forschenden, der wissenschaftlichen Gemeinschaft (z.B. in Form von Publikationen, Konferenzbeiträgen, Auszeichnungen), bis hin zum Transfer in die Gesellschaft und verschiedene Organisationen. Jede Dimension wird dabei über verschiedene Daten erhoben, zu Indikatoren verdichtet und kann dann auf der Ebene der Dimension aggregiert oder spezifischer analysiert werden.

Mehrdimensionale Ansätze für ein klares Profil

Für die Kommunikations- und Medienwissenschaft ist dieser mehrdimensionale Ansatz daher geeigneter als ein Vergleich über einen eindimensionalen Indikator. Zur Erstellung von Profilen werden die Indikatoren zunächst an die Grösse der Forschungseinheit angepasst und dann normalisiert, um eine Vergleichbarkeit herstellen zu können. Anschliessend können die Indikato-

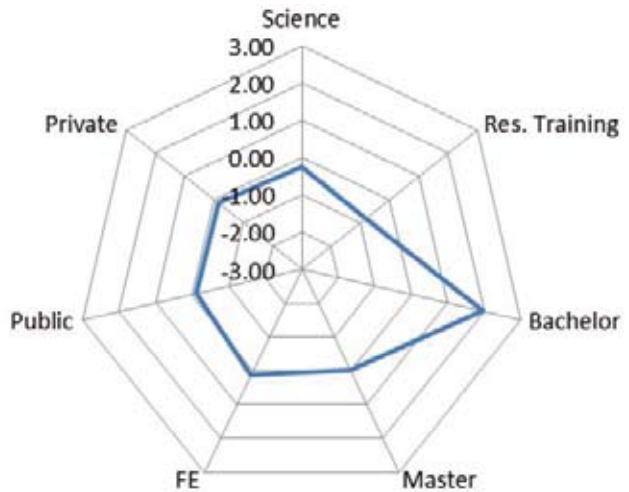
ren durch Mittelwert- oder Medianberechnung in der jeweiligen Dimension aggregiert werden. Grafiken 1 und 2 zeigen beispielhaft die ersten Resultate für ausgewählte Dimensionen zweier Forschungseinheiten, aus denen sich die unterschiedliche Profilierung sehr deutlich zeigt. So liegt u.a. bei der ersten Forschungseinheit ein klarer Schwerpunkt in der BA-Ausbildung, während die zweite Forschungseinheit u.a. auf Output für die wissenschaftliche Gemeinschaft fokussiert («Science») und in der Doktorandenausbildung stark engagiert ist («Res. Training»).

Zusätzlich zum gemessenen Profil wird die Leitung der jeweiligen Forschungseinheiten in semistrukturierten Interviews auch nach der Selbsteinschätzung und der zukünftigen Entwicklung gefragt, sodass die Forschungseinheiten im Analysebericht eine wertvolle Ergänzung erhalten, inwiefern sich ihre Selbsteinschätzung mit dem Status quo deckt. In einer späteren Folgerhebung könnten so Rückschlüsse auf mögliche Unterschiede in der gewünschten/prognostizierten im Vergleich zur tatsächlichen Entwicklung gegeben werden.

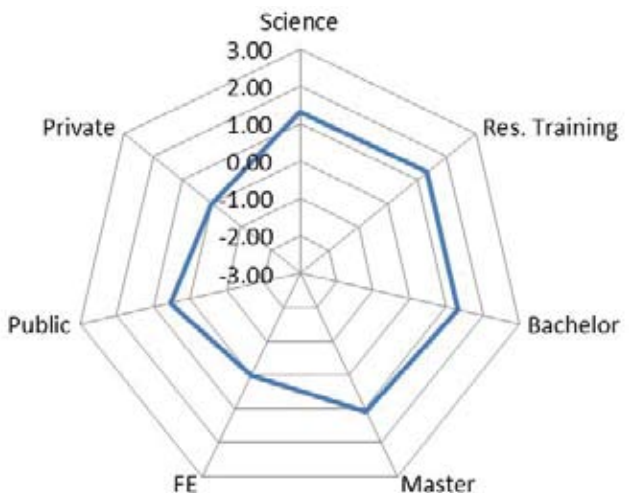
Kein perfektes Instrument, aber ein ganzheitliches Bild

