

Fachgutachten, Panel oder Würfeln?

Wie über die **Förderwürdigkeit von Forschung** entschieden werden kann und soll, diskutierten die Teilnehmenden der Podiumsdiskussion der Jahrestagung des Netzwerks Forschungs- und Transfermanagement Mitte März an der Universität Potsdam. Protokoll: Benita von Behr

Armin Himmelrath: Wir reden heute über Forschungsbegutachtung und -bewertung. Welche Erfahrungen haben Sie damit in der Volkswagen-Stiftung? Was ist der spannendste Bereich, mit dem Sie in den letzten Jahren zu tun hatten?

Dr. Henrike Hartmann: Wir wissen alle darum, dass der Druck, Drittmittel einzuwerben, enorm gestiegen ist. Das führt dazu, dass die Antragszahlen steigen und die Nachteile von Peer-Review sichtbar werden. Wir müssen sehen, wie wir die Begutachtung so aufsetzen, dass sie dem Ziel gerecht wird. Ich denke, es ist enorm wichtig, dass wir die Begutachtungsverfahren immer wieder neu testen, ausprobieren und reflektieren, wo Schwierigkeiten liegen. Wenn Entscheidungen sehr schnell getroffen werden, etwa: Wo geht es um Fragen des Implicit Bias? Wo kommen Scheinkriterien hinein, die nicht wissenschaftlich sind, nur für ei-

ne Legitimation der Entscheidungsfindung? In einem Programm machen wir jetzt das Experiment, dass wir Peer-Review um ein Losverfahren ergänzen. Wenn wir neue Ansätze ausprobieren, ist uns eine Begleitforschung wichtig, mit der wir überprüfen, was wirklich funktioniert.

Himmelrath: Herr Lange, der Wissenschaftsrat hat ein Papier mit Ideen dazu verfasst, wie es gehen kann. Macht es Sie nervös, wenn Sie hören, dass die Volkswagen-Stiftung die Vergabe von Forschungsgeldern auslösen will?

Dr. Rainer Lange: Nein, nervös macht uns das überhaupt nicht. Im Gegenteil, wir haben das mit großem Interesse und auch großer Zustimmung zur Kenntnis genommen. Als wir das von Ihnen erwähnte Papier [Ergänzung der Red.: Wissenschaftsrat: Begutachtung-

Auf dem Podium

(v. l.) **Rainer Lange**, Leiter der Abteilung Forschung des Wissenschaftsrats; **Henrike Hartmann**, Leiterin der Abteilung Förderung und Mitglied der Geschäftsleitung der Volkswagen-Stiftung; **Stefan Hornbostel** von der Humboldt-Universität zu Berlin und Abteilungsleiter am Deutschen Zentrum für Hochschul- und Wissenschaftsforschung, **Klaus Wehrberger**, Gruppenleiter für Programm- und Infrastrukturförderung der Deutschen Forschungsgemeinschaft (DFG), **Armin Himmelrath**, Wissenschaftsjournalist und Moderator





gen im Wissenschaftssystem. Positionspapier, 2017, abrufbar unter: www.wissenschaftsrat.de/download/archiv/6680-17.pdf] vorbereitet haben, ging es uns erst einmal darum, einen gewissen empirischen Überblick zu schaffen, wo eigentlich überall begutachtet wird. Das ist unglaublich vielfältig. Es betrifft nicht nur die Forschungsförderung, sondern es werden Gutachten geschrieben über Doktorarbeiten und Publikationen, für Berufungsverfahren und wenn Menschen Reise-mittel haben möchten. Unsere Sorge war, dass die Verfahren möglicherweise alles zu sehr über einen Kamm scheren. In manchen Kontexten muss man sich erst einmal klarmachen: Welche Art von Fehler will ich eigentlich vermeiden? Will ich nur Verschwendung verhindern? Oder muss ich eher aufpassen, dass ich wirklich originelle Dinge nicht unterdrücke? Im letzteren Fall ist die Randomisierung möglicherweise ein guter Weg. Denn wenn nach einer ersten Qualitätsauswahl immer noch zu viel Förderungswürdiges zur Wahl steht und die Entscheidung in einem Gremium getroffen wird, ist das Risiko sehr hoch, dass sich die Beteiligten auf Mainstream-Projekte einigen. Weil die eben am ehesten Konsens sind.

Himmelrath: Was Sie beschreiben, ist ja eher eine Idealsituation – also, dass man in Ruhe reflektieren kann: Welche Fehler wollen wir vermeiden, was wollen wir fördern, wie wollen wir zum Beispiel Kreativität

ermöglichen, ohne wissenschaftliche Standards abzubauen. Herr Wehrberger, besteht im Alltag überhaupt die Chance, sich mit solchen Fragen intensiv auseinanderzusetzen?