Auch wenn es kein perfektes Instrument zur Messung von Forschung geben wird, zeigt unsere Studie insgesamt, dass mittels mehrdimensionaler Indikatoren und aggregierter Profile ein ganzheitlicheres Bild über verschiedene Output-Arten entsteht, welches nicht nur den Status quo abbildet, sondern auch den Kontext mit berücksichtigt und mögliche Ursachen für das Ergebnis aufzeigt (wie z.B. starke Auslastungen in einzelnen Bereichen wie der Lehre). Um valide und reliable Indikatoren zu ermitteln, die auch für die Forschungsgemeinschaft relevant sind, wurden die jeweiligen Entwicklungsschritte durch eine Expertengruppe aus der Schweizerischen Gesellschaft für Medien- und Kommunikationswissenschaft begleitet sowie weitere internationale Experten aus dem Bereich der Indikatorenbildung involviert. Dieser Zugang ermöglicht, die Ergebnisse in der Fach-Community abzusichern und unterschiedlich strukturierte Forschungseinheiten in ihren jeweiligen Besonderheiten zu betrachten und im jeweiligen Kontext zu vergleichen. Damit ist ein erster, wichtiger Schritt gemacht, die Vielfalt in der Diskussion um akademischen Output zu berücksichtigen.

Grafik 1



Grafik 2



Wie «gut» ist juristische Forschung in der Schweiz?

Andreas Lienhard, Qualitätsbeauftragter, und Fabian Amschwand, Qualitätsassistent, Rechtswissenschaftliche Fakultät, Universität Bern.

46

Ist die Qualität rechtswissenschaftlicher Forschungsleistungen messbar? Forschende haben an einer Fachtagung der Universität Bern die Notwendigkeiten und Methoden der Forschungsevaluation in den Rechtswissenschaften diskutiert.

Vorweg: Die im Titel gestellte Frage lässt sich heute nicht abschliessend beantworten. Der Grund dafür sind in erster Linie fehlende einheitliche, wissenschaftlich abgestützte Evaluationsverfahren. So finden juristische Publikationen kaum Eingang in die internationalen bibliometrischen Datenbanken, wie beispielsweise das von Thomson Reuters geführte «Web of Science». Wie weit ist dies von Nachteil? Bibliometrische Analysen, Peer Reviews und weitere Evaluationsverfahren sind betreffend ihre Aussagekraft und ihren Anreizwirkungen umstritten und verursachen administrativen Aufwand. Dieser Grundtenor dominierte zumindest an einer Fachtagung der schweizerischen Rechtsfakultäten. Forschende und weitere interessierte Kreise diskutierten am 25. Juni 2010 im Rahmen eines Projektes der Rektorenkonferenz der Schweizer Universitäten (CRUS) Fragen der Forschungsevaluation. Der Tagungsbericht dieser Fachtagung wurde unlängst publiziert.¹

Akzeptierte Notwendigkeit

Trotz aller Zweifel an der Messbarkeit der Qualität juristischer Forschungsleistungen anerkennen die Forschenden das Interesse an deren Sichtbarmachung: Evaluationen sind ein wichtiges Element der Rechenschaftsablage gegenüber Politik und Öffentlichkeit. Die Ergebnisse ermöglichen ausserdem die Reflexion eigener Leistungen. Und auch Justizbehörden und Rechtsanwendende des öffentlichen und privaten Sektors wollen sich ein Bild über die Qualität der rechtswissenschaftlichen Publikationen machen können, auf welche sie sich stützen. Inwiefern Evaluationen zur Stärkung der Wettbewerbsposition notwendig sind, wurde an der Tagung

kontrovers diskutiert. Der Kampf um Drittmittel steht in den wenig kostenintensiven Rechtswissenschaften zumindest in der Schweiz nicht im Vordergrund. Der Wettbewerb um renommierte Professorinnen und Professoren sowie um talentierten wissenschaftlichen Nachwuchs und um Studierende schon eher.

Keine Musterlösung

Die mehrheitliche Publikation in Amtssprachen, das Fehlen von «rated journals», eine starke Verbreitung von Monographien, ein inhaltlich und regional stark segmentiertes Forschungsgebiet sowie die daraus resultierende grosse Zahl von Zeitschriftentiteln mit kleinen Auflagen verunmöglichen eine unbesehene Übernahme von Verfahren aus anderen Wissenschaftsbereichen. Die Fachbibliographien, Bibliothekskataloge und verschiedenen Forschungsdatenbanken der Universitäten bilden zudem den Forschungsoutput für sich allein nur unvollständig ab. Damit fehlt in den Rechtswissenschaften die Basis für aussagekräftige bibliometrische Analysen.

Eigener Weg

Notwendig ist somit nebst der Schaffung einer anschlussfähigen bibliographischen Datenbank die Entwicklung geeigneter Qualitätskriterien und -indikatoren. Diese haben sich an konkreten Fragestellungen zu orientieren. Ein Erfordernis, dem bisher nach Ansicht vieler zu wenig Rechnung getragen wurde. Die Verantwortung für die Entwicklung tauglicher, die Besonderheiten juristischer Forschung berücksichtigender Instrumente kann dabei nicht den Evaluationsexperten überlassen werden. Die Teilnahme der juristischen Forschungsgemeinschaft am entsprechenden akademischen Diskurs ist eine unabdingbare Voraussetzung und liegt im Hinblick auf die Stellung und das Ansehen der Rechtswissenschaften in Gesellschaft, Politik und Wissenschaft in ihrem ureigenen Interesse.

¹ www.forschungsevaluation-rechtsfakultaeten.unibe.ch

Opening the black box of evaluation: How quality is recognized by peer review panels

Michèle Lamont, *Department of Sociology and African and African-American Studies, Harvard University*
Katri Huutoniemi, *Department of Social Research, University of Helsinki*

In two recent contributions to the study of peer review we went beyond stating the obvious that peer review produces valid judgments. We draw on in-depth analyses of five fellowship competitions in the United States, and of four grant panels organized by the Academy of Finland. We analyze and compare the intersubjective understandings academic experts create and maintain in making collective judgments on research quality. More specifically, we analyze the social conditions that lead panelists to an understanding of their choices as fair and legitimate, and to a belief that they are able to identify the best and less good proposals.

Our studies contest the common notion that one can separate cognitive from non-cognitive aspects of evaluation, as we describe the evaluative process as deeply interactional, emotional, and cognitive, and as mobilizing the self-concept of evaluators as much as their expertise. Studies of the internal functioning of peer review have revealed various “intrinsic biases” in peer review like “cognitive particularism” (Travis and Collins 1991), “favoritism for the familiar” (Porter and Rossini 1985), or “peer bias” (Chubin and Hackett 1990; Fuller 2002). These effects show that peer review is not a socially disembodied quality-assessing process in which a set of objective criteria is applied consistently by various reviewers. In fact, the particular cognitive and professional lenses through which evaluators understand proposals necessarily shape evaluation. It is in this context that the informal rules peer reviewers follow become important, as are the lenses through which they understand proposals and the emotions they invest in particular topics and research styles. Thus, instead of contrasting “biased” and “unbiased” evaluation, we aim to capture how evaluation unfolds, as it is carried off and understood by emotional, cognitive, and social beings who necessarily interact with the world through specific frames, narratives, and conventions, but who nevertheless develop expert views concerning what de-

finer legitimate and illegitimate assessments, as well as excellent and less stellar research.

In the two studies, we interviewed academic professionals serving on peer review panels that evaluate fellowship or grant proposals. During the interviews, panelists were asked to describe the arguments they made about a range of proposals, to contrast their arguments with those of other panelists, and to explain what happened in each case. Throughout the interviews, we asked panelists to put themselves in the role of privileged informer and to explain to us how “it” works. They were encouraged to take on the role of the native describing to the observer the rules of the universe in which they operate. We also had access to the preliminary evaluations produced before panel deliberations by individual panelists and to the list of awards given.

Pragmatic fairness is produced by informal rules

“How Professors Think” came out more than a year ago and has been debated within various academic communities, as it takes on several aspects of the evaluation in multidisciplinary panels in the social sciences and humanities. It is based on an analysis of twelve funding panels organized by important national funding competitions: those of the Social Science Research Council, the American Council for Learned Societies, the Woodrow Wilson Fellowship Foundation, a Society of Fellow at an Ivy League university and an important social science foundation in the social sciences. It draws on 81 interviews with panelists and program officers, as well as on observation of three panels.

A first substantive chapter describes how panels are organized. A second one concerns the evaluative culture of various disciplines, ranging from philosophy to literary studies, history, political science, and economics. A third chapter considers how multidisciplinary panels reach consensus despite variations in disciplinary evaluative cultures. This is followed by two chapters that focus on criteria of evaluation. One analyzes the formal criteria of evaluation provided by the funding organization to panelists (originality, significance, feasibility, etc.) as well as informal criteria (elegance, display of cultural capital, fit between theory and data, etc.). The following chapter considers how cognitive criteria are meshed with extra-cognitive ones (having to do with diversity and interdisciplinarity). We discover that institutional and disciplinary diversity loom much larger than gender and racial diversity in decision making. A concluding chapter considers the implications of the study of evaluation cultures across national contexts, including in Europe.