Dr. Klaus Wehrberger: Die Deutsche Forschungsgemeinschaft (DFG) ist ein großer Nachfrager nach

”

Unsere Sorge war, dass die Verfahren möglicherweise alles zu sehr über einen Kamm scheren.

“

Gutachten. Vor einem Jahr haben wir in einer Analyse unsere Begutachtungsaktivität von 2008 und 2016 verglichen. Ein Ergebnis war, dass wir zwar mehr Gutachten einholen, aber nicht im gleichen Ausmaß, wie wir mehr Anträge bekommen. Wir holen also weniger Gutachten pro Antrag ein. Das ist schon mal eine gute Entwicklung, denke ich. Außerdem gehen wir mehr in die Breite, wir fragen mehr Personen um Gutachten. Es sind nicht immer die Gleichen, die ganz oft gutachten. Das muss sich weiterentwickeln. Wir wissen: Be-

gutachtung ist nicht der Stein der Weisen. Es ist nur das Beste, was derzeit „auf dem Markt“ ist, um unter Projektanträgen auszuwählen. Wir wissen, es ist subjektiv, es können Vorurteile einfließen, bewusst oder unbewusst. Wir wissen, es ist aufwendig und auch fehlerbehaftet. Es ist ein Blick in die Zukunft und damit ist immer das Risiko verbunden, Fehler zu machen. Das muss man bewusst eingehen. In der Summe ist es schon so, dass das Begutachtungssystem funktioniert, aber wir müssen ständig daran arbeiten.

Kein Verfahren passt für alles

Prof. Dr. Stefan Hornbostel: Wir haben seit mehr als 30 Jahren eine Debatte um Begutachtungsverfahren, das Peer-Review und dessen Güte. Seitdem wissen wir auch, dass der Zufallsanteil in diesen Verfahren rela-



Als Förderer hat man eine Verantwortung, an den Begutachungskriterien mitzuwirken.



tiv hoch ist. Wenn man in die jüngere Vergangenheit guckt, sieht man die Folgen des tatsächlich enorm angewachsenen Volumens an Drittmitteln. Wir haben seitdem auch eine Evaluationsbewegung. Nicht nur Forschungsanträge, sondern ganze Einrichtungen werden evaluiert, Personen werden begutachtet. In den letzten zehn Jahren gab es eine Fülle von Experimenten mit anderen Verfahren. Bisher kann man das schwer resümieren. Wir haben sehr unterschiedliche Begutachtungsgegenstände, -verfahren und Kriterien, die in Anschlag gebracht werden sollen. Deshalb gibt es kein One-Size-Fits-All-Verfahren.

Lange: Auch wenn sie nicht der Treiber ist, denke ich, dass die DFG in gewisser Hinsicht eine prägende Rolle in Deutschland hat. Denn sie übernimmt die Sozialisation von Gutachterinnen und Gutachtern. Die meisten Leute in Begutachtungsverfahren haben selbst schon einmal an einem DFG-Verfahren teilgenommen und orientieren sich an dem, was sie dort gelernt haben.

Das führt natürlich schon zu einer Tendenz, alles so zu machen wie bei der DFG.

Reflektion der Kriterien und Ziele

Hartmann: Ich denke auch, dass One-Size-Fits-All nicht passt. Deshalb ist es umso wichtiger, dass man bei Begutachtungen gemeinsam mit den Gutachtern zuerst auf eine Metaebene geht und fragt: Was tun wir hier eigentlich? Was sind in diesem speziellen Fall die Kriterien und worum geht es uns? Als Förderer hat man durchaus eine Verantwortung, an dieser Gestaltung mitzuwirken. Die Volkswagen-Stiftung hat das zum Beispiel für die Freigeist-Fellowships eingeführt. Am Abend bevor es losgeht, geben wir den Gutachtern ein ganz explizites Briefing. Wir verständigen uns auf einen Konsens, worum es geht. Das hat unserer Wahrnehmung nach enorm positive Einflüsse darauf, wie die eigentliche Begutachtung verläuft.

Wehrberger: Auch die DFG kennt solche Vorbesprechungen, in denen die Gutachter darauf hingewiesen werden, was die Kriterien sind und was eben nicht. Bei sehr hohem Wettbewerbsdruck ist es eine Möglichkeit, dass man ein zweistufiges Verfahren einführt, um es sowohl für die Antragsstellenden als auch für die Gutachter effizienter zu gestalten. Im ersten Schritt erfolgt eine Auswahl auf der Basis von kurzen Beschreibungen, in einem zweiten Schritt stellen die dabei Ausgewählten einen Antrag, um die Vorhaben in der Tiefe zu betrachten.

Wie finden neue Impulse Eingang ins System?