The book is concerned not only with disciplinary compromise, but also with the pragmatic rules that panelists say they abide by, which lead them to believe that the process is fair (this belief is shared by the vast majority of academics interviewed). “How Professors Think” details a range of rules, which include for instance the notion that one should defer to expertise, and that methodological pluralism should be respected.

Rules vary across evaluation settings

In her forthcoming article with Huutoniemi, Lamont explores whether these customary rules apply across contexts, and how they vary with how panels are set up. Thus, “Comparing Customary Rules of Fairness” is based on a dialogue between “How Professors Think” and a parallel study conducted by Huutoniemi of the four panels organized by the Academy of Finland. These panels concern: Social Sciences; Environment and Society; Environmental Sciences; and Environmental Ecology. Unlike Lamont’s study, this analysis was explicitly concerned with the effects of the mix of expertise on panels on how customary rules were enacted. The idea was to compare panels with varying degrees of specialization (unidisciplinary – multidisciplinary panels) and with different kinds of expertise (specialist experts – generalists). However, in the course of comparing results from the two studies, other points of comparison beyond expert composition emerged – whether panelists “rate” or “rank” proposals, have an advisory or decisional role, come from the social sciences and humanities fields or from more scientific fields, etc. Our exploratory analysis points to some important similarities and differences in the internal dynamics of evaluative practices that have gone unnoticed to date and that shed light on how evaluative settings enable and constrain various types of evaluative conventions.

Among the most salient customary rules of evaluation, deferring to expertise and respecting disciplinary sovereignty manifest themselves differently based on the degree of specialization of panels: we find that there is less deference in unidisciplinary panels where the expertise of panelists more often overlap. Overlapping expertise makes it more difficult for any one panelist to convince others of the value of a proposal when opinions differ; unlike in multidisciplinary panels, insisting on sovereignty would result in intense conflict for scientific authority. There is also less respect of disciplinary sovereignty in panels composed of generalists rather than experts specialized in particular disciplines, and panels concerned with topics such as Environment and Society that are of interest to wider audiences. In such panels, we find more explicit reference to general arguments and to the role of intuition in grounding decision-making.

While there is a rule against the conspicuous display of alliances across all panels, strategic voting and so-called “horse-trading” appear to be less frequent in panels that “rate” as opposed to “rank” proposals, and in those that have an advisory as opposed to a decisional role. The evaluative technique imposed by the funding agency thus influences the behavior of panelists. Moreover, the customary rules of methodological pluralism and cognitive contextualism are more salient in the humanities and social science panels than they are in the pure and applied science panels, where disciplinary identities may be unified around the notion of scientific consensus, including the definition of shared indicators of quality. Finally, a concern for the use of consistent criteria and the bracketing of idiosyncratic taste is more salient in the sciences than in the social sciences and humanities, due in part to the fact that in the latter disciplines evaluators may be more aware of the role played by (inter)subjectivity in the evaluation

process. While the analogy of democratic deliberation appears to describe well the work of the social sciences and humanities panels, the science panels may be best described as functioning as a court of justice, where panel members present a case to a jury.

Conclusion: Practices matter

The customary rules of fairness are part of “epistemic cultures” and essential to the process of collective attribution of significance. In this context, considering reasons offered for disagreement, how those are negotiated, as well as how panelists interpret agreement is crucial to capture fairness as a collective accomplishment. Together, these studies demonstrate the necessity for more comparative studies of evaluative processes and evaluative culture. This remains a largely unexplored but promising aspect of the field of higher education, especially in a context where European research organizations and universities aim to standardize evaluative practices. We look forward to interacting with colleagues as this research area develops.

This article was first published in Q&A magazine, nr. 2, september 2010, pp. 7-10

49

- *
 - monograph “How Professors Think” by Michèle Lamont (Harvard University Press 2009)
 - article “Comparing Customary Rules of Fairness: Evidence of Evaluative Practices in Peer Review Panels”, by Michèle Lamont and Katri Huutoniemi (forthcoming in an edited volume “Social Science in the Making”, University of Chicago Press)

Qualitäts- und Leistungsmessung in den Geisteswissenschaften – Eine Pilotstudie des deutschen Wissenschaftsrats

Christian Mair (unter Mitarbeit von Elke Lütkemeier und Veronika Khlavna), Englisch Seminar, Albert-Ludwigs-Universität Freiburg; Mitglied des deutschen Wissenschaftsrats

50

Der deutsche Wissenschaftsrat (WR) erarbeitete Ende 2004 Empfehlungen für ein mehrdimensionales Forschungsrating. Die Entwicklung angemessener Kriterien und Bewertungsverfahren für die Geisteswissenschaften stellte angesichts der Eigenständigkeit und Heterogenität der beteiligten Fachgebiete eine besondere Herausforderung dar. Doch nun liegen erste Ergebnisse vor: Die Bewertung geisteswissenschaftlicher Forschung soll anhand mehrerer Kriterien erfolgen: «Forschungsqualität», «Forschungsermöglichung» und «Transfer an ausserwissenschaftliche Adressaten».

Forschungsrankings sind trotz ihrer weiten Verbreitung und ihres öffentlichen Echos methodisch oft unangereift und daher als Steuerungsgrundlage für eine verantwortungsvolle Forschungsplanung nicht geeignet. Deshalb hat der deutsche Wissenschaftsrat (WR) Ende 2004 Empfehlungen für ein Forschungsrating erarbeitet, das (1) wissenschaftsgeleitet, (2) disziplinspezifisch und (3) in der Lage ist, Leistungen entlang mehrerer Parameter getrennt auszuweisen (z.B. Forschung, Nachwuchsförderung, Transfer in Wirtschaft und Gesellschaft). An die Stelle der (nur scheinbar präzisen) Listenplätze in Rankings treten somit als Ergebnisse differenzierte Leistungsprofile, die über die Institutionen hinweg verglichen werden können.

Unerwartet konfliktreiches «Forschungsrating Geschichte»

Bislang sind zwei Pilotstudien (Chemie, Soziologie) weitgehend erfolgreich beendet worden, eine dritte (Elektro- und Informationstechnik) nähert sich dem Abschluss. Der Versuch, die Methode auf das Fach Geschichte anzuwenden, führte jedoch zu einer Reihe von Komplikationen, die ihre Ursache zumindest zum Teil in der spezifischen Forschungskultur der Geisteswissenschaften haben und deshalb vielleicht auch für ein Lesepublikum ausserhalb Deutschlands von Interesse sein könnten. Das von WR und Historikerverband zu-

erst einvernehmlich gesuchte Gespräch erwies sich – insbesondere auch nach teilweiser Verlagerung der Debatte in das Feuilleton – als unerwartet konfliktreich und endete mit dem Rückzug des Verbandes. Das Pilotprojekt «Forschungsrating Geschichte» war gescheitert, bevor es begonnen hatte. Der Schaden für die deutsche Geschichtswissenschaft mag sich, will man den Argumenten des Verbandes folgen, in Grenzen halten, da das Fach nach eigener Aussage über funktionierende Mechanismen der Qualitätsprüfung und Reputationszuweisung verfüge. Die Herausforderung für den WR allerdings war, dem Eindruck vorzubeugen, dass ein – gemessen an Studierendenzahlen und öffentlicher Wirkung – wesentlicher Teil des Wissenschaftssystems im geplanten Rating nicht in den Griff zu bekommen sein würde.

Spezielle Bedürfnisse in den Geisteswissenschaften

In der sachlich und kommunikativ-atmosphärisch etwas verfahrenen Situation, die nach dem «Historikerstreit» entstanden war, verbot es sich, einfach an die nächste geisteswissenschaftliche Disziplin heranzutreten. Stattdessen wurde eine Arbeitsgruppe eingesetzt, der Vertreterinnen und Vertreter fast des gesamten geisteswissenschaftlichen Spektrums (Sprach- und Literaturwissenschaften, Philosophie, Kunstwissenschaft), insbesondere auch Vertreter der sogenannten «Kleinen Fächer» (und sogar vier Geschichtswissenschaftler) angehörten und die vor einem weiteren Pilotversuch allgemeine Fragen der Messung und des Vergleichs von Forschungsleistung in den Geisteswissenschaften bearbeiten sollte. Die Arbeitsgruppe hat ihre Ergebnisse im Sommer 2010 vorgelegt und damit die Basis geschaffen, auf der sich mittlerweile die einschlägigen Fachverbände zur Teilnahme an der Pilotstudie «Forschungsrating Anglistik/Amerikanistik» bereit erklärt haben.

Ausgehend von der Annahme, dass die universitären und die ausseruniversitären Einrichtungen neben

Forschung im engeren Sinne auch weitere forschungsbezogene Aufgaben wahrnehmen, soll die Bewertung geisteswissenschaftlicher (im konkreten Fall anglistischer) Forschung anhand mehrerer Kriterien erfolgen: «Forschungsqualität», «Forschungsermöglichung» und «Transfer an ausserwissenschaftliche Adressaten».