Himmelrath: Das Spannende aus meiner Sicht ist, was passiert, wenn Gutachterinnen und Gutachter eigene Impulse miteinbringen, weil sie vielleicht irgendwo anders sozialisiert wurden oder in einem anderen Wissenschaftssystem gearbeitet haben. Eines der bekanntesten Beispiele ist wahrscheinlich die Begutachtung für die erste Runde der Exzellenzinitiative, wo auf internationalen Druck hin das Thema Frauenförderung plötzlich eine Rolle spielte.

Wehrberger: Das hat die DFG ja dann auch aktiv genutzt und die Rolle dieses Faktors in den Kriterien gestärkt, auch in der Erwartung, dass die Gutachter aus dem Ausland Deutschland den Spiegel vorhalten und uns sagen, wo und wie wir noch besser werden können.



Die Podiumsdiskussion im Hörsaal der Universität Potsdam stieß auf reges Interesse.

Himmelrath: Herr Hornbostel, gibt es andere Beispiele, wo Veränderungsimpulse von außen kamen und ins Gutachtersystem hineingetragen wurden?

Hornbostel: Nach meinem Wissen ist es eher selten, dass die Gutachter selbst auf einmal Kriterien setzen. Die sind normalerweise im Verfahren vorgegeben. Aber es passiert, dass sehr unterschiedliche Begutachtungsstile zum Tragen kommen. Es gibt durchaus nationale Traditionen, in denen man entweder schärfer oder etwas lockerer und nachsichtiger ist. Auch die Art und Weise, wie man Kritik oder Lob formuliert, ist national sehr unterschiedlich. Das lässt sich schon in kurzer Distanz erleben. In Österreich zum Beispiel ist man sehr viel zurückhaltender – nicht weniger kritisch, aber die Kritik formuliert man völlig anders. Das Gleiche gilt auch für amerikanische Gutachter. Die Unterschiedlichkeit ist zwar erfrischend, aber manchmal auch problematisch, weil die Verständigung nicht klar ist. Wenn jemand seine Kritik sehr höflich und versteckt artikuliert, verstehen deutsche Kollegen das mitunter eher als positives Urteil.

Hartmann: Ähnliches gilt, wenn Gutachter aus unterschiedlichen Disziplinen an einem Tisch sitzen. Da

sind auf der einen Seite die Lebenswissenschaftler, die sehr euphorisch sein können, aber auch gnadenlos hart, und auf der anderen Seite die Geistes- und Gesellschaftswissenschaftler, die für sich in Anspruch nehmen, alles sehr differenziert zu sehen. Ich würde auch nicht sagen, dass Gutachter Kriterien einbringen, aber wir haben schon die Erfahrung gemacht, dass durch eine internationale Zusammensetzung und eine gemeinsame Reflektion, was man erreichen möchte, Impulse hineinkommen. Dazu ein Beispiel: Wir finden Wissenschaftskommunikation sehr wichtig und fordern deshalb seit längerer Zeit von den Antragstellenden, dass sie einen fiktiven Zeitungsartikel über ihr Projekt schreiben. Durch Gutachter kam der Impuls, dass das doch auch ein Video sein könnte. Das machen wir jetzt so und die Anzahl derer, die dies nutzen, steigt.

Wehrberger: Es ist, glaube ich, ziemlich umstritten, inwieweit das Show-Element einfließen sollte in die Frage, ob ein Projekt gut oder nicht so gut ist. Wir erleben es immer wieder, dass es sehr spannend wird, wenn verschiedene Fach-Communitys mit ihren unterschiedlichen Kulturen aufeinandertreffen, auch wenn internationale Gutachter dazukommen.



Foto: Nico Fischbein

Die Teilnehmenden fragten interessiert nach.

Die Frage nach dem Impact

Lange: In einem Idealzustand fände ich es schon schön, wenn man mit den Gutachterinnen und Gutachtern auch ein Stück gemeinsam an der Interpretation eines Kriterienrasters arbeiten könnte. Die Menschen, die daran teilnehmen, machen sich das Verfahren in ganz anderer Weise zu eigen. Ich möchte noch ein aktuelles Beispiel nennen, wo uns Impulse aus dem Ausland immer wieder begeben: Wenn wir uns mit Institutionen oder Infrastrukturen beschäftigen – das tun wir ja beim Wissenschaftsrat häufig –, werden wir von Gutachtern aus den Niederlanden oder aus Skandinavien in den letzten Jahren vermehrt nach dem gesellschaftlichen Impact gefragt: Wollt ihr den nicht auch bewerten? Das ist für die ein viel zentraleres Thema als in Deutschland. Das bringt uns schon zum Nachdenken,

weil es offenkundig in anderen Systemen inzwischen zu einem Kriterium geworden ist.