Forschungsqualität

Abweichend von den bisherigen Pilotstudien soll die Bewertung des Kriteriums «Forschungsqualität» hauptsächlich auf der Grundlage selektiver Lektüre ausgewählter Publikationen erfolgen (und nicht etwa quantifizierend durch Berechnung von Impact-Faktoren und Zitationshäufigkeiten). Informationen über die Quantität der Publikationen sowie über die etwa durch Preise und Auszeichnungen belegte Reputation treten hinzu. Es wird der Bewertungsgruppe «Anglistik» freistehen, diese Datengrundlage um weitere, fachspezifische Aspekte anzureichern – etwa in Hinblick auf die Gewichtung englischsprachiger gegenüber deutschsprachigen Publikationen oder die Bewertung von Einzel-Autorschaft gegenüber Ko-Publikation.

Forschungsermöglichung

Schon die noch nicht etablierte Bezeichnung «Forschungsermöglichung» für die zweite Bewertungsdimension weist auf eine Umbruchssituation in Teilen der Geisteswissenschaften hin. Während für manche der Einzelforscher die umfangreiche thematische Monographie die reputierlichste Form der Publikation und nach wie vor die Regel ist, gewinnen in anderen Gebieten das Forschungsteam und die Arbeit auf der Basis von teuren apparativen und digitalen Infrastrukturen an Bedeutung. Die Bewertungsdimension «Forschungsermöglichung» erlaubt es, Erfolge bei der Einwerbung von Drittmitteln, der Förderung des wissenschaftlichen Nachwuchses oder dem Aufbau von Infrastrukturen zu honorieren, ohne all diese forschungsrelevanten Verdienste direkt in die Bewertung der Forschungs-

qualität einzurechnen, denn nach wie vor ist höchste Forschungsqualität in den Geisteswissenschaften auch ohne Drittmittel möglich.

Transfer

Vergleichsweise einfach gestaltet sich hingegen die Anpassung der Dimension «Transfer». Je nach Fach liesse sich an Ausstellungen, öffentliche Vorträge, Beiträge in den Medien oder Lehrerfortbildung denken.

Die Herausforderungen annehmen

Die Entwicklung angemessener Kriterien und Bewertungsverfahren für die Geisteswissenschaften stellt angesichts der Eigenständigkeit und Heterogenität der beteiligten Fachgebiete eine besondere Herausforderung dar. In diesem Sinne ist dem Pilotprojekt «Forschungsrating Anglistik/Amerikanistik» zu wünschen, dass es anders als sein Vorgänger zu einem erfolgreichen Abschluss kommt. Als doppelt Betroffener – der Verfasser dieser Zeilen ist Anglist und Mitglied des WR – hoffe ich, dass die breite Mehrheit der Kolleginnen und Kollegen im Lauf der Übung zumindest zur Überzeugung gelangt, dass, wenn schon «geratet» wird, die Verantwortung wenigstens in der Hand des Faches selbst liegen soll. Das Rating wird jedoch dann zum vollen Erfolg, wenn eine noch konstruktivere Einsicht Platz greift: Wichtig ist nicht primär, wie man selbst im Vergleich zur Nachbarprofessur am selben Standort abschneidet, sondern wie sich der gesamte Standort im nationalen Vergleich präsentiert. Vor allem auf dieser Grundlage könnte das Rating sinnvolle Forschungsplanung und Hochschulpolitik unterstützen.

Download

«Empfehlungen zur vergleichenden Forschungsbewertung in den Geisteswissenschaften»

<http://www.wissenschaftsrat.de/download/archiv/10039-10.pdf>

Yes we should; research assessment in the Humanities

Wiljan van den Akker, *Modern Poetry*, Dean of Humanities
Utrecht University

52

A shorter version of the “Humanities Attitude” towards research assessment could be: “We don’t want it, because we don’t have to, because we don’t need it, because we are not like the others, and therefore we don’t like it, and they shouldn’t force us, because they don’t know us, because they don’t understand us, because they don’t love us.” The image of the humanist working in solitude in the attic, is still alive and kicking.

The humanities have developed several defense mechanisms against research assessment in general. 1. The argument of intuition: the quality of our research is not measurable, not quantifiable. We know quality when we see it. It is easy to see that this argument is highly irrelevant. One could turn it around and say: this makes research assessment only easier. 2. The second mechanism is: the humanities are different from all the other forms of science. This is not true. The humanities are made of different disciplines and fields each with their own academic positions. Some are very familiar to empirical physics, like linguistics, some to social sciences, like some historical disciplines. Parts of philosophy claim the same domain as mathematics. 3. Thirdly: it is impossible to compare us to other parts of the body of knowledge. Not a strong argument, since it also holds true for what we call the sciences. Just think of the social sciences where the anthropological and the empirical approach are highly competing each other.