Wehrberger: Das beobachten wir auch – allerdings ebenso, dass dann meistens eine gewisse Erleichterung zu verspüren ist, wenn wir sagen, dass das bei uns kein Kriterium ist und es nur auf die wissenschaftliche Qualität ankommt. Es ist auch eine Aufgabe, die Gutachter aus dem Ausland mit dem hiesigen System vertraut zu machen. Wenn wir jemanden dann zum dritten Mal einladen, funktioniert es schon ganz reibungsfrei.

Hornbostel: Ich würde das etwas skeptischer sehen. Es ist in der Tat schwer, so etwas wie gesellschaftliche Wirkung zu messen und vernünftige Kriterien dafür anzuwenden. Aber den Hintergrund dafür, dass diese Debatte so stark aufgekommen ist, darf man nicht vergessen. Wir haben eine lang anhaltende Diskussion in den Lebenswissenschaften, es werden Forschungsprojekte, Publikationen und vieles andere in aufwendigen Peer-Review-Verfahren begutachtet und schließlich stellt man fest, dass Ergebnisse nicht reproduzierbar sind, dass die Umsetzung in die Anwendung nicht funktioniert. Ähnliches konnten wir in der Ökonomie sehen. Es werden Artikel in hochkarätigen Journals veröffentlicht und anschließend kommt eine Finanzkrise und die ganze Disziplin steht mehr oder weniger ratlos da. Die internen Begutachtungsprozesse, das reine Fokussieren darauf, ob es wissenschaftlich gut oder schlecht ist, lässt einen großen Bereich draußen, nämlich die Frage: Ist das Wissen, das hier produziert wird, auch in anderen Kontexten anwendbar? Funktionieren die Orientierungen, die aus dem Gutachtersystem kommen?

Noch etwas anderes: Man darf nicht vergessen, dass die Begutachtungen nicht gleich verteilt sind im Wissenschaftssystem. Wenige Gutachter begutachten extrem viel. Die Zeitspanne, die begehrten Gutachtern pro Gutachten zur Verfügung steht, ist erschreckend gering. Außerdem: Komischerweise gibt es keinen Weg, systematisch Begutachtungen zu lernen. Alle gehen davon aus, dass Wissenschaftler das irgendwie können, weil sie ja im Laufe ihrer Sozialisation dauernd irgendetwas bewerten. Das finde ich problematisch. Auch Begutachten muss man lernen.

Schulungen zur Begutachtungskompetenz?

Hartmann: Ich denke auch, dass das ein wichtiger Punkt ist. Wir haben überlegt, ob wir mal eine solche Schulung anbieten, und fanden es interessant, dass die eine Hälfte „Ja“ sagt und die andere Hälfte „Nein, das brauchen wir nicht“.

Wehrberger: Die DFG hat in den letzten Jahren 80 Prozent der Gutachterinnen und Gutachter weniger – teilweise deutlich weniger – als einmal im Jahr angefragt. Es ist also nicht so, dass nur ein paar große Kapitäne entscheiden, wo das Schiff langgeht. Schulungen bieten wir tatsächlich im kleinen Maßstab an, vielleicht sollten wir es mehr tun. Es gibt jedes Jahr ein Jahrestreffen der Emmy-Noether-Geförderten, das ist ja Hochschullehrenachwuchs. Die Teilnehmer bekommen einen Workshop zum Thema Begutachtung angeboten.

Lange: Die Ungleichverteilung im System und das Thema Schulungen hängen ja möglicherweise zusammen. Wir haben auch diese Konzentration wahrgenommen. Das Problem ist, dass die Zahl von Begutachtungen insgesamt wächst und diese sich bei denselben Leuten konzentrieren. Das ist unsere Beobachtung, die ich aber nicht quantifizieren kann. Dafür gibt es mindestens zwei Erklärungsmöglichkeiten. Einerseits orientiert man sich bei der Suche nach Gutachtern gerne an bewährten Personen. Andererseits wissen wir alle, dass Begutachtungen manchmal ja im Wesentlichen einen legitimierenden Effekt erzeugen sollen. Die Entscheidung soll hinterher unangreifbar sein und da suche ich mir natürlich am besten eine Person mit viel Reputation.

Wir haben deshalb in dem bereits erwähnten Positionspapier empfohlen, dass es wichtig wäre, auch jüngere Wissenschaftlerinnen und Wissenschaftler frühzeitig an das Begutachten heranzuführen. Man könnte darüber hinaus noch die Frage stellen: Ist es eigentlich wirklich so, dass die besten Gutachten von den besten Forscherinnen und Forschern geschrieben werden? Ich bin da nicht hundertprozentig sicher. Wenn wir in ein anderes gesellschaftliches Teilsystem gucken, zur Musik: Niemand würde behaupten, dass Musiker auch die besten Musikkritiker sind. //



Foto: Taru Huhkio / unsplash.com