Courage to play the game

These defense mechanisms are not effective for today’s world and especially not for the future of the humanities. We cannot and should not insist on being “different” just to shy away from any form of research assessment. If we continue doing that, we will be the young sister or brother who is tolerated at the dining table, at the mercy of the food that the rest of the family thinks it can spare and always looked down upon.

In the near future, in a world that increasingly asks for justification of public funds, in a world where at the same time public money becomes scarcer and less amounts have to be distributed among more players, in a world where research funds are being concentrated and distributed on a highly competitive basis, we as humanists have to take the stand and declare that we are grown-ups and want to play the game.

Seven aspects of research assessments within the humanities

In discussing research assessment within the humanities, we are dealing with a complicated matter. I will name seven aspects.

(1) There are substantial differences in scientific practice between the several disciplines within the humanities and they will have consequences for the selection of quality indicators. There are areas where groups of scholars work together on a common project and therefore publish together in journals. An analysis of citations can be useful. In other areas individuals work on topics and therefore publish individually. An analysis of citations can be less useful. (2) Contrary to other fields of science, much of what humanists produce can have an effect in the long(er) run and will not be outdated within two or three years. (3) The goals and products of research are different in different areas of the humanities. Research in the humanities often has the societal intention or assignment to guard, disclose, save and interpret (inter)national heritage. If we don’t do it, who else will? The products of research will be the construction of databases and data collections, or exhibitions with catalogues, or excavations of archeological sites. Data collections, including books, are laboratories for the humanities. (4) The target group of the humanities is diverse. On the one hand peers, on the other hand a large, non-academic audience. One of the problems is to define this wider group, and to justify our relations with it. (5) The publication channels of the

humanities will vary. Books will be written next to articles in journals. (6) A highly controversial aspect is the language of our research. Often there is a nationalistic side to the discussion. Because of the linguistic nature of the field of study, there have to be journals in a language other than English. Tied to this is the more sentimental reasoning: a country has its own cultural heritage, and academia should honour the uniqueness of it, by allowing scholarly work to be written in the mother language. First it is a mistake to think that most scholarly work is written in English. It looks and sounds like English, but it is not. It is at best Scholarly English, like Latin was centuries ago. The language of the object of research has nothing to do with the language in which we scholarly communicate about it. The one reason for the relation between the language of the object and the language skills of the scholar, is the mastering of the skills to read the relevant documents. Nationalism is a killer in the world of science, also in the humanities. I am fully aware of the new nationalism that spreads its bad seeds all over Europe. (7) Finally there is the level of organization within the humanities. Or the lack of it. More and more organization becomes a substantial factor. Several fields could and should be well or better organized. The field of the discipline or subdiscipline,

the field of the managerial organization – departments, schools, research institutes – the field of the faculties of humanities within the university, the field of the national endowments of the humanities, the European field such as the European Science Foundation and the European Research Council. With regard to the humanities, some fields are highly successful and well organized at the same time, like archeology, linguistics and social-economic history. It is no coincidence that these disciplines are working with laboratories and large data collections. We should not automatically copy this for all the disciplines within the humanities, but we definitely need more research projects, more research institutes within the universities. We should stop telling the world that we are only *different*. The most important thing is: let's take the lead ourselves.

Conclusion

The recognition of the humanities as a distinct member of the body of academic knowledge, leads to the conclusion that humanists should take the steering wheel in developing adequate forms of research assessment themselves. If we leave it to others, the humanities will look like arms attached to the feet.

European Educational Research Quality Indicators (EERQI)

Ingrid Gogolin, Erziehungswissenschaft, Psychologie und Bewegungswissenschaft, Universität Hamburg

54

Weltweit verändern sich Strukturen und Steuerungsmechanismen öffentlich finanzierter Forschung. Zu den Ursachen dafür gehört die Beschwörung der «internationalen Wettbewerbsfähigkeit». Eine stets mitschwingende Metapher dabei ist «Qualität»: Die Verbesserung und Förderung der Forschungsqualität sei ein treibendes Motiv für die Einführung von Kontroll- bzw. Legitimationssystemen, auf deren Basis sich Forschung steuern, beurteilen und finanzieren lasse. Da liegt die Frage nahe, was unter Qualität verstanden und wie sie gemessen wird. Analysen zu dieser Frage standen am Ausgangspunkt für die Einwerbung des Forschungsprojekts «European Educational Research Quality Indicators (EERQI)» im 7. Forschungsrahmenprogramm der Europäischen Union (vgl. www.eerqi.eu).

Im Fokus der Analysen standen Verfahren, die für die Prüfung erziehungswissenschaftlicher Forschung eingesetzt werden, die in Europa stattfindet. Die Erziehungswissenschaft kann als prototypisch für weite Bereiche der «Social Sciences and Humanities» gelten. Das Fach vereint ein weites Spektrum – vom primär philosophisch-historischen Forschungszugriff in den Geisteswissenschaften bis zur Gewinnung und quantitativen Analyse von grossen Datenmengen. Es weist also Merkmale der Wissensproduktion auf, die auch in anderen geistes- und sozialwissenschaftlichen Disziplinen anzutreffen sind.

Bisherige Verfahren sind ungeeignet

Das Ergebnis der Auseinandersetzung mit der Frage, wie und woran Qualität der erziehungswissenschaftlichen Forschungsleistungen gemessen wird, lautete: Es handelt sich um nicht valide Verfahren, denn sie prüfen nicht, was sie zu prüfen vorgeben. Ein Beispiel ist die Qualitätsmessung über Zitations- und Journal-Rankings, wie sie weithin üblich sind. Ihr zentrales Kriterium ist das der «internationalen Sichtbarkeit»,

die sich ausdrücke durch Publikation in gut beleumundeten Fachzeitschriften und in der Zahl der Zitierungen dieser Publikation – so z.B. beim Social Science Citation Index. Die Dokumentation der darin vertretenen Zeitschriften zeigt (für das Jahr 2009 und das Fach Erziehungswissenschaft nach dem «Journal Citation Report»):

Von insgesamt 201 Zeitschriften erscheinen ca. 52% in US-amerikanischen Verlagen. Weitere 24% stammen aus britischen Verlagshäusern. Die nächst«grossen» Nationen sind die Niederlande (mit 4%) und Deutschland (mit 3%). Insgesamt sind 15 Nationen im Ranking vertreten. Eine etwas andere Betrachtungsweise nach der Frage der Sprachen, in denen die Journale erscheinen, ergibt: 89% der Publikationen verwenden die englische Sprache. Die nächst«grössten» Sprachen mit 2,5% bzw. je 2% sind Deutsch, Spanisch und Türkisch. Insgesamt sind 11 Sprachen im Index vertreten. Französisch gehört nicht dazu. Das Verfahren liefert also keine validen Informationen, denn die angestrebte internationale Relevanz der einbezogenen Publikationen kann nicht nachgewiesen werden.

Vom EERQI entwickelte Verfahren

Das Forschungsprojekt EERQI hat sich nun nicht vorgenommen, den wirtschaftlich mächtigen Lieferanten solcher Produkte Konkurrenz zu bieten. Vielmehr besteht sein Anliegen darin, prozessbegleitende Verfahrensweisen zu entwickeln und zu erproben, die den Vorgang einer Begutachtung unterstützen können. Entwickelt wird ein Rahmenwerk von Verfahren, welche die vom Menschen vollzogene Beurteilung von Qualität unterstützen können. Diese sollen auch dann angewendet werden können, wenn Forschungsarbeiten nicht in Zeitschriften, sondern in anderen Medien erscheinen, und wenn sie in anderen Sprachen als Englisch publiziert werden.

Zu den Verfahren und Produkten des Projekts gehören:

- eine Suchmaschine mit multilingualer Funktion (exemplarisch für die Sprachen Englisch, Deutsch, Französisch und Schwedisch), die imstande ist, erziehungswissenschaftliche Texte zu erkennen – also das Potenzial für die Prüfung von Qualität zu liefern;
- ein Verfahren der multilingualen elektronischen semantischen Analyse, das Leser bei der raschen Erfassung zentraler Aussagen eines Textes unterstützt;
- Verfahren der bibliometrischen Analyse, die eine Näherung an Qualitätsaussagen vermittelt der Prüfung äußerlicher Textmerkmale erlauben, und
- ein empirisch überprüftes Instrument für Peer-Review-Begutachtungen, das relevante Aussagen zur Qualität eines Textes anhand von drei generischen Indikatoren und ihrer Operationalisierung ermöglicht.

Zur Verdeutlichung der Möglichkeiten des Einsatzes und der Funktion dieser Verfahren wurden «Use case Szenarien» entwickelt: häufig vorkommende Konstellationen der Qualitätsprüfung und Hinweise darauf, welches Verfahren bzw. welche Kombination von Vorgehensweisen jeweils geeignet sind.

Das EERQI-Projekt beendet seine Arbeit an der Entwicklung eines prototypischen Instrumentariums zur Jahresmitte 2011; es stellt seine Ergebnisse bei einer Konferenz am 15. und 16. März 2011 in Brüssel vor